XLV INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION, COMMUNICATION AND ENERGY SYSTEMS AND TECHNOLOGIES

www.icestconf.org

ICEST 2010

23 - 26 June, Ohrid, Macedonia

Proceedings of Papers

VOLUME 1

Bitola, 2010

XLV INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION,

COMMUNICATION AND ENERGY SYSTEMS AND TECHNOLOGIES

- ICEST 2010 -

Proceedings of Papers	**– Volume 1 of 2 volumes**
Editor:	**Prof. Dr. Cvetko Mitrovski**
Technical Editor:	**Jove Pargovski**
Published by:	**Faculty of Technical Sciences – Bitola**
Printed by:	**OFFICE 1 - BITOLA**
Print run:	**50**
ISBN:	**978-9989-786-57-0**

CIP - Каталогизација во публикација
Национална и универзитетска библиотека "Св. Климент Охридски", Скопје

621.39(063)
004.383.3(063)
004.73/.78(063)
006.91:[004.85:004.7(063)
621.313(063)
681.51(063)
621.38(063)
620.97(063)

XLV International Scientific Conference on Information, Communication and Energy Systems and Technologies (45 ; 2010 ; Ohrid) ICEST 2010 : proceedings of papers / XLV International Scientific
Conference on Information, Communication and Energy Systems and Technologies, 23-26 Juni, Ohrid, Macedonia ; [editor Cvetko Mitrovski]. - Bitola : Faculty of Technical Sciences, 2010. - 2 св. (956 стр.) : илустр. ; 29 см

Фусноти кон текстот. - Библиографија кон трудовите. - Регистри

ISBN 978-9989-786-57-0 (вол. I)
ISBN 978-9989-786-58-7 (вол. II)
1. Mitrovski, Cvetko [уредник]
а) Телекомуникации - Собири б) Дигитална обработка на сигнали - Собири в) Компјутерски мрежи - Интернет - Собири г) Метрологија - Системи за далечинско управување - Собири д) Електрични машини - Собири ѓ) Автоматски контролни системи - Собири ђ) Електроника - Собири е) Обновливи извори на енергија - Собири
COBISS.MK-ID 84633354

WELCOME TO ICEST 2010

Dear Colleagues,

On behalf of the Technical Program Committee it is my pleasure to invite you to attend the XLV International Scientific Conference on Information, Communication and Energy Systems and Technology **ICEST-2010**, which will be held from 23 to 26 June at the Faculty of Tourism and Organizational Sciences - Ohrid, Macedonia.

The ICEST Conference gained its acronym as a successor of EIST Conference in 2002, when the Faculty of Electronics from Niš, Serbia, joined the Faculty of Communication and Communication Technologies from Sofia, Bulgaria and the Faculty of Technical Sciences from Bitola, Macedonia, in organizing the series of Conferences.

As to the previous five ICEST Conferences, we have ever increasing number of papers and authors from institutions all over the world. This year we have 193 authors (and coauthors) from 15 countries with more than 270 scientific and application oriented papers out of which the Conference International Reviewer Board selected 223 papers for oral and poster presentations.
In spite of the tight Conference schedule I hope that there will be enough time for social activities which will enable further strengthening of the professional and personal relations among the ICEST participants for the benefit of the institutions and the countries they are coming from.

At last, in the name of the Conference Technical and Program Committee I want to express my gratitude to the authors and reviewers who have contributed to maintain the high standards for ICEST, and to all others who were involved in organization of the Conference.

I am sure that the beautiful and picturesque Ohrid will give you all, an inspiration for fruitful work and pleasant time during the ICEST Conference.

I am looking forward for seeing you in Ohrid.

Prof. D-r Cvetko Mitrovski

Conference Chairman

XLV INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION, COMMUNICATION AND ENERGY SYSTEMS AND TECHNOLOGIES

www.icestconf.org

ICEST 2010

23 - 26 June, Ohrid, Macedonia

Organized by

Faculty of Technical Sciences, Bitola, Macedonia

Faculty of Electronic Engineering, Niš,Serbia

Faculty of Communications and Communication Technologies, Sofia, Bulgaria

Under auspices of

Macedonian Ministry of Education and Science

In cooperation with:

Macedonian IEEE section

TECHNICAL PROGRAM COMMITTEE

General Chairman:

Mitrovski C.,	University “St. Kliment Ohridski” Bitola, Macedonia

Vice Chairmen:

Milovanović B.,	University of Niš, Serbia
Arnaudov R.,	Technical University of Sofia, Bulgaria

Members:

Aćimović-Raspopović V.,	University of Belgrade, Serbia
Altimirsky E.,	Technical University of Sofia, Bulgaria
Antić D.,	University of Niš, Serbia
Arsov Lj.,	University "Ss. Cyril and Methodius", Skopje, Macedonia
Bogdanov M.,	University "Ss. Cyril and Methodius", Skopje, Macedonia
Deskovski S.,	University “St. Kliment Ohridski” Bitola, Macedonia
Dimcev V.,	University "Ss. Cyril and Methodius", Skopje, Macedonia
Dimitrijević B.,	University of Niš, Serbia
Dimitrov D.,	Technical University of Sofia, Bulgaria
Dobrev D.,	Technical University of Sofia, Bulgaria
Dončov N.,	University of Niš, Serbia
Doseva A.,	Technical University of Sofia, Bulgaria
Gavrovski C.,	University "Ss. Cyril and Methodius", Skopje, Macedonia
Goleva R.,	Technical University of Sofia, Bulgaria
Gušev M.,	University "Ss. Cyril and Methodius", Skopje, Macedonia
Iliev I.,	Technical University of Sofia, Bulgaria
Janković D.,	University of Niš, Serbia
Jeftić M.,	University of Niš, Serbia
Katić V.,	University of Novi Sad, Serbia
Kuntchev R.,	Technical University of Sofia, Bulgaria
Lutovac M.,	University of Novi Pazar, Serbia
Markovski A.,	University “St. Kliment Ohridski” Bitola, Macedonia
Mirtchev S.,	Technical University of Sofia, Bulgaria
Mitrevski P.,	University “St. Kliment Ohridski” Bitola, Macedonia
Nedelkovski I.,	University “St. Kliment Ohridski” Bitola, Macedonia
Paunović Đ.,	University of Belgrade, Serbia
Pencheva E.,	Technical University of Sofia, Bulgaria
Perić Z.,	University of Niš, Serbia
Pleshkova-Bekiarska S.,	Technical University of Sofia, Bulgaria
Pranchov R.,	Technical University of Sofia, Bulgaria
Pulkov V.,	Technical University of Sofia, Bulgaria
Reljin I.,	University of Belgrade, Serbia
Stanković Z.,	University of Niš, Serbia
Stojčev M.,	University of Niš, Serbia
Stojmenov L.,	University of Niš, Serbia
Stoyanov G.,	Technical University of Sofia, Bulgaria
Tasić D.,	University of Niš, Serbia
Trpezanovski Lj.,	University “St. Kliment Ohridski” Bitola, Macedonia
Tsankov B.,	Technical University of Sofia, Bulgaria
Uzunov I.,	Technical University of Sofia, Bulgaria
Zieleznik L.,	Technical University of Sofia, Bulgaria

ORGANIZING COMMITTEE

Acevski N.,	University “St. Kliment Ohridski” Bitola, Macedonia
Atanasovski M.,	University “St. Kliment Ohridski” Bitola, Macedonia
Ceselkoska V.,	University “St. Kliment Ohridski” Bitola, Macedonia
Dimitrijević T.,	University of Niš, Serbia
Jolevski I.,	University “St. Kliment Ohridski” Bitola, Macedonia
Kostov M.,	University “St. Kliment Ohridski” Bitola, Macedonia
Lubih L.,	Technical University of Sofia, Bulgaria
Milijić M.,	University of Niš, Serbia
Pargovski J.,	University “St. Kliment Ohridski” Bitola, Macedonia
Petkovski M.,	University “St. Kliment Ohridski” Bitola, Macedonia
Radevska P. M.,	University “St. Kliment Ohridski” Bitola, Macedonia
Stevanovski B.,	University “St. Kliment Ohridski” Bitola, Macedonia

CORRESPONDENCE

Conference Chairman:
Prof. Dr. Mitrovski Cvetko
Faculty of Technical Sciences
Ivo Lola Ribar bb, 7000 Bitola, Macedonia
phone: +389 47 207 711
e-mail: cvetko.mitrovski@uklo.edu.mk

Organizing Committee Chairman:
Prof. Dr. Mitrevski Pece
Faculty of Technical Sciences
Ivo Lola Ribar bb, 7000 Bitola, Macedonia
phone: +389 47 207 149
fax: +389 47 203 370
e-mail: pece.mitrevski@uklo.edu.mk

Technical Editor:
Assist. Mr. Pargovski Jove
Faculty of Technical Sciences
Ivo Lola Ribar bb, 7000 Bitola, Macedonia
phone: +389 47 207 711
fax: +389 47 203 370
e-mail: jove.pargovski@uklo.edu.mk

UPDATED INFORMATION

Updated information can be obtained from the official Conference web site:
http://www.icestconf.org

LIST OF ICEST 2010 REVIEWERS

Aleksandar Bekiarski, assoc. prof.
Technical University of Sofia, Bulgaria

Aleksandar Markovski, Prof. Dr.
University “St. Kliment Ohridski” Bitola, Macedonia

Aleksandar Risteski, Prof. Dr.
University "Ss. Cyril and Methodius", Skopje, Macedonia

Aleksandar Tsenov, assoc. prof.
Technical University of Sofia, Bulgaria

Anelia Doseva, assoc. prof.
Technical University of Sofia, Bulgaria

Antoaneta Popova, assoc. prof.
Technical University of Sofia, Bulgaria

Boncho Bonev, assoc. prof.
Technical University of Sofia, Bulgaria

Boris Tsankov, Prof. Dr.
Technical University of Sofia, Bulgaria

Borislav Pankov, Prof.Dr
Technical University of Sofia, Bulgaria

Boyanka Nikolova, assoc. prof.
Technical University of Sofia, Bulgaria

Božidar Dimitrijević, Prof. Dr.
University of Niš, Serbia

Bratislav Milovanović, Prof. Dr.
University of Niš, Serbia

Cvetan Gavrovski, Prof. Dr.
University "Ss. Cyril and Methodius", Skopje, Macedonia

Dimitar Dimitrov, Prof. Dr.
Technical University of Sofia, Bulgaria

Dimitar Taskovski, Prof. Dr.
University "Ss. Cyril and Methodius", Skopje, Macedonia

Dobri Dobrev, Prof. Dr.
Technical University of Sofia, Bulgaria

Đorđe Paunović, Prof. Dr.
University of Belgrade, Serbia

Dragan Antić, Prof. Dr.
University of Niš, Serbia

Dragan Janković, Prof. Dr.
University of Niš, Serbia

Dragan Tasić, Prof. Dr.
University of Niš, Serbia

Georgi Stoianov, Prof. Dr.
Technical University of Sofia, Bulgaria

Goce Arsov , Prof. Dr.
University "Ss. Cyril and Methodius", Skopje, Macedonia

Goce Shutinoski , Prof. Dr.
University "Ss. Cyril and Methodius", Skopje, Macedonia

Igor Nedelkovski, Prof. Dr.
University “St. Kliment Ohridski” Bitola, Macedonia

Ilia Iliev, assoc. prof.
Technical University of Sofia, Bulgaria

Irini Reljin, Prof. Dr.
University of Belgrade, Serbia

Ivailo Atanasov, assoc. prof.
Technical University of Sofia, Bulgaria

Josif Kosev , Prof. Dr.
University "Ss. Cyril and Methodius", Skopje, Macedonia

Jovan Stefanovski , Dr.
JP Strezevo Bitola

Leonid Stojmenov, Prof. Dr.
University of Niš, Serbia

Lidia Jordanova, assoc. prof.
Technical University of Sofia, Bulgaria

Liljana Gavrilovska , Prof. Dr.
University "Ss. Cyril and Methodius", Skopje, Macedonia

Ljuben Janev, Prof. Dr.
University "Ss. Cyril and Methodius", Skopje, Macedonia

Ljupco Panovski , Prof. Dr.
University "Ss. Cyril and Methodius", Skopje, Macedonia

Ljupco Trpezanovski, Prof. Dr.
University “St. Kliment Ohridski” Bitola, Macedonia

Marija Cundeva-Blajer, Prof. Dr.
University "Ss. Cyril and Methodius", Skopje, Macedonia

Marin Nedelchev, assoc. prof.
Technical University of Sofia, Bulgaria

Metodija Kamilovski , Prof. Dr.
University "Ss. Cyril and Methodius", Skopje, Macedonia

Mile Stojčev, Prof. Dr.
University of Niš, Serbia

Milun Jeftić, Prof. Dr.
University of Niš, Serbia

Mirka Radevska P., Prof. Dr.
University "St. Kliment Ohridski" Bitola, Macedonia

Miroslav Lutovac, Prof. Dr.
University of Novi Pazar, Serbia

Mitko Kostov, Prof. Dr.
University "St. Kliment Ohridski" Bitola, Macedonia

Momcilo Bogdanov , Prof. Dr.
University of Skopje, Macedonia

Nebojša Dončov, Prof. Dr.
University of Niš, Serbia

Nikolce Acevski, Prof. Dr.
University "St. Kliment Ohridski" Bitola, Macedonia

Pece Mitrevski, Prof. Dr.
University "St. Kliment Ohridski" Bitola, Macedonia

Rosen Miletiev, assoc. prof.
Technical University of Sofia, Bulgaria

Rositsa Goleva, assoc. prof.
Technical University of Sofia, Bulgaria

Roumen Arnaudov, Prof. Dr.
Technical University of Sofia, Bulgaria

Roumen Kounchev, Prof. Dr.
Technical University of Sofia, Bulgaria

Roumen Mironov, assoc. prof.
Technical University of Sofia, Bulgaria

Seferin Mirchev, assoc. prof.
Technical University of Sofia, Bulgaria

Sofija Bogdanova , Prof. Dr.
University "Ss. Cyril and Methodius", Skopje, Macedonia

Tomislav Dzhekov , Prof. Dr.
University "Ss. Cyril and Methodius", Skopje, Macedonia

Ventsislav Trifonov, assoc. prof.
Technical University of Sofia, Bulgaria

Veska Georgieva, assoc. prof.
Technical University of Sofia, Bulgaria

Vesna Ceselkoska, Prof. Dr.
University "St. Kliment Ohridski" Bitola, Macedonia

Vladanka Aćimović-Raspopović, Prof. Dr.
University of Belgrade, Serbia

Vladimir Dimcev, Prof. Dr.
University "Ss. Cyril and Methodius", Skopje, **Macedonia**

Vladimir Katić, Prof. Dr.
University of Novi Sad, Serbia

Vladimir Poulkov, assoc. prof.
Technical University of Sofia, Bulgaria

Zoran Hadzi-Velkov , Prof. Dr.
University "Ss. Cyril and Methodius", Skopje, Macedonia

Zoran Ivanovski , Prof. Dr.
University "Ss. Cyril and Methodius", Skopje, Macedonia

Zoran Perić, Prof. Dr.
University of Niš, Serbia

Zoran Stanković, Prof. Dr.
University of Niš, Serbia

TABLE OF CONTENTS

VOLUME 1

Telecommunications Systems and Technology I

Telecommunications Systems and Technology II

Radio Communications, Microwave Technique and Antennas I

Radio Communications, Microwave Technique and Antennas II

Signal Processing I

Signal Processing II

Computer Systems and Internet Technologies I

Computer Systems and Internet Technologies II

Computer Systems and Internet Technologies III

Control Systems

Metrology and Remote Sensing

Electrical Machines

POSTER SESSION I - Control Systems

SESSION TST I

Telecommunications Systems and Technology I

CT-MAC: Energy-Efficient Contention-based MAC Protocol for Wireless Sensor Networks

Milica D. Jovanovic[1] and Goran Lj. Djordjevic[2]

***Abstract* – This paper presents Control-Tone MAC (CT-MAC), a scheduled contention-based medium access protocol especially designed for Wireless Sensor Networks (WSNs). Similarly to other contention-based MAC protocols with common active periods (e.g. S-MAC, T-MAC, and SCP-MAC), CT-MAC coordinates sensor nodes into sleep/wakeup schedules, allowing them to remain awake only for brief contention periods. Unlike most of the current solutions, CT-MAC employs short control tones, instead of control packets (e.g. RTS/CTS) in order to realize an energy-efficient contention resolution mechanism in multi-hop networks. The simulation results demonstrate that CT-MAC significantly reduces energy waste due to collisions, overhearing and idle listening in respect to SCP-MAC.**

***Keywords* – wireless sensor networks, medium access control, energy efficiency, contention protocol.**

I. INTRODUCTION

Recent advances in wireless communications, low-power design, and MEMS-based sensor technology have enabled the development of relatively inexpensive and low power wireless sensor nodes. The common vision is to create a large wireless sensor network (WSN) through *ad-hoc* deployment of hundreds or thousands of such tiny devices able to sense the environment, compute simple task and communicate with each other in order to achieve some common objective, like environmental monitoring, target tracking, detecting hazardous chemicals and forest fires, monitoring seismic activity, military surveillance [1]. The primary objective in WSN design is maximizing node/network lifetime, leaving the other performance metrics as secondary objectives. Since the communication of sensor nodes is more energy consuming than their computation, it is a primary concern to minimize communication while achieving the desired network operation.

A medium access control (MAC) protocol decides when competing nodes may access the radio channel, and tries to ensure that no two nodes are interfering with each other's transmissions. MAC protocols for WSN usually trade off performance (latency, throughput, fairness) for cost (energy efficiency, reduced algorithmic complexity), while providing a good scalability and some limited adaptability for topology changes [2]. Collisions, overhearing, and idle listening are the main types of energy waste for sensor nodes that occur during medium access. Collision occurs if a node receives multiple transmissions at the same time. Idle listening occurs if a node listens to the medium when there is no transmission, whereas an overhearing happens when a node receives a data packet transmission even if it is not the intended recipient of this transmission.

One important approach is based on common active/sleep periods. Nodes use active periods for communication and the sleep periods for saving energy. At the beginning of each active period, nodes contend for the medium using contention-based approaches. Only nodes participating in data transfer remain awake after contention periods, while others can sleep. Contention-based media access mechanisms of various kinds are employed in a plurality of different MAC protocol for WSNs. For instance, S-MAC employs an explicit contention mechanism which requires carrier sense and the use of control packets, such as Request to Send (RTS) and Clear to Send (CTS) [3]. Although RTS/CTS can alleviate the hidden terminal problem, it incurs high overhead because data packets are typically very small in WSN [4]. SCP-MAC replaces RTS/CTS control packets with short wake up tones [5]. When a node wakes up during the common active schedule and does not find a tone, it goes back to sleep. In order to improve the contention performances, SCP-MAC introduce a two level contention window. Before sending the tone, a node performs carrier sense by randomly select a slot within the first contention window. If the channel is idle, then the node sends the tone to wake up the receiver. Only nodes that successfully send wakeup tones will enter the second contention window. Such nodes randomly select a slot in the second contention window and then perform a carrier sense; if they find the channel idle then they transmit the data. The major advantage of spitted contention phase is lower collision probability with shorter overall contention time. Although the SCP-MAC is more energy efficient than S-MAC, it does not implement an appropriate mechanism to alleviate the hidden terminal problem that normally exists in multi-hop networks.

In this paper, we present CT-MAC, abbreviated from Control Tone MAC, which uses short signal tone transmissions to implement an energy-efficient contention mechanism. In CT-MAC we use control tones not only to wakeup intended receivers but to implement efficient handshake mechanism among competing nodes arranged in a multi-hop network.

II. THE PROTOCOL DESCRIPTION

In our contention protocol, competition is accomplished by exchanging short *control tones* among competing nodes. Control tones play a similar role as RTS/CTS packets in traditional collision avoidance handshaking mechanism with

[1, 2] Milica D. Jovanovic and Goran Lj. Djordjevic are with Faculty of Electronic Engineering, University of Nis, Aleksandra Medvedeva 14, P.O. Box 73, 18000 Nis, Serbia, E-mails: {milica.jovanovic, gdjordj}@elfak.ni.ac.rs

the following two important differences: (a) control tone is not a packet; it is a simple flag signal and encodes no data, and (b) control tones may collide with one another without affecting their functionality. Control tones are transmitted over a shared wireless medium and nodes within the transmission range individually decide how to react to those tones, based on their own states at that time. Thus, different from a handshake in other contention protocols, the handshake in CT-MAC is implicit in a sense that a node that hears a control tone does not know the identity of the sending node neither it knows whether it hears one tone or a mix of several control tones.

The contention period is slotted. In each slot, a node can either transmits the control tone, listen for a tone, or stay inactive. The duration of the slot is equal $T_s = T_d + T_g$, where T_d is the duration of the control tone, and T_g is the guard time which compensate time discrepancy among nodes. Note that the control tone can be very short in duration, just enough for the listening node to check if the medium is busy (typically 2 – 3 ms). The guard time depends on clock drift rate and the efficiency of synchronization mechanism employed. Typical value of the guard time is in order of several milliseconds.

During contention, a node can be put into one of the following five states: P_SEND, P_REC, IDLE, SEND, and REC. A node enters contention period either in state P_SEND or P_REC. The P_SEND is for potential sender, i.e. node which has data packet to send in the current frame, and the P_REC is for potential receiver, i.e. node which does not have data to send. Node leaves the contention period in one of three final states, IDLE, SEND or, REC, which controls its behavior during data transmission phase. A node in IDLE state will be put into sleep mode until the next frame; a node in state SEND is a potential sender which is chosen to send its data packet, and node in state REC is a node that is instructed to enter data transmission phase in the listening mode in order to receive a data packet intended to it (if any).

Slots of the contention period are arranged into two contention windows, CW1 and CW2, with N slots in CW1, and $2M$ slots in CW2, where N and M are network-wide constants. Nodes that survive the contention window CW1 enter the CW2. Nodes compete for the medium by following a different contention procedure in each contention window. In what follows, we provide a detailed description of these two contention procedures.

Contention procedure for CW1. CT-MAC requires that each node is assigned a listening slot in contention window CW1. Each node also knows the numbers of listening slots assigned to all neighbors in its transmission range. A potential sender announces its intention to send a data packet by transmitting the control tone in the listening slot of the intended receiver. On the other hand, a node treats the presence of a control tone in its listening slot as an indication that there is **at least one** adjacent node that has packet for it.

The contention procedure is illustrated by flow diagram in Fig 1. A node enters this procedure only for its own listening slot and listening slots of neighboring nodes. A potential receiver, say node R, skips all slots (i.e. stays inactive) until its listening slot (denoted with *MySlotID* in Fig. 1). If a control tone is not heard during this slot, node will modify its state to IDLE. Otherwise, if a control tone is heard, the node R will change its state to REC. Once put in state REC, the node will start to monitor the channel during listening slots of neighboring nodes. If a control tone is heard in any of monitored slots, note R will change its state to IDLE and quit the competition.

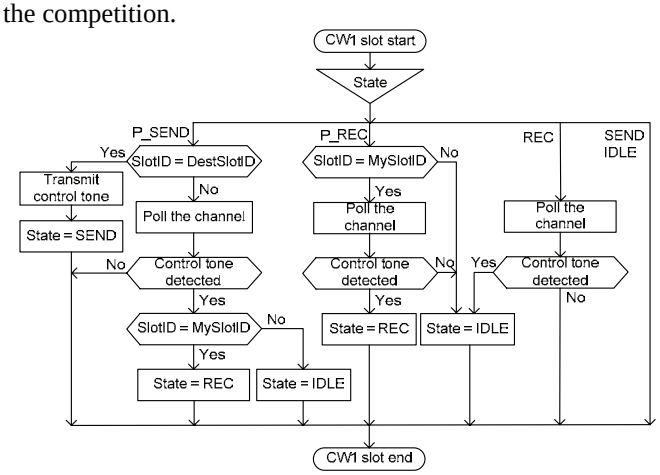


Fig. 1. Flow diagram for CW1 contention procedure.

While waiting for the listening slot of its intended receiver (denoted with *DestSlotID* in Fig. 1), a potential sender, say node S, monitors the channel for control tone transmissions during listening slots of neighboring nodes. If a control tone is heard, node S will quit the competition in IDLE state. An idle channel will allow the node S to proceed until the listening slot of its intended receiver, when it transmits the control tone and modifies its state to SEND. Once in SEND state, the node S will skip all slots until the end of CW1. It may happen that the listening slot of the node S precedes the listening slot of its intended receiver. In that case, S will poll the channel in its own listening slot. If a control tone is not heard, S will resume its operation as potential sender. Otherwise, if control tone is heard, the S will change its state to REC and continues slot monitoring, but now as a potential receiver.

The contention procedure for CW1 in a network with 8 nodes is illustrated in Fig 2(a) represents an initial condition. Potential senders are represented with dark gray circles, and potential receivers are represented with light gray circles. Every two nodes within the transmission range of one another are connected with an edge. Arrow on the edge points to the intended receiver. The numbers enclosed with brackets attached to each node denote the number of the listening slot assigned to that node. The first control tone transmission occurs in slot 2 (i.e. in the listening slot of node N_2). In this slot, node N_1 transmits the control tone to notify node N_2, along with all potential senders within the range, of its intent to send a data packet to N_2. As a consequence, N_1 changes its state to SEND. Presence of a control tone in its own listening slot moves N_2 into REC state. In contrast to N_1, which will stay inactive until the end of CW1, N_2 continues to monitor listening slots of neighboring nodes. Slot 2 is also monitored by N_2's surrounding potential senders, N_3 and N_4. Since N_3 hears the tone it changes its state to IDLE - giving up its data transmission in order to prevent a possible collision at node N_2. On the other hand, node N_4 does not hear anything and continues its operation as potential sender. The next control tone is transmitted by node N_7 during the slot 3. This

transmission affects only nodes N_7 and N_8 since there is no other node that hears the tone. Note that N_5 is potential receiver which will stay inactive until its own listening slot. The third and the final control tone transmission takes place in slot 4. Now we have two potential senders, N_4 and N_6, transmitting control tone in the same slot. In spite of tone collision, N_5 detects channel activity and changes its state to REC. Nodes N_4 and N_6 are not aware of one another transmission, and both set its state to SEND. Given that potential senders N_3 and N_7 have already determined their final state in CW1, the node N_2 is the only N_5's neighbor that monitors its listening slot. Since N_2 hears the tone transmitted by N_4, it becomes aware that there are now two senders in its neighborhood and that the collision of their data packet is inevitable. This is the reason why it changes its state from REC to IDLE. Until the end of CW1 there are no control tone transmissions, and no state changes. The node state distribution after CW1 can be seen in Fig 2(b). Note that if data transmission starts immediately after CW1, there will be one successful data transfer (from N_7 to N_8) in addition to one collision (at node N_5) and one unnecessary data packet transmission (N_1).

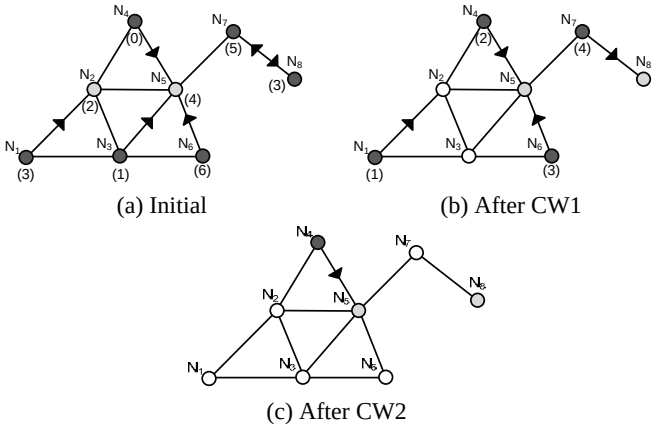


Fig. 2. Distribution of sending, receiving and idle nodes during contention period of CT-MAC – an example.

Contention procedure for CW2. Although contention in CW1 eliminates a number of conflicting conditions, there are still situations that may cause energy loss during data transfer. This is the reason why we introduce the second contention window, CW2. Active participants in this round of contention are nodes that survive competition in CW1, i.e. those in SEND and REC state. CW2 is divided in two equal sized sub-windows, CW2a and CW2b. Slots in both sub-windows are numbered form 0 to M-1. The task of a sending node in CW2a is to randomly pick a slot number from the range [0, M-1], remember that number for future use, and then transmit control tone in the selected slot. The operation of a receiving node is a little bit involved. A receiving node wakes up at the beginning of CW2a and continues to listen in each slot until it hears a control tone. It remembers the number of the corresponding slot, and then skips remaining slots in CW2a.

During sub-window CW2b, competing nodes change their roles - nodes in REC state are those who will transmit, and nodes in state SEND are those who will listen for control tones. The only responsibility of a receiving node in CW2b is to transmit a control tone in slot *i*, where *i* is the number of slot in CW2a during which it was heard the tone. On the other hand, a sending node wakes up at the beginning of CW2a and continues to listen in each slot until it hears a control tone. If a control tone is heard in the slot with the same number as one that it was used to transmit a control tone in CW2a, the sending node will keep SEND state skipping the remaining slots in CW2b. Otherwise, if a control tone is heard in some other slot, the sending node will quit contention in IDLE state.

The contention procedure for CW2 is primarily designed to eliminate collisions that occur when multiple senders transmit data packets to their common neighbor. This situation is illustrated in Fig. 2(b) where both nodes N_4 and N_6 are about to send data packet to node N_5. The numbers enclosed with brackets now denote the slot numbers in CW2a chosen by sending nodes. The collision will be prevented if these two nodes chose to transmit their control tones during different slots in CW2a. The winning sender is one which selects the slot with the smallest number in CW2a.

The reason why a sending node listens not only during the slot of CW2b in which it expects a response from its intended receiver, but also during all previous slots is to prevent collisions that its data packet may induce at some other receiving nodes within its transmission range. Let consider operation of node N_7 in Fig. 2(b). While waiting for a response from its intended receiver N_8, node N_7 is in slot monitoring mode. In slot 2, node N_5 transmits a control tone as a response to the tone that it was heard in the corresponding slot of CW2a. Since N_7 is within the transmission range of N_5, it will hear that tone and modify its state to IDLE. Node N_8 cannot know that N_7 is no longer in SEND state, and it will keep its REC state after transmitting confirmation control tone in slot 4. This creates a condition for idle listening. However, idle listening is less energy costly than a packet collision that would happen if N_7 was not given up its data transmission.

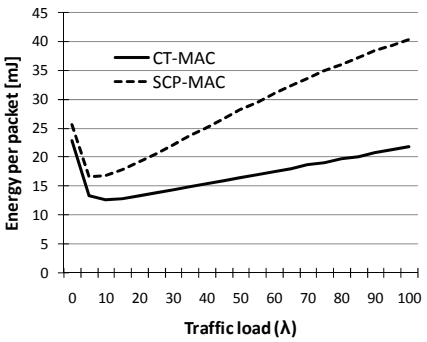
Finally, let see what happens with node N_1. Node N_1 enters CW2 in SEND state, even though its intended receiver, N_2, is already in IDLE state. Even though N_1 transmits a control tone in CW2a (during slot 1), there is no confirmation control tone in CW2b since N_2 is already quit its activity in this frame. Because of that, N_1 quits the contention and postpones its data transmission until the next frame. The final outcome of the contention resolution can be seen in Fig. 2 (c). The only survival sender-receiver pair is N_4-N_5 in addition of one receiving node, N_8, predestinated to idle listening.

III. Performance Evaluation

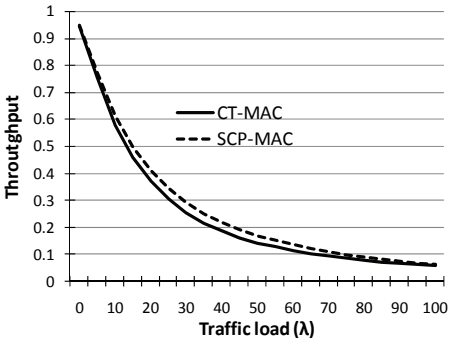
We implement CT-MAC and SCP-MAC protocols in a custom WSN simulator build in C++, and conduct several experiments to evaluate their performances. Our evaluation is based on the simulation of 200 nodes randomly distributed in an area of 100x100 m^2. The average number of nodes within radio transmission range, i.e. the average size of one-hop neighborhood, in this network is 6. Data rate of 20 Kbps is assumed, and data packet length is fixed to the value of 32 bytes. Both analyzed protocols divide contention period into two contention windows, CW1 and CW2. We configure both protocols with the 32 slots in CW1 and 16 slots in CW2. The

Parameter	Value
Time needed to poll channel once	3 ms
Time to transmit a control tone	5 ms
Time to transmit or receive a data packet	40 ms
Time to transmit/receive an ACK packet	18 ms
Time to receive MAC header	16 ms
Idle channel timeout period	8 ms

(a) Experimental parameters for CT-MAC



(b) Energy consumption per transferred packet.



(c) Throughput over varying traffic load.

Fig. 3. Results of performance evaluation.

receiver responds with an acknowledgement (ACK) packet after each successful data packet transmission. Also, the receiver performs overhearing avoidance by examining the destination address of a packet immediately after receiving its MAC header. If packet is destined to another node, it immediately stops the reception. Relevant timing parameters used in our analysis are shown in Fig 3(a).

We first compare the energy performance of CT-MAC and SCP-MAC by varying the traffic load. To vary the degree of traffic load, we vary the percentage, λ, of nodes in the network that generate data packet during each frame. During simulation we count successfully transferred data packets, and record all nodes' activities in order to estimate energy used by each node. Figure 3(b) shows the average energy consumption per transferred data packet as the traffic load increases.

There are two types of energy waste, energy consumed during contention period, and energy consumed during data transfer period. The contention energy is present even when data traffic is absent since each node must poll the channel once during each contention period. Under the light traffic load, almost all generated data packet are successfully transferred with small increase of node activity during contention period. As a result, the energy per packet decreases with the increase of traffic load. Further increase of traffic load results in steady increase of energy per packet. From one hand, contention energy increases since a larger number of nodes are involved in contention. From the other hand, energy consumed during data transfer period increases too, since collisions, packet overhearing and idle listening happen more often. It is clear from Fig. 3(b) that, with a more aggressive network-wide suppression of collisions and packet overhearing, the CT-MAC always outperforms SCP-MAC in terms of the energy efficiency, especially when the traffic load is high. CT-MAC uses only 10% less energy than SCP-MAC to handle the traffic load less than λ=5%. As traffic load increases, the energy saving is greater. For example, when λ is 30%, CT-MAC consumes 35% less energy than SCP-MAC per transferred data packet. Under the maximum traffic load, CT-MAC uses only about half the energy of SCP-MAC.

In the second experiment, the results of which are shown Fig. 3(c), we study the influence of the traffic load on data throughput. We define the throughput as the ratio between the number of successfully transferred packets and the number of packets that are generated per frame. Note that, even though the throughput falls, the throughput that is achieved by CT-MAC is constantly higher than in SCP-MAC case. This is due to the fact that SCP-MAC ignores the effect of hidden terminals, which limits its efficiency in a multi-hop scenario. On the other hand, CT-MAC utilizes more elaborate contention mechanism which prevents most of collisions that might be caused by hidden terminals. This not only saves the energy but also increases data throughput.

IV. Conclusion

In this paper, we describe energy-efficient, contention-based MAC protocol, called CT-MAC. The novelty of our approach resides in its use of short control tones instead of control packets in order to implement a contention scheme in a WSN. Although CT-MAC does not alleviate the hidden terminal problem entirely, it significantly reduces probability of collisions and lowers energy waste due to overhearing and idle listening in respect to SCP-MAC. In this paper, we do not consider the problems of node synchronization and contention slot assignment. This remains for a future work.

Acknowledgement

This work was supported by the Serbian Ministry of Science and Technological Development, project No. TR - 11020 - "Reconfigurable embedded systems".

References

[1] F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: A survey", *Computer Networks*, vol. 38, no. 3, pp. 393–422, 2002.

[2] W. Ye, J. Heidemann, and D. Estrin, "Medium Access Control with Coordinated Adaptive Sleeping for Wireless Sensor Networks", *IEEE/ACM Trans. Net.*, vol. 12, no. 3, pp. 493-506, June 2004.

[3] W. Ye, J. Heidemann, and D. Estrin, "An energy-efficient MAC protocol for wireless sensor networks", In *21st Conference of the IEEE Computer and Communications Societies (INFOCOM)*, vol. 3, pp. 1567–1576, June 2002.

[4] A.Woo and D.Culler, "A transmission control scheme for media access in sensor networks", In ACM MobiCom 2001, pp. 221-235, 2001.

[5] W. Ye, F. Silva and J. Heidemann, "Ultra-Low Duty Cycle MAC with Scheduled Channel Polling", In Proc. of ACM SenSys, Boulder, CO, November 2006.

Dynamic Admission Control for Improving Utilization of Parlay X Gateway

Evelina Pencheva[1] and Ivaylo Atanasov[2]

Abstract - **The paper presents a model of Parlay X gateway used for evaluation of its performance. The model considers the distributed architecture of the gateway and applies mechanisms for admission control and load balancing to prevent from overloading A new admission control algorithm is proposed to improve quality of service and node gateway utilization. Simulations are used to evaluate the behavior and the benefits of the proposed algorithm.**

Keywords – admission control, load balancing, web-based traffic, performance evaluation

I. INTRODUCTION

Parlay X is technology that allows open access to network functions. The access is provided through application programming interfaces which hide for the application developers the underlying network specifics and complexity. Third party applications can access network information and control functions by interface method invocation instead of programming network protocols. The "intelligence" is concentrated in a Parlay X gateway which provides interfaces towards applications and "talks" the control protocols toward the network.

In the context of service architecture, a common problem is overload. To avoid congestion usually an admission control mechanism is used. In [2] the authors propose a load control mechanism aimed at supporting constraints imposed by the distributed Parlay X gateway architecture. The mechanism uses preliminary defined threshold values to predict the load and decide if the message should be accepted or rejected. A paper that treats overload control for distributed web-based applications is [3]. The authors suggest a control algorithm that self-configures a dynamic constraint on the rate of incoming new sessions in order to guarantee the fulfillment of the quality requirements specified in service level agreement (SLA). In [4], a staged event-driven architecture is proposed which decomposes a complex, event-driven application into a set of stages connected by queues avoiding the high overhead associated with thread-based concurrency models, and decouples event and thread scheduling from application logic. The proxy-based overload control for web applications presented in [4] is based on measurements of metrics such as response time, throughput, and resource utilization.

[1]Evelina N. Pencheva is with the Faculty of Telecommunications, TU-Sofia, 7 Kl. Ohridski blvd, 1000 Sofia, Bulgaria, E-mail: enp@tu-sofia.bg

[2]Ivaylo I. Atanasov is with the Faculty of Telecommunications, TU-Sofia, 7 Kl. Ohridski blvd, 1000 Sofia, Bulgaria, E-mail: iia@tu-sofia.bg

This paper presents a model of Parlay X gateway providing interfaces for access to location information. The model considers the distributed gateway architecture and applies dynamic control strategy for admission of incoming application traffic. The aim is to optimize the gateway utilization considering the current application demands.

The paper is structured as follows. First in the rest of the paper, we present related work where a dynamic model for the token bucket algorithm is used in packet networks. In Section III, we present a model of Parlay X gateway and in Section IV we describe the traffic model and define node utilization function. In Section V, the dynamic control with feedback for traffic policing is presented. Section VI describes simulation parameters and Section VII discusses the obtained numerical results. At the end, we present our conclusion mentioning issues for future work.

II. RELATED WORK

A token bucket (TB) is well known mechanism used for admission control and packet filtering. In [1], the authors present a dynamic model for the TB algorithm. In a general traffic model, the traffic is observed at regular time intervals $[t_{k-1}, t_k)$ for k=1, 2, …, K. The observation periods are equal and small enough during which only one packet may arrive. The input traffic from several traffic sources is multiplexed in an access node. To prevent the node from overloading, admission control mechanism is used. The incoming traffic of each source is policed by a TB. Usually, the rate of token accumulation is constant in each observation interval. The authors present a balance equation for the state of the i-th TB and express the conforming and non conforming traffic. They suggest a feedback control strategy for dynamic change of the number of tokens transmitted to the TB per time unit. The dynamic control grants tokens according to traffic source demands.

In [2], the overload control in Parlay X gateway providing interfaces for 3rd party call control is studied. An algorithm based on the priority and utility function is proposed. However, the rejection of new service messages in terms of priority will sometimes lead to over-control, which results in part of the resources being idle.

Based on research in [1] and [2], we suggest a model of Parlay X gateway providing access to location information for 3rd parties. The model is used to evaluate the traffic load of the gateway. Our model considers the distributed architecture of Parlay X gateway with a number of processing nodes called PX-M converters. A PX-M converter converts the application request into Mobile Application Part (MAP) request and MAP response into application response. The converters work in parallel. The traffic load is balanced between PX-M

converters using Round Robin algorithm. Model includes admission control implemented by a number of TBs. We apply adaptive control to token accumulation rate to optimize the gateway utilization and to improve the quality of service.

III. Model of Parlay X gateway

The incoming traffic is generated by Service providers (SPs) which have contracts with network operator. The contract defines constraints that have to be fulfilled. The constraints include the peak and average number of application requests that should be accepted per time unit, and the maximum delay between application request and response. To be able to fulfill the constraints and to avoid congestion at the PX-M converters an Admission Control/Load Balancing (AC/LB) mechanism is used. The aim with load balancing is to distribute uniformly the load between the converters. The admission control rejects the nonconforming requests. When a SP sends a request to the Parlay X gateway, it is received by the LB/AC. The admission control, which is modeled by TB, decides whether to accept or reject the request. If the request is accepted, the load balancing decides to which PX-M converter has to be forwarded. Each accepted request has to be answered. The PX-M converter maps the request onto MAP message. Each of the PX-M converters is modeled as a single FIFO buffer with limited size. The buffer size is restricted by the maximal delay between a request and its response.

The modeled PX gateway system consists of n SPs and m PX-M converters as shown in Fig.1.

We assume that all PX-M converters have same capacity C. The Round Robin algorithm is used for load balancing because of the equal capacity of the PX-M converters. Each SP might include several applications but as the contracts are agreed between the SPs and the Parlay X gateway, the number of applications is not important. A single SP is a traffic generator to the LB/AC as far as each SP is connected to its own LB/AC.

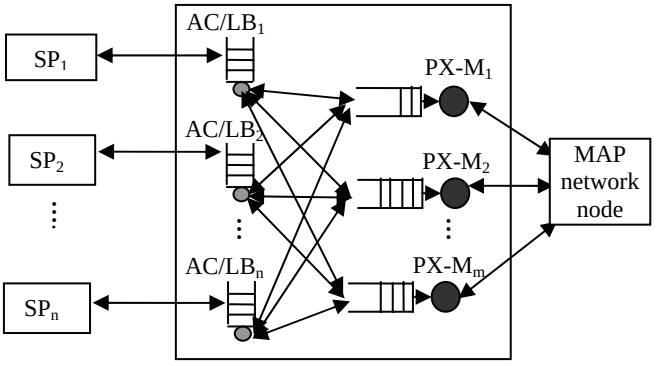


Fig.1 Model of distributed Parlay X Gateway architecture

IV Traffic model

The traffic generated by each SP is a random process observed at time t_k, as shown in Fig.2.

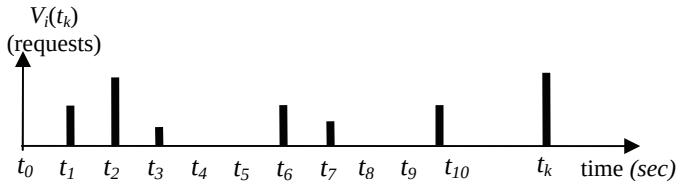


Fig.2 The traffic model of the i-th SP

The status of the token buckets is given by a vector valued function $\mu \equiv (\mu_1, \mu_2, \ldots, \mu_n)'$ of dimension n . The vector valued function μ is described by a system of equations where the dynamic token accumulation for i-th TB is given by:

$$\mu_i(t_k) = \mu_i(t_{k-1}) + [\min(u_i(t_k), T_i - \mu_i(t_{k-1}))] - V_i(t_k).I\{ V_i(t_k) \le \mu_i(t_{k-1}) + [\min(u_i(t_k), T_i - \mu_i(t_{k-1}))]\} \quad (1)$$

The indication function $I(S)$ is used to represent the truth of the statement S where $I(S) = 1$ if the statement S is true, or $I(S) = 0$ otherwise.

The conforming traffic is also given by a vector valued function of dimension n, which we denote by $G \equiv (G_1, G_2, \ldots, G_n)'$, where

$$G_i(t_k) = V_i(t_k).I\{ V_i(t_k) \le \mu_i(t_{k-1}) + [\min(u_i(t_k), T_i - \mu_i(t_{k-1}))]\} \quad (2)$$

The vector of nonconforming traffic is given by $R \equiv (R_1, R_2, \ldots, R_n)'$, where

$$R_i(t_k) = V_i(t_k) - G_i(t_k) \quad (3)$$

To formulate the status of each PX-M converter, we have to consider the load balancing mechanism. With $\sigma_i(t_k)$ we denote the state of the i-th LB that is the PX-M converter to which the conforming application request has to be forwarded, and it is given by:

$$\sigma_i(t_k) = 1 + (\sigma_i(t_{k-1}) + G_i(t_k)) \bmod m \quad (4)$$

The state of i-th LB models the control variable which depends on its previous value, the conforming traffic for the period of observation, and finally the count of converters. The state reflects the robust Round Robin principle of balancing.

With $H_{ij}(t_k)$ we denote the part of the conforming traffic distributed by the i-th access controller to the j-converter. It is given by:

$$H_{ij}(t_k) = \lfloor G_i(t_k))/m \rfloor + I[j \le \sigma_i(t_k)] \quad (5)$$

The conforming traffic forwarded to the j-th PX-M converter is given by:

$$F_j(t_k) = \sum_{i=1}^{n} H_{ij}(t_k) \quad (6)$$

The status of the j-th PX-M converter is given by the size of the queue at the PX-M converter denoted by $q_j(t_k)$ and waiting for service, that is, for onward transmission to the network. This variable is governed by the following equation:

$$q_j(t_k) = \max [q_j(t_{k-1}) - C_j(t_k - t_{k-1})/2, 0] + \min\{F_j(t_k), Q_j - \max [q_j(t_{k-1}) - C_j(t_k - t_{k-1}), 0]\} \quad (7)$$

where with Q_j we denote the buffer size of the j-th PX-M converter, and C_j is the capacity of the j-th PX-M converter. Each conforming application request has to be answered, so we assume that the capacity is measured as requests per second (the capacity for responses is the same). The first component in the right hand side of the above expression describes the leftover traffic at time t_k after the PX-m converter has sent the converted request to the network during the period $[t_{k-1}, t_k)$. The second component represents the

traffic accepted from all load balancers during the same period. The traffic accepted by the *j*-th PX-M converter is given by the smaller of the available (empty) space in the buffer and the sum of the conforming traffic.

It is expected that the *j*-th PX-M converter may not be able to convert all forwarded application requests because of its buffer size limitation Q_j and capacity limitation C_j. Thus the traffic lost at the *j*-th PX-M converter is given by:

$L_j(t_k) = F_j(t_k) - \min\{F_j(t_k), Q_j - \max[q_j(t_{k-1}) - C_j(t_k - t_{k-1}), 0]\}$ (8)

From this equation we may define Parlay X gateway utilization as:

$$\eta = \frac{1}{C(t_K - t_0)} \sum_{k}^{K} \left(\sum_{i=1}^{n} V_i(t_k) - \left(\sum_{i=1}^{n} R_i(t_k) + \sum_{j=1}^{m} L_j(t_k) \right) \right) \quad (9)$$

Another measure of Parlay X gateway efficiency is the average throughput which is intimately related to the gateway utilization as defined above.

$\Delta = C\eta = \frac{1}{(t_K - t_0)} \sum_{k=0}^{K} \Delta(t_k)$, where $\Delta(t_k)$ denotes the throughput during the period $[t_{k-1}, t_k)$.

V. Objective function and load control

The objective function for the network provider is given by:

$$J(u) = \sum_{j=1}^{m} \sum_{k=0}^{K} \alpha_j(t_k) L_j(t_k) + \sum_{i=1}^{n} \sum_{k=0}^{K} \beta_i(t_k) R_i(t_k) + \sum_{j=1}^{m} \sum_{k=0}^{K} \gamma_j(t_k) q_j(t_k) \quad (10)$$

where $u = (u_1, u_2, \ldots, u_n)'$ is the control vector that appears in the dynamic model of the TBs.

The parameters $\{\alpha_j, \beta_i, \gamma_j, i = 1...n, j = 1...m\}$ are nonnegative functions of time assigning relative weights given to various losses. The first parameter imposes a penalty on losses in the *j*-th converter, the second parameter imposes a penalty on lost traffic at the admission control for the *i*-th TB and the third parameter is an approximate measure of waiting time or delay at the *j*-th PX-M converter before being served. The problem is to find a control policy that minimizes this function.

We denote with $W(t_k)$ the resources that have to be distributed between SPs in time t_k as a sum of requests that will be processed by all converters and the available places in the converter buffers:

$$W(t_k) = \sum_{j=1}^{m} (\tau . C_j + (Q_j - q_j(t_k))) \quad (11)$$

The control policy for tokens granted to different SPs considers their current demands and it is defined by:

$$u_i(t_k) = ug_i(t_k) + \frac{V_i(t_k)}{\sum_{i=1}^{n} V_i(t_k)} . (W(t_k) - \mu_i(t_k)) \quad (12)$$

The left side of the equation (12) presents the tokens that will be granted to *i*-th SP for the interval $[t_k, t_{k+1})$. The first component in the right side is the number of token that the network operator has been guaranteed to provide to the *i*-th SP, and the second component represents the part of all available resources that are distributed proportionally to the demands of the SPs.

VI. Simulation parameters

To demonstrate the usefulness and effectiveness of the model, we implement the model and the adaptive TB control algorithm in Java. We have used Java IDE Eclipse.

The simulation duration is 600 *s*. During simulation, statistic data are gathered at intervals of 1 *s*.

In the simulation 8 SPs and 4 PX-M converters are used. The capacity of the Parlay X gateway is 800 requests per second which is equally distributed between the converters.

We assume that 2 of the SP generate priority traffic with higher peak and guaranteed rates in comparison with the rest 6 SPs. The peak rate for the high priority traffic is 130 requests per second and the guaranteed rate is 100 requests per second. The peak rate for the low priority traffic is 100 requests per second and the guaranteed rate is 75 requests per second. The behavior of each SP is modeled by Markov Modulated Poisson Process (MMPP) as the arrival process in the context of web services [5]. New application requests are generated according to four-state MMPP, where mean rates for the higher priority SPs are 0, 50, 90, 130 requests per second and for the low priority SPs the mean rates are 0, 30, 60, 100. Changes between different states are uniformly distributed and occurred according to Poisson process with mean 4 seconds. The SP traffic is policed by 8 TBs whose conforming outputs are multiplexed between 4 PX-M converters. The token rate is equal to the guaranteed rate and the bucket size is 30 for high priority SP and 25 for low priority SP. Initially $\mu_i(t_0) = T = 30(25)$, $q_i(t_0) = 0$.

The observable intervals $[t_k - t_{k-1})$ are equal (100 *ms*). The processing time for a single request/response in a converter is 5 *ms*. Given the constraint of 100 ms between given request and its response, the maximum waiting time in the buffer in each direction is *n*.5 = 45*ms* where with *n* we denote the buffer size (9 requests/responses). We assume that the networks delay is constant (1 *ms*) in order to correlate the time between request and its response spent in the gateway.

When choosing values for the relative weights given to various losses, we stress on losses in the gateway where the conformed traffic is lost due to the converter overload. We choose $\alpha_j(t_k) = 10, \beta_i(t_k) - 1, \gamma_j(t_k) = 1, i = 1...n, j = 1...m$ for all t_k.

VII. Numerical results

We define evaluation function *Z* which represents the ratio of the gateway utilization in case of constant rate of token accumulation (without control) to the throughput in case of adaptive rate of token accumulation (with control). The evaluation function *Z* is presented by $Z = \frac{\eta_{static}}{1 + \eta_{dynamic}}$.

Fig.3 shows the changes of the evaluation function *Z* in time. The gateway utilization in case of adaptive control is higher than the utilization without control. The average throughput gain is 8.378827%.

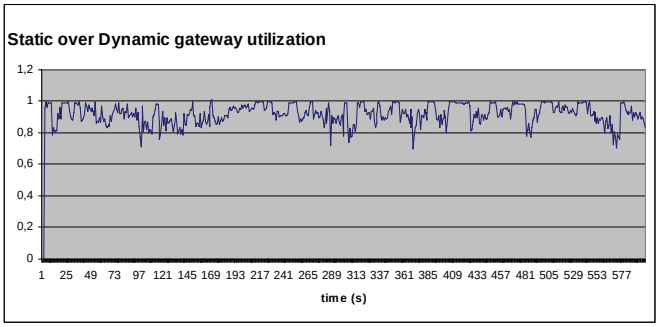


Fig.3. Ratio of gateway throughput for constant and adaptive token rates

To assess the losses, we define evaluation function *Y* which represents the ratio of the losses in case of static token rate to the losses in case of dynamic token rate. The evaluation function *Y* is presented by $Y = \frac{J_{static}}{1 + J_{dynamic}}$.

Fig.4 shows the changes of the evaluation function *Y* in time. The losses in case of adaptive control are less than the losses without control. The average ratio is 5.42972.

Fig.5 shows the ratio of measured delays (the time between request and its response) in case of static and dynamic rate of tokens. As it can be seen, the average delay in case of no control of token rate is less than the average delay in case of adaptive control. This is because of the control strategy which accepts more requests, which means that the buffers before the converters are fuller in comparison with the static token rate. The average delay without control is 49 *ms*, and with applying control is 49,5 *ms*. The measured maximum delay is less than 100 *ms*, as has to be expected due to the limited buffer size.

VIII. CONCLUSION

The paper presents a model for evaluation of the traffic load of a Parlay X gateway with distributed architecture. The numerical results can be used for setting values of quality of service parameters agreed with network operator. The suggested algorithm for adaptive control improves the gateway performance.

The proposed model is suitable for transaction traffic pattern that means that the requests do not correlate each other. Our future work will consider the session traffic pattern where several requests are related to a session.

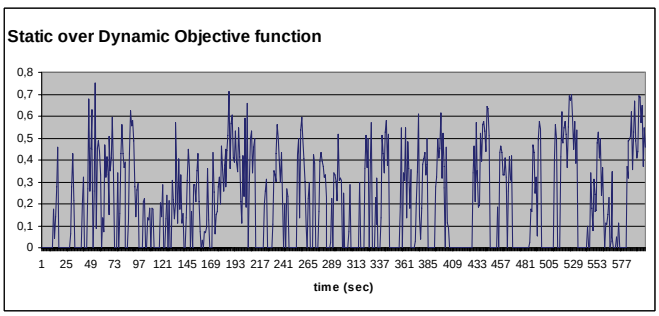


Fig.4. Ratio of losses for constant and adaptive token rates

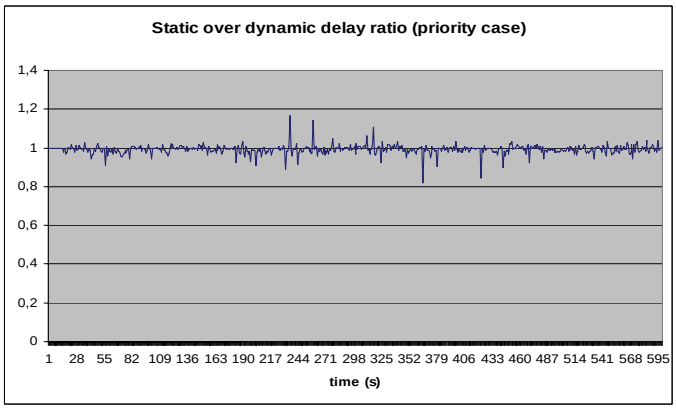


Fig.5 Ration between delays without and with control of the token rates

ACKNOWLEDGEMENT

The work is conducted under the grant of the Project DO 02-135/2008 funded by Bulgarian Ministry of Education and Science.

REFERENCES

[1] Ahmed, N., Wahg, Q., Barbosa, O., Systems Approach to Modeling the Token Bucket Algorithm in Computer Networks, Mathematical problems in Engineering, 2002, vol. 8(3), pp.265-279

[2] Anderson, J., Kihl, M., Sobirk D., Overload Control of a Parlay X Application Server, Proc. of SPECTS'04, pp.821-828, 2004.

[3] Bartolini, N., Bongiovanni, G., Silvestri S., Self- through self-learning: Overload control for distributed web systems, Computer Networks, 53, 2009, pp.727–743

[4] Mathur, V., Dhopeshwarkar , S., Apte, V., MASTH Proxy: An Extensible Platform for Web Overload Control, http://www2009.org/proceedings/pdf/p1113.pdf, 2009

[5] Muscariello, M., Mellia, M., Meo, M., Marsan, M., Lo Cigno, R., Markov models of internet traffic and a new hierarchical MMPP model, Computer Communications vol. 28 , issue 16, 2005, pp. 1835-1851

Outage Probability of Dual SC Over Correlated Rician Fading Channels in the Presence of Multiple CCIs

Dragan Lj. Draca[1], Aleksandra S. Panajotovic[2], Mihajlo C. Stefanovic[3], Nikola M. Sekulovic[4] and Zorica B. Nikolic[5]

***Abstract* – Outage probability is important measure to control the cochannel interference (CCI) level helping the designers of wireless communications systems to readjust the systems operating parameters. The analysis in this paper considers dual selection combining (SC) diversity schemes and it is not limited to a single CCI, but rather assumes the presence of multiple CCIs. Namely, analytical expression for the outage probability for the case of dual signal-to-interference ratio (SIR)-based SC receiver over correlated Rician fading channels experiencing an arbitrary number of Rayleigh CCIs is derived. The various performance evaluation results are graphically presented to show the effect of fading severity, level of correlation and number of CCIs to the system's outage performance.**

***Keywords* – SC diversity system, cochannel interference, Rician fading, Rayleigh fading, outage probability.**

I. INTRODUCTION

Transmission in wireless communications systems is subjected to both fading and cochannel interference (CCI). Fading is the result of multipath propagation and CCI is the result of frequency reuse which is essential in increasing system capacity. Diversity combining [1] is one of the most practical, effective and widely employed technique in digital communications receivers for mitigating the effects of fading and CCI in order to improve the offered quality-of-service (QoS) and to increase systems capacity.

Space diversity techniques combine input signals from multiple receive antennas [2]. Over the years, a variety of diversity techniques, such as maximal ratio combining (MRC), equal gain combining (EGC) or selection combining (SC) have evolved. MRC and EGC systems require all or some of the channel state information (fading amplitude, phase and delay) from all received signals. In addition, a separate receiver chain is needed for each diversity branch which increases their complexity. In opposition to MRC and EGC, SC receiver is much simpler for practical realization since it processes only one of the diversity branches. Specifically, in its conventional form, the SC combiner chooses the branch with the highest signal-to-noise ratio (SNR). Efficient cellular system designs are interference-limited [2], i.e. the level of CCI is sufficiently high as compared to the level of thermal noise, so the thermal noise effect may be ignored. In that case, SC receiver selects the branch with the highest signal-to-interference ratio (SIR; SIR-based selection diversity).

There are different models describing the statistical behavior of the multipath fading envelopes depending on the nature of the radio propagation environment. The most frequently used models are Nakagami, Rayleigh, Rician and Weibull. In a microcellular environment, CCIs usually experience significantly deeper fading than the desired signal. Therefore, different fading models are needed to characterize the desired and undesired signals [3]. The desired signal usually experiences Rician fading since a line-of-sight (LoS) path between transmitter and receiver within a microcell may exist [4]. In a microcell environment, an undesired signal from a distant cochannel cell may well be modeled by Rayleigh statistics [5]. Most of the published papers study diversity by considering the effect of only the strongest CCI [6-10] assuming that the remaining CCIs can be ignored. In general, the performance analysis of cellular mobile radio systems should include the influence of multiple CCIs from two or more distant cells [3-5, 11-12].

The performance of dual SC diversity receiver operating over correlated Rician fading channels in the presence of arbitrary number of correlated Rayleigh distributed CCIs is studied in this paper. Actually, the outage probability, as important performance metric, is obtained. The derivation of analytical expressions for the probability density function (PDF) and cumulative distribution function (CDF) of the output SIR has been preceded to assess the influence of various parameters such as fading severity, level of correlation and number of CCIs on SC system performance.

II. SYSTEM AND CHANNEL MODEL

We consider a cellular mobile radio system with dual SIR-based SC receiver. The desired signal envelopes on two diversity branches are assumed to follow the correlated Rician distribution, due to insufficient antennas distance, whose PDF is given by [13]

[1]Dragan Lj. Draca is with the Faculty of Electronic Engineering, University of Nis, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: dragan.draca@elfak.ni.ac.rs

[2]Aleksandra S. Panajotovic is with the Faculty of Electronic Engineering, University of Nis, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: aleksandra.panajotovic@elfak.ni.ac.rs

[3]Mihajlo C. Stefanovic is with the Faculty of Electronic Engineering, University of Nis, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: mihajlo.stefanovic@elfak.ni.ac.rs

[4]Nikola M. Sekulovic is with the Faculty of Electronic Engineering, University of Nis, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: sekulani@bankerinter.net

[5]Zorica B. Nikolic is with the Faculty of Electronic Engineering, University of Nis, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: zorica.nikolic@elfak.ni.ac.rs

$$p_{A_1A_2}(A_1,A_2)=\frac{A_1A_2(K+1)^2}{\beta^2(1-r^2)}\exp\left(-\frac{\left(A_1^2+A_2^2\right)(K+1)+4K\beta(1-r)}{2\beta(1-r^2)}\right)$$
$$\times\sum_{i=0}^{\infty}\varepsilon_i I_i\left(\frac{A_1A_2r(K+1)}{\beta(1-r^2)}\right)I_i\left(\frac{A_1}{1+r}\sqrt{\frac{2K(K+1)}{\beta}}\right)I_i\left(\frac{A_2}{1+r}\sqrt{\frac{2K(K+1)}{\beta}}\right), \quad (1)$$
$$\varepsilon_i=\begin{cases}1, & i=0\\ 2, & i\neq 0\end{cases},$$

where r is the branch correlation coefficient, K is the Rician factor defined as the ratio of the signal power in the dominant component over the scattered power, $2\beta=\overline{A_1^2}=\overline{A_2^2}$ is the average desired signal power and $I_i(\cdot)$ is the modified Bessel function of the first kind and ith order.

The desired signal is corrupted by arbitrary number (n) of Rayleigh distributed CCIs having the same average power, $\Omega=\overline{r_j^2}$, $j=1,2,...,n$. In practice, assumption that CCIs have the same average power is not unreasonable since they originated from approximately the same distance from the receiver [12]. The power sum of CCIs on lth diversity branch is $R_l^2=r_{l,1}^2+r_{l,2}^2+\mathrm{K}+r_{l,n}^2$, $l=1,2$. In that case, the joint PDF of R_1 and R_2 is [14]

$$p_{R_1R_2}(R_1,R_2)=\frac{4(R_1R_2)^n}{\Gamma(n)(1-r^2)r^{n-1}\Omega^{n+1}}\exp\left(-\frac{R_1^2+R_2^2}{\Omega(1-r^2)}\right)I_{n-1}\left(\frac{2rR_1R_2}{\Omega(1-r^2)}\right). \quad (2)$$

Let $\mu_1=A_1^2/R_1^2$ and $\mu_2=A_2^2/R_2^2$ be the instantaneous SIRs on the two diversity branches of SC receiver. The joint PDF of these random variables can be obtained as

$$p_{\mu_1\mu_2}(\mu_1,\mu_2)=\frac{1}{4\sqrt{\mu_1\mu_2}}\int_0^{\infty}\int_0^{\infty}R_1R_2\,p_{A_1A_2}\left(R_1\sqrt{\mu_1},R_2\sqrt{\mu_2}\right)p_{R_1R_2}(R_1,R_2)\,dR_1dR_2. \quad (3)$$

Substituting (1) and (2) in (3) and using the infinite-series representation of the modified Bessel function [15]

$$I_i(x)=\sum_{n=0}^{\infty}\frac{x^{2n+i}}{2^{2n+i}n!\Gamma(n+i+1)} \quad (4)$$

and [16, Eq. (3.326/2)], $p_{\mu_1\mu_2}(\mu_1,\mu_2)$ can be expressed as

$$p_{\mu_1\mu_2}(\mu_1,\mu_2)=\frac{\exp\left(-\frac{2K}{1+r}\right)}{\Gamma(n)}\sum_{i,j,l,s,p=0}^{\infty}\varepsilon_i\frac{r^{i+2(j+p)}\left(1-r^2\right)^{i+l+s+n+1}}{2^{2(i+j+1)+l+s}(1+r)^{2(i+l+s)}}$$
$$\times\frac{K^{i+l+s}(K+1)^{2(i+j+1)+l+s}\Omega^{2(i+j+1)+l+s}}{\beta^{2(i+j+1)+l+s}} \quad (5)$$
$$\times\frac{\Gamma(i+j+l+n+p+1)\Gamma(i+j+s+n+p+1)}{j!l!s!p!\Gamma(i+j+1)\Gamma(i+l+1)\Gamma(i+s+1)\Gamma(n+p)}$$
$$\times\frac{\mu_1^{i+j+l}\mu_2^{i+j+s}}{\left(\frac{(K+1)\mu_1}{\gamma}+1\right)^{i+j+l+n+p+1}\left(\frac{(K+1)\mu_2}{\gamma}+1\right)^{i+j+s+n+p+1}},$$

where γ is the average SIR defined as $\gamma=2\beta/\Omega$.

The joint bivariate CDF of μ_1 and μ_2 can be derived as

$$F_{\mu_1\mu_2}(\mu_1,\mu_2)=\int_0^{\mu_1}\int_0^{\mu_2}p_{\mu_1\mu_2}(x_1,x_2)\,dx_1dx_2. \quad (6)$$

which, by substituting (5) and using [16, Eq. (3.194/1)], yields

$$F_{\mu_1\mu_2}(\mu_1,\mu_2)=\frac{\exp\left(-\frac{2K}{1+r}\right)}{\Gamma(n)}\sum_{i,j,l,s,p=0}^{\infty}\varepsilon_i\frac{r^{i+2(j+p)}\left(1-r^2\right)^{i+l+s+n+1}}{(1+r)^{2(i+l+s)}\gamma^{2(i+j+1)+l+s}}$$
$$\times\frac{K^{i+l+s}(K+1)^{2(i+j+1)+l+s}\mu_1^{i+j+l+1}\mu_2^{i+j+s+1}}{j!l!s!p!}$$
$$\times\frac{\Gamma(i+j+l+n+p+1)\Gamma(i+j+s+n+p+1)}{\Gamma(n+p)\Gamma(i+j+1)\Gamma(i+l+1)\Gamma(i+s+1)(i+j+l+1)(i+j+s+1)} \quad (7)$$
$$\times{}_2F_1\left(i+j+l+n+p+1,\,i+j+l+1;\,i+j+l+2;\,-\frac{(K+1)\mu_1}{\gamma}\right)$$
$$\times{}_2F_1\left(i+j+s+n+p+1,\,i+j+s+1;\,i+j+s+2;\,-\frac{(K+1)\mu_2}{\gamma}\right),$$

where ${}_2F_1(a,b;c;d)$ is the Gaussian hypergeometric function.

The SC receiver chooses and outputs the branch with the highest SIR, i.e. $\mu_{SC}=\max\{\mu_1,\mu_2\}$. The CDF of μ_{SC} can be derived from (7) by equating arguments.

$$F_{\mu_{SC}}(\mu_{SC})=F_{\mu_1\mu_2}(\mu_{SC},\mu_{SC}). \quad (8)$$

III. Outage Probability

The outage probability is important, reliable and widely accepted performance measure for diversity systems operating in fading environments. Therefore, this performance criterion has been thoroughly studied by many researchers. In interference-limited environments, the outage probability is defined as the probability that the output SIR falls below a given outage threshold μ_{th}, also known as a protection ratio, which depends on expected QoS.

$$P_{out}=P_R(\mu<\mu_{th})=\int_0^{\mu_{th}}p_{\mu_{SC}}(\mu)d\mu=F_{\mu_{SC}}(\mu_{th}). \quad (9)$$

A system planner can use the outage probability calculations to assess the effects of various systems and channels parameters on the QoS provided by the system.

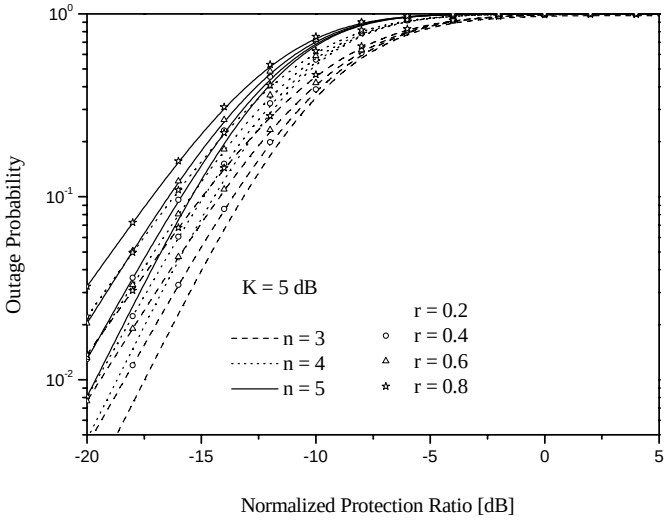
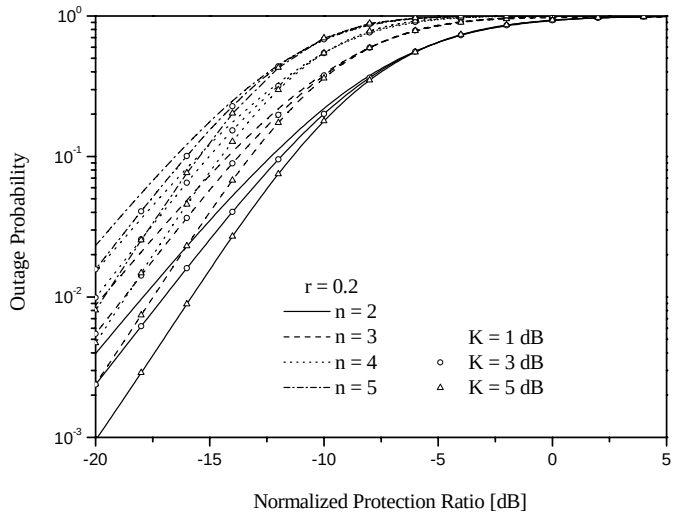
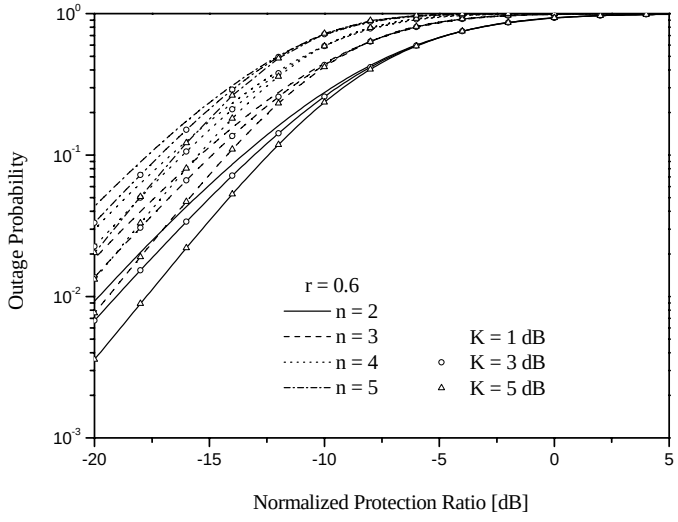


Fig. 1. Outage probability versus the normalized protection ratio for a dual SC receiver for several values of correlation coefficient in the presence of different number of CCIs



a)



b)

Fig. 2. Outage probability versus the normalized protection ratio for a dual SC receiver for several values of Rician factor in the presence of different number of CCIs a) r=0.2 b) r=0.6

IV. NUMERICAL RESULTS

Using the previous mathematical analysis, the outage probability as a function of the normalized protection ratio μ_{th}/γ for several values of correlation coefficient in the presence of different number of CCIs is plotted in Fig. 1. The obtained results show clearly that outage performance degrades with increase of branch correlation (decrease spatial separation between the received antennas in terminals) and/or number of CCIs. Fig. 2 depicts that the outage probability improves when the Rician factor increases. This happens because a higher value of the Rician factor means that the desired signal contains a large LoS component and a small diffuse-scattered component, i.e. the desired signal suffers less from severe fading. Comparison of Figs. 2*a* and Fig. 2*b* shows greater influence of interferences number *n* on the outage probability in the case of lower correlation coefficient. As an illustration, if the relative deterioration due to increase of number of CCIs is defined as

$$\Delta_{n,n+1} = \left[1 - \frac{P_{out\ n}}{P_{out\ n+1}}\right] \cdot 100\%,$$

for K = 5 dB and normalized protection ratio of -15 dB, $\Delta_{2,3}$ = 62.1% for r = 0.2 and $\Delta_{2,3}$ =53.27% for r = 0.6.

V. CONCLUSION

In cellular mobile radio systems, the main factors limiting performance are CCIs from neighboring cells and multipath fading. Dual SC receiver operating over correlated Rician fading channels in the presence of arbitrary number of correlated Rayleigh distributed CCIs has been studied in this paper. The outage probability has been derived in form of infinite-series representation. The various graphically presented evaluation results demonstrate that system performance improves when Rician factor increases (fading severity decreases) and/or correlation coefficient decreases. Moreover, the system shows better outage performance for smaller number of CCIs. A system planner can efficiently use the presented expression for the outage probability to simulate different system and channel conditions and to readjust system parameters to provide required QoS.

ACKNOWLEDGEMENT

This work was supported in part by the Ministry of Science of Serbia within the project "Development and realization of new generation software, hardware and services based on software radio for specific purpose applications" (TR-11030).

REFERENCES

[1] M. K. Simon and M.-S. Alouini, *Digital Communication Over Fading Channels*, 1st ed. New York: Wiley, 2000.

[2] A. Goldsmith, *Wireless communications*, Cambridge University Pres: New York, 2005.

[3] Y.-D. Yao and A. U. H. Sheikh, "Investigations into Cochannel Interference in Microcellular Mobile radio Systems", IEEE Trans. Veh. Technol., vol. 41, no. 2, pp. 114-123, 1992.

[4] R. J. C. Bultitude and G. K. Bedal, "Propagation characteristics on microcellular urban mobile radio channels at 910 MHz", IEEE J. Select. Areas Commun., vol. 7, no. 1, pp. 31-39, 1989.

[5] Y.-D. Yao and A. U. H. Sheikh, "Outage probability analysis for microcell mobile radio systems with cochannel interferers in Rician/Rayleigh fading environment", Electron. Lett., vol. 26, no. 13, pp. 864-866, 1990.

[6] S. Okui, "Effects of SIR selection diversity with two correlated branches in the m-fading channel", IEEE Trans. Commun., vol. 48, no. 10, pp. 1631-1633, 2000.

[7] G. K. Karagiannidis, "Performance analysis of SIR-based dual selection diversity over correlated Nakagami-m fading channels", IEEE Trans. Veh. Technol., vol. 52, no. 5, pp. 1207-1216, 2003.

[8] M. Č. Stefanović, D. M. Milović, A. M. Mitić and M. M. Jakovljević, "Performance analysis of system with selection combining over correlated Weibull fading channels in the

presence of cochannel interference", Int. J. AEÜ, vol. 62, no. 9, pp. 695-700, 2008.

[9] A. S. Panajotović, M. Č. Stefanović and D. Lj. Drača, "Performance analysis of system with selection combining over correlated Rician fading channels in the presence of cochannel interference", Int. J. AEÜ, vol. 63, no. 12, pp. 1061-1066, 2009.

[10] M. Č. Stefanović, D. Lj. Drača, A. S. Panajotović and N. M. Sekulović, "Performance analysis of system with L-branch selection combining over correlated Weibull fading channels in the presence of cochannel interference", Int. J. Commun. Syst., vol. 23, no. 2, pp. 139-150, 2010.

[11] K. W. Sowerby and A. G. Williamson, "Outage probability calculations for a mobile radio system having multiple Rayleigh interferers", Electron. Lett., vol. 23, no. 11, pp. 600-601, 1987.

[12] A. A. Abu-Dayya and N. C. Beaulieu, "Outage probabilities of Cellular Mobile Radio Systems with Multiple Nakagami-m Interferers", IEEE Trans. Veh. Technol., vol. 40, no. 4, pp. 757-768, 1991.

[13] M. K. Simon, "Comments on "Infinite-series representations associated with the bivariate Rician distribution and their applications", IEEE Trans. Commun., vol. 54, no. 8, pp. 1511-1512, 2006.

[14] M. Nakagami, "The m-distribution - A general formula of intensity distribution of rapid fading", in *Statistical Methods in Radio Wave Propagation*, Pergamon Press, Oxford, U.K., pp. 3-36, 1960.

[15] M. Abramovitz and I. A. Stegun, *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, New York: Dover publications Inc., 1970.

[16] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 5th ed. New York: Academic, 1994.

Effects of Carrier Phase Error on QPSK Receiver over Nakagami-*m* Fading/Gamma Shadowing

Goran T. Djordjevic[1], Dejan N. Milic[2], Bojana Z. Nikolic[3] and Aleksandra M. Cvetkovic[4]

***Abstract* – In this paper, taking the imperfect reference signal extraction into account we analyze the reception of quaternary phase-shift keying (QPSK) signals over the composite fading channel. The composite fading channel is described by Nakagami-*m* probability density function (PDF) whose average power is also a random process with gamma distribution. The effects of multipath fading severity, shadowing and imperfect reference signal recovery on bit error rate (BER) performance are studied.**

***Keywords* – Bit error rate, Fading, Shadowing.**

I. INTRODUCTION

In wireless communication channels, short-term (multipath) fading caused by combination of delayed, reflected scattered and diffracted signal components, and long-term fading (shadowing) due to local topography in receiver surrounding occur simultaneously leading to the composite fading [1]-[3]. Composite fading is often modeled by Nakagami-*m* probability density function (PDF) whose mean power is random process with lognormal distribution. However, a composite PDF, obtained in this way, is in integral form and it is not convenient for further analysis and mobile radio system design. For this reason, equivalent gamma distribution is more often used for describing slow fading. In [2], [3], the relations between lognormal and equivalent gamma distribution was determined. Consequently, random fast fluctuations of the signal envelope in composite fading channel can be described by Nakagami-*m* PDF while the average power is also a random process with gamma PDF.

In previous papers, concerning coherent signal detection in composite fading channel [1], [2], it was strictly assumed that reference signal carrier is perfect i.e. the phase of the reference signal carrier is perfectly extracted. It is, however, clear that results, obtained in this way, are idealized and they should be taken as the most optimistic in given circumstances. In practical realizations of the receivers, due to the presence of a noise and fading in a channel, the extraction of the reference signal carrier is imperfect i.e. there is a certain difference between the phase of the incoming signal and the phase of the extracted carrier.

The influence of the imperfect reference signal carrier extraction has been considered in previous papers only in multipath fading channels, without taking into account a shadowing effect [4], [5]. In [4] the error probability has been determined in the case of digital binary and quaternary phase-shift keying (BPSK and QPSK) signal detection in the presence of Gaussian noise, Nakagami-*m* fading and imperfect reference signal carrier extraction. It was assumed that phase error has uniform distribution. In [5] the derivation procedure of the general explicit analytical expression for average symbol error probability for MPSK signal detection in the presence of Nakagami-*m* fading, Gaussian noise and imperfect phase reference has been presented. It was assumed that reference signal recovery is performed by phase-locked loop (PLL) from unmodulated pilot signal.

The contribution of this paper is in the estimation of the bit error rate (BER) in detection of QPSK signal, propagating over the composite fading channel. A reference signal carrier extraction is performed by PLL from unmodulated signal and is assumed imperfect. A certain phase error exists and represents the difference between the phase of the incoming signal and the phase of the recovered reference signal carrier. This phase error is a random process whose instantaneous values have Tikhonov PDF [5]. The BER performance of QPSK receiver are obtained and discussed.

II. CHANNEL AND RECEIVER MODEL

A. Channel model

The PDF of the envelope *r* in the Nakagami-*m* channel is

$$p_r(r)=\frac{2}{\Gamma(m_m)}\left(\frac{m_m}{\Omega_m}\right)^{m_m} r^{2m_m-1}\exp\left(-\frac{m_m}{\Omega_m}r^2\right) \qquad (1)$$

where $\Gamma(.)$ is the Euler Gamma function [6], m_m is the severity parameter of multipath fading and Ω_m is the mean envelope power $\Omega=E\{r^2\}$, where $E\{.\}$ denotes the expectation operator. The more the value of m_m, the less the fading severity is. When there is no shadowing, the average power

[1]Goran T. Djordjevic is with Faculty of Electronic Engineering, Department of Telecommunications, Aleksandra Medvedava 14, 18000 Nis, Serbia, E-mail: goran@elfak.ni.ac.rs

[2]Dejan N. Milic is with Faculty of Electronic Engineering, Department of Telecommunications, Aleksandra Medvedava 14, 18000 Nis, Serbia, E-mail: dejan.milic@elfak.ni.ac.rs

[3]Bojana Z. Nikolic is with Faculty of Electronic Engineering, Department of Telecommunications, Aleksandra Medvedava 14, 18000 Nis, Serbia, E-mail: bojana.nikolic@elfak.ni.ac.rs

[4]Aleksandra M. Cvetkovic is with Faculty of Electronic Engineering, Department of Telecommunications, Aleksandra Medvedava 14, 18000 Nis, Serbia, E-mail: caleksandra@gmail.com

Ω_m is constant, but when shadowing is present, Ω_m is the stochastic variable, and consequently eq. (1) can be rewritten in the form

$$p_{r/\Omega_m}(r/\Omega_m)=\frac{2}{\Gamma(m_m)}\left(\frac{m_m}{\Omega_m}\right)^{m_m} r^{2m_m-1}\exp\left(-\frac{m_m}{\Omega_m}r^2\right), \quad (2)$$

where Ω_m has lognormal distribution

$$p_{\Omega_m}(\Omega_m)=\frac{1}{\sqrt{2\pi\sigma_{sh}{}^2}\Omega_m}\exp\left(-\frac{(\ln\Omega_m-\mu)^2}{2\sigma_{sh}{}^2}\right), \quad (3)$$

where $\mu=\ln P_r$ and σ are mean value and standard deviation of $\ln\Omega_m$. In this formula, μ and σ are expressed in nepers. The average received signal power is denoted by P_r. By σ_{sh} we denote the shadow standard deviation. Standard deviation expressed in decibels (shadowing spread) is given by $\sigma_{shdB}=8.686\sigma_{sh}$.

The PDF of envelope when multipath fading and shadowing are simultaneously present is given by

$$p_r(r)=\int_0^\infty p_{r/\Omega}(r/\Omega)p_\Omega(\Omega)d\Omega \quad (4)$$

that is not a closed form solution and is inconvenient for further analysis. Therefore, the lognormal distribution is approximated by equivalent gamma PDF

$$p_{\Omega_m}(\Omega_m)=\frac{1}{\Gamma(m_s)}\left(\frac{m_s}{\Omega_s}\right)^{m_s}\Omega_m{}^{m_s-1}\exp\left(-\frac{m_s}{\Omega_s}\Omega_m\right). \quad (5)$$

By replacing (2) and (5) in (4), the PDF of the composite envelope of the Nakagami-*m*/gamma shadowing signal is derived as

$$p_r(r)=\frac{4}{\Gamma(m_m)\Gamma(m_s)}\left(\frac{m_m m_s}{\Omega_s}\right)^{(m_m+m_s)/2}$$

$$\times r^{m_m+m_s-1}K_{m_s-m_m}\left(2r\sqrt{\frac{m_m m_s}{\Omega_s}}\right), \quad (6)$$

where $K_\nu(.)$ is the modified Bessel function of the second kind and order ν [6].

The relations among parameters of original lognormal distribution (3) and equivalent gamma distribution (5) can be established by using two different approaches. In [3] by using moment matching method the relations among parameters of lognormal and gamma PDF are

$$m_s=\frac{1}{\exp(\sigma_{sh}{}^2)-1}, \quad \Omega_s=P_r\sqrt{\frac{m_s+1}{m_s}}. \quad (7)$$

i.e.

$$\sigma_{sh}=\sqrt{\ln\frac{m_s+1}{m_s}}, \quad \mu=\ln\left(\Omega_s\sqrt{\frac{m_s}{m_s+1}}\right). \quad (8)$$

In [2], the parameters of lognormal and gamma PDF are related through

$$\sigma_{sh}=\frac{10}{\ln 10}\sqrt{\psi'(m_s)}\ \text{(dB)}, \quad \mu=\frac{10}{\ln 10}\left(\ln\frac{\Omega_s}{m_s}+\psi(m_s)\right)\ \text{(dB)}, \quad (9)$$

where $\psi(.)$ is the digamma function and $\psi'(.)$ is the trigamma function defined as [6]

$$\psi(z)=\frac{d}{dz}\ln\Gamma(z)=\frac{\Gamma'(z)}{\Gamma(z)}, \quad \psi'(z)=\frac{d^2}{dz^2}\ln\Gamma(z). \quad (10)$$

The illustration of signal propagation between transmitter and receiver in the composite fading environment is given in Fig. 1. Three statistically phenomena should be noted: deterministic path loss, slow shadowing and fast multipath fading.

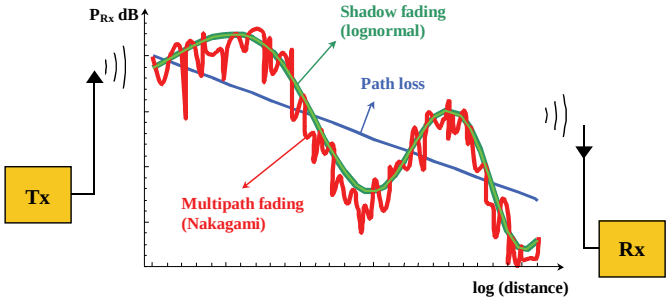


Fig. 1. Illustration of composite fading channel

B. Receiver model

The input signal of the receiver can be written in the form

$$s(t)=r(t)\cos(\omega_0 t+\gamma(t)+\phi_n)+n(t), \quad (14)$$

where $r(t)$ is a signal envelope, $\gamma(t)$ is a random phase of the received signal, which appeared due to multipath propagation and ϕ_n is a information bearing signal that can take one of the following values: {π/4, 3π/4, -3π/4, -π/4}. Additive white Gaussian noise (AWGN) with zero mean value and variance σ^2 is denoted by $n(t)$. Since the signal propagates over the composite fading channel, the envelope $r(t)$ of that signal is a random process and its instantaneous values have PDF given by (6).

The multipath Nakagami-*m* fading parameter is denoted by m_m. A greater parameter m_m indicates a lower fading severity. The range of this parameter's values is: $0.5\le m_m<\infty$. The shadowing parameter is denoted by m_s. The greater the parameter m_s the smaller the shadowing is. The average signal

power is defined as $\overline{r^2} = \int_0^\infty r^2 p_r(r)dr = \Omega_s$. It can be shown that the PDF of the instantaneous signal-to-noise ratio (SNR) is given by

$$p_\gamma(\gamma) = \frac{2}{\Gamma(m_m)\Gamma(m_s)}\left(\frac{m_m m_s}{\gamma_0}\right)^{(m_m+m_s)/2} \times \gamma^{\frac{m_m+m_s-2}{2}} K_{m_s-m_m}\left(2\sqrt{\frac{m_m m_s}{\gamma_0}\gamma}\right), \quad (15)$$

where γ_0 is an average symbol SNR. The relation between the average symbol and bit SNR is: $\gamma_0 = \gamma_{0b}\log_2 M$, where M denotes the number of phase levels and it is worth $M = 4$ for QPSK. The average bit SNR is represented by γ_{0b}.

The signal $s(t)$ is first filtered by a band pass filter and then led into the upper and lower branch of the receiver. In upper branch the signal is multiplied by the recovered signal from the PLL, $2\cos(\omega_0 t + \hat{\gamma}(t))$, and the resulting signal is then filtered by a low pass filter. The signal in the lower branch is multiplied by the recovered signal from the PLL, phase shifted for the value of $\pi/2$, $-2\sin(\omega_0 t + \hat{\gamma}(t))$, and then filtered by the low pass filter. It should be noticed that the estimated phase is denoted by $\hat{\gamma}(t)$. A difference between the phase of the incoming signal and the estimated phase is $\varphi(t) = \gamma(t) - \hat{\gamma}(t)$. This phase difference is a random process. Based on the signal values in the upper and lower branch of the receiver, a decision which symbol is sent by the transmitter has been made. The signals in the upper and lower branch of the receiver are

$$z_u(t) = r(t)\cos(\phi_n + \varphi(t)) + x(t), \quad (16)$$

$$z_l(t) = r(t)\sin(\phi_n + \varphi(t)) + y(t). \quad (17)$$

where $x(t)$ and $y(t)$ are Gaussian noise components with zero mean value and variance σ^2.

After the analysis of the receiver operating, an expression for the conditional BER can be derived

$$P_{e/\gamma,\varphi}(\gamma,\varphi) = 0.25\left\{erfc\left(\sqrt{\gamma/2}(\cos\varphi - \sin\varphi)\right), + erfc\left(\sqrt{\gamma/2}(\cos\varphi + \sin\varphi)\right)\right\}, \quad (18)$$

where erfc(.) is the complementary error function [6]. This is the BER conditioned on the phase error, φ, and the instantaneous SNR, γ.

When the reference signal carrier phase is extracted from unmodulated pilot signal, the phase error has Tikhinov PDF [5]

$$p_\varphi(\varphi) = \frac{\exp(\gamma_{PLL}\cos\varphi)}{2\pi I_0(\gamma_{PLL})}, \quad -\pi \le \varphi < \pi, \quad (19)$$

where γ_{PLL} is the SNR in PLL circuit and it is related to standard deviation of phase error by

$$\gamma_{PLL} = \frac{1}{\sigma_\varphi^2}. \quad (20)$$

The average BER can be obtained by averaging the conditional BER over the instantaneous SNR, γ, and the phase error, φ,

$$P_e = \int_0^\infty \int_{-\pi}^{\pi} P_{e/\gamma,\varphi}(\gamma,\varphi)p_\gamma(\gamma)p_\varphi(\varphi)d\varphi d\gamma. \quad (21)$$

III. Numerical results

In this Section, we present the numerical results obtained by applying numerical integration in (21) and simulation results obtained by Monte Carlo simulations.

Fig. 3 presents the influence of standard deviation of phase error on BER performance. It should be noticed that stochastic phase error causes appearance of irreducible error floor. When SNR tends to infinity this BER floor remains constant and depends only on phase noise standard deviation.

Fig. 4. presents the influence of multipath fading severity on BER performance of partially coherent QPSK receiver. The BER values decrease with m_m increasing (severity decreasing).

Fig. 5. illustrates the influence of shadowing on BER performance. The shadowing has considerable influence on the BER. In the case of $\sigma_\varphi = 10^o$, for $\gamma_{0b} = 22$ dB, BER is 8.6×10^{-5} when $m_s = 10$ (light shadowing) and 3.6×10^{-3} when $m_s = 1$ (heavy shadowing). It should be also noticed that as m_s increases, the BER performance becomes less sensitive to m_s.

Fig. 6 shows the average SNR values in the channel that are required for achieving BER of 10^{-4}, for different severities of multipath fading and shadowing intensities. For example, if the m_m = 3 and m_s = 5.25, the minimum required value of SNR in the channel is 19 dB.

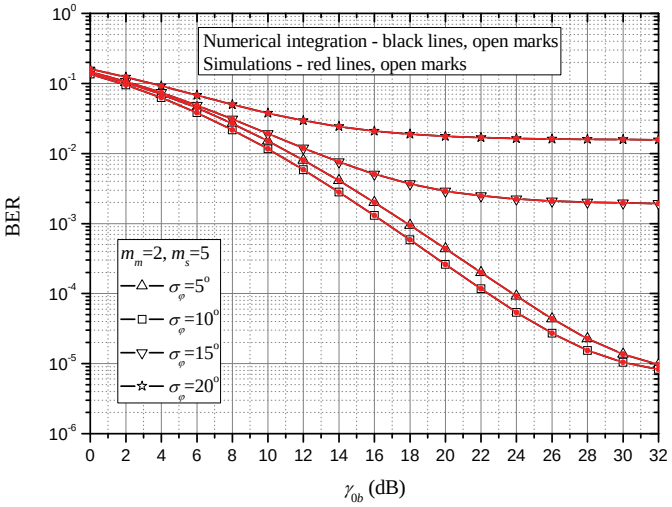


Fig. 1. BER versus SNR for different values of standard deviation of phase noise

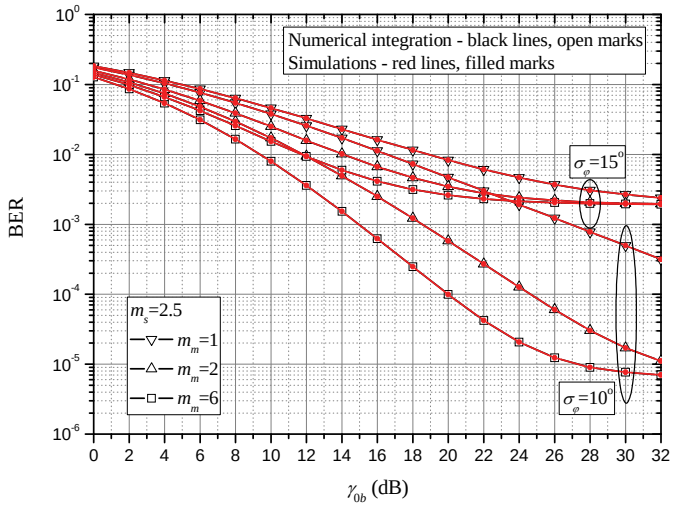


Fig. 4. BER versus SNR for different values of multipath fading severity

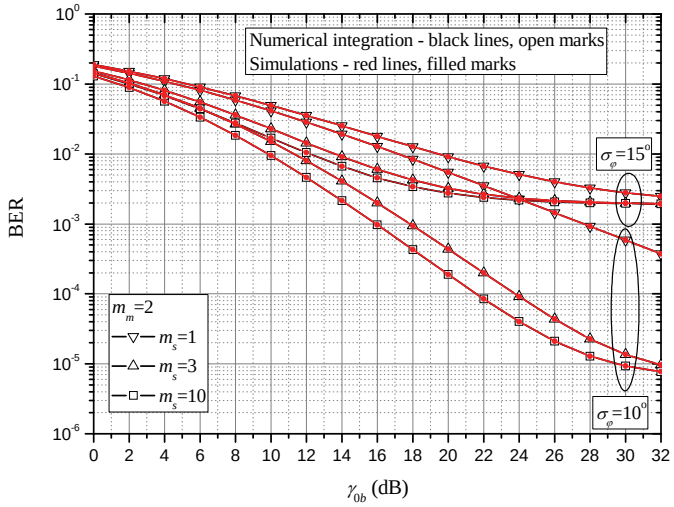


Fig. 5. BER versus SNR for different values of shadowing

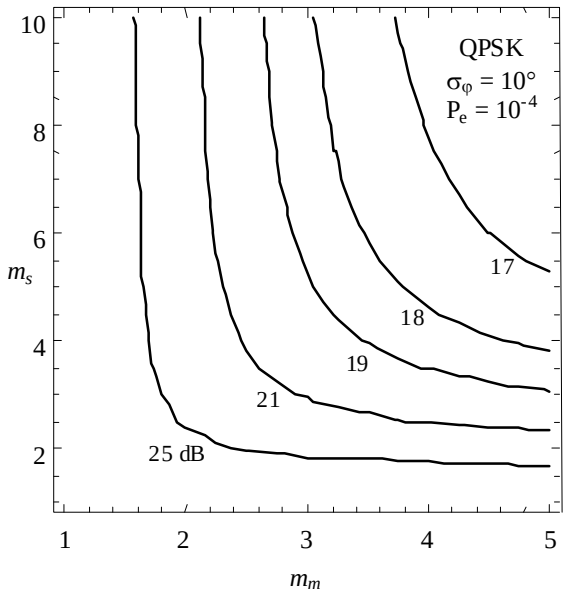


Fig.6. Required values of average SNR in order to achieve BER of 10^{-4}

IV. Conclusion

In this paper, we have presented theoretical basics and simulation model for composite fading channel. We have observed Nakagami-*m*/gamma shadowing channels. Furthermore, the error performance of QPSK receiver operating over composite fading channel has been determined. The BER degradations caused by simultaneous influences of imperfect reference signal extraction, multipath fading severity and shadowing have been determined. There are good match between results obtained by numerical integration and results obtained by Monte Carlo simulations.

Acknowledgement

This work was supported in part by the Ministry of Science of Serbia within the Project "Development and realization of new generation software, hardware and services based on software radio for specific purpose applications" (TR-11030).

References

[1] P. S. Bithas, N. C. Sagias, P. K. Mathiopoulos, G. K. Karagiannidis, A. A. Rontogiannis, "On the performance analysis of digital communications over generalized-K fading channels", IEEE Communications Letters, vol. 10, no. 5, pp. 353-355, May 2006.

[2] P. M. Shankar, "Error rates in generalized shadowed fading channels", Wireless Personal Communications, vol. 28, no. 4, pp. 223-238, Feb. 2004.

[3] I. M. Kostic, "Analytical approach to performance analysis for channel subject to shadowing and fading", IEE Proceedings, vol. 152, no. 6, pp. 821-827, Dec. 2005.

[4] C. M. Lo, W. H. Lam, "Average BER of BPSK and QPSK systems with noisy phase reference over Nakagami-m fading channels", IEICE Transactions on Communications, vol. E84-B, no. 6, pp. 1687-1689, 2001.

[5] I. M. Kostic, "Average SEP for M-ary CPSK with noisy phase reference in Nakagami fading and Gaussian noise", European Transactions on Telecommunications, vol. 18, no. 2, pp. 109-113, 2007.

[6] I. S. Gradsteyn, I. M. Ryzhik, *Table of integrals, series and products*, 7th ed. San Diego, CA: Academic, 1994.

[7] I. M. Kostic, "A method for error rate analysis in the presence of rapid and slow fading", Proc. XLVI ETRAN Conference, vol. II, pp. 62-64, Banja Vrucica – Teslic, June 4-7, 2002.

[8] G. Marsaglia, W. W. Tsang, "A simple method for generating gamma variables", ACM Transactions on Mathematical Software, vol. 26, no. 3, pp. 363-372, September 2000.

[9] D. Kundu, R. D. Gupta, "A convenient way of generating gamma random variables using generalized exponential distribution", Computational Statiscics & Data Analysis, vol. 51, no. 6, pp. 2796-2802, March 2007.

[10] M. C. Jeruchim, P. Balaban, K. Sam Shanmugan, Simulation of Communication Systems – Modeling, Methodology, and Techniques, New York: Kluwer, 2000.

Optimal Configuration of Enterprise Storage Capacity for a Telecom Company

Mario O. Krastev[1], Ludmila H. Raikovska[2] and Kiril I. Dakov[3]

***Abstract* – A research on enterprise storage systems optimal disk configuration is carried out for maximum performance and disaster recovery purposes of online applications in a Telecommunication company. The described solution is based on high performance, security and storage capacity demands needed for a typical telecommunication operator. The results can be used for planning of similar solutions and developing data protection scenarios for any modern telecommunication company.**

***Keywords* – Storage systems, Telecom applications, Data protection, capacity optimization**

I. Introduction

Telecommunication operators use a variety of on-line applications, running on different HA systems in complex heterogonous IT infrastructure. The aim of the current paper is to present a methodology for the optimal enterprise storage array disk space configuration based on telecom applications requirements. The proposed solution follows the best practices and standards for data protection; it could be successfully implemented for business continuity and disaster recovery purposes [1, 2]. The main purpose of the research is to show how the proposed configuration and design lead to higher performance and storage utilization results that could be achieved without any additional hardware or software investments.

II. Methodology for Optimal Configuration of Enterprise Storage Capacity

A. Enterprise storage systems requirements

The most important business critical enterprise applications and their requirements for performance, capacity and protection level, that are typical for any Telecom operator and are most important for its daily work, are presented in Table I. These systems' data should be hosted on enterprise storage arrays. The requirements for the main application databases and their typical capacity needs, number of IOPS and needs for local and remote replication are summarised in Table I.

Table I. Main applications in telecom operator

Appli-cation	Protec-tion	Size	Max IOPS	I/O avg size	Type I/O	Read/ Write	Rep licat ion	B C V
Billing	high	3 TB	9000	32 k	Ran-dom	3:1	yes	3 pairs
CRM	high	1,5 TB	9000	32 k	Ran-dom	3:1	yes	3 pairs
ERP	high	2 TB	6000	32 k	Ran-dom	3:1	yes	3 pairs
Develop ment	normal	10 TB	3000	32 k	Ran-dom	3:1	yes	2T 20%
Develop ment, Test, Data ware-house	normal	30 TB	2000	32 k	Sequ ential	4:1	no	no

B. Choice of drive types

The proper choice of disk drives is important for meeting the requirement of all business critical applications. Correct choice and configuration of drives could lead to significant improvement of response time and performance of the applications. In Table II the basic disk drives characteristics of EMC DMX3 storage array [3] are presented.

Table II. Drives specifications

Specification	73, 146, 300 GB 15k rpm	73, 146, 300 GB 10k rpm	500 GB 7,2k rpm
Internal data rate (MB/s)	68-114	47-94	47-94
Avg rotation latency (ms)	2	3	4,17
Avg read seek (ms)	3,5	4,7	8,5
Avg write seek (ms)	4	5,4	9,5
Read single track seek (ms)	0,27	0,20	0,8
Write single track seek (ms)	0,45	0,50	1
Read full stroke seek (ms)	7,4	9,7	16
Write full stroke seek (ms)	7,9	10,4	16,55

In most cases the main factor for drive selection is capacity and rotation speed but this approach leads to many performance and configuration issues. For example if we have an equal hyper volume size over FC drives of 300 GB 15k rpm and 146 GB 10k rpm the higher performance will be achieved with the volume with 146 GB drives although drives are slower. This is because of the fact that 4 drives at 300 GB are needed for one hyper volume or 8 drives at 146 GB. In the first case all IO operation will be distributed over 4 disks and

[1] Mario O. Krastev is with the Faculty of Telecommunications, TU-Sofia, 1000 Sofia, Bulgaria, e-mail: mkrastev@tu-sofia.bg

[2] Ludmila H. Raikovska is with the Faculty of Telecommunications, TU-Sofia, 1000 Sofia, 8 Kliment Ohridski blvd., Bulgaria, e-mail: l.raikovska@snt.bg

[3] Kiril I. Dakov is with S&T Bulgaria eood, 1528 Sofia, 7 Iskarsko shosse blvd., Bulgaria, e-mail: k.dakov@snt.bg

over 8 in the second and thus the slower drives would feature better performance [4].

Therefore not only drive size and rotation speed should be taken into account in the layout configuration design of the storage system, but also latency, access times and configuration features. In most cases is not recommended to use the biggest capacity disk because it will cause performance degradation.

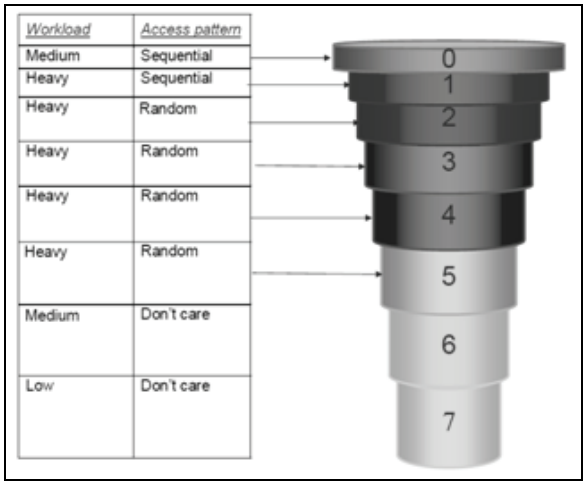


Fig.1. Disk drive workload

The access pattern of a single drive is presented on Fig.1. One of the drive parameters that is most often neglected is the seek time. This is the time for moving drive heads from current position to the proper track. If the volumes are not correctly configured, the drives' heads will move from inner to outer end and vice verse very frequently during random IO operations. It is of major importance to take this into account especially when we have random read/write operations.

For performance oriented applications with a lots of random requests it is recommended that only the middle part of drive plates should be used and not to use about 25% in the inner and outer part [3]. The outer parts are rotating faster than the inner, so they are very suitable for consequential operations.

In the discussed case there are a lot of random queries for most applications so it will be better to use bigger drives, with fast rotation speed and to use only the middle part of the drive. This will minimize the latency of moving of the heads.

The number of maximal theoretical IOPS for each drive can be obtained from the following expression:

$$\frac{1000ms}{AvgRotationLatency + Avr\,\mathrm{Re}\,adSeek} = IOPS \quad (1)$$

Multiplying this IOPS by the number of drives we obtain the maximal theoretical operations number which this part of the array can serve. Maximal IOPS weight from host side is calculated from the formula:

$$\mathrm{ReadOp}.MW + \mathrm{WriteOp}.K.MW = IOPS \quad (2)$$

where

MW – maximal IOPS which a host can handle

ReadOp – percentage of read operations in all IOPS to this host

WriteOp - percentage of write operations in all IOPS to this host

K – RAID protection level coefficient. It shows the number of parallel writes that are done over drives for one single operation. The most common values are K=1 for RAID-0 – no redundancy; K=2 for RAID-10, because all data are written simultaneously on two drives; K=3 for RAID 5 (3+1) – data are written over 3 drives.

Therefore the total number of IOPS which a single host can handle is:

$$MW = \frac{IOPS}{\mathrm{ReadOp} + \mathrm{WriteOp}.K} \quad (3)$$

The proportion between read and write operations for different applications could be obtained from Table I. The above formulas represent only a theoretical evaluation but in a real production system these values should be corrected. This correction is done during the configuration of the storages.

In a configuration process following the best practices of the storage manufacturer EMC [3], one hot spare drive is dedicated for each 30 drives and two drives per director pair are dedicated for service and system information.

C. Optimal drive configuration

In order to obtain a solution, fully compliant to all specific Telco requirements, we propose/choose to divide the storage capacity into different groups. For a Telecommunication company usually at least two groups are necessary. The first group will contain fast volumes for most performance oriented applications with more than 3000 IOPS and the second group will consist of slow volumes for the rest of the applications. Our solution is significantly different from the standard way proposed by the storage vendor. The main purpose of the default configuration is to achieve maximum workload for the whole system and to distribute equally all drives and IOPS between directors, which does not take into account the specific performance needs of Telco applications.

III. TELECOM STORAGE SYSTEM EXAMPLE

The proposed method has been implemented in a real working environment. The infrastructure consists of SAN environment, servers running under IBM AIX operation systems and the storage arrays are EMC Symmetrix DMX3 [3]. During the planning phase of the IT solution 40% increase of the usable capacity and 10% increase of performance needs within the next three years have been taken into account.

The IT infrastructure for information protection and disaster recovery includes three identical storage arrays situated at three different geographical sites, working in concurrent storage based synchronous and asynchronous replication between them [2]. Additional capacity for BCV – business continuity volumes (similar to snapshots and clones) has been

also calculated and the different scenarios for IO workload, needed to achieve storage optimization without additional recourses have been examined.

In the primary data centre the main IT resources are located. Local and remote replications are performed on the main storage array. Local replicas are for tests, backups and statistical needs without impact on the production systems. The proxy site is located about 25 km away from the main data centre (DC) and is synchronized in real time with the primary data centre. In case of failures in the primary DC, all applications except „Development, Test, and Data warehouse" could be started there. The third site is identical with the primary. It is located at about 800 km apart from the central DC and there is asynchronous replication between storages at primary and proxy sites and disaster site. The main idea is that in case of disaster the proxy site should assure synchronous copy of the production data until the disaster site becomes ready to failover all operations and business applications [2]. At the disaster site stand by servers and storages have been located, ready in critical situations to take over all operations. There are secondary independent systems for Data warehouse Test and Development at the disaster site. Connections between primary DC and proxy site are over CWDM, and between primary DC and remote disaster site - over DWDM. All lines are doubled and independent on each other.

In this example the configuration and optimal disk space configuration for the primary DC storage has been examined. This system is the most important in whole Telco IT infrastructure and serves not only as storage for business critical applications but also it takes part in replications, backup and assures business continuity plan for the organization. The configurations of the other two systems are more or less the similar.

A. Fast group

The fast group needs 6,5 TB storage capacity and should handle at least 24000 IOPS at peak workload. There is planned 40% growth of the capacity and 10% performance increase. Due to the fact that there are big requirement for IOPS, two disk directors' pairs would be dedicated for this group, with 16 DAE connected to each, which means that maximum 240 disk drives will be dedicated to the fast group. All drives will be configured in RAID 10, striped meta-volumes with 8 hyper-volumes. Each hype-volume will be distributed on different slice from the directors.

According DMX system recommendation [3] eight drives are dedicated as global hot spares and 4 drives are for service and system information, so for customer data 224 drives will be used. The drive type for the fast group will be 146GB 15k rpm. This implies about 16535 GB usable capacity. To achieve maximum performance we need only about 40% to 70% of disk capacity, which will be used at the middle part of the plates (as shown on Fig. 1). Therefore the real usable capacity will be between 40-60% or from 6000 GB to 9800 GB. These values fully satisfy requirements for the fast group as well as the predicted capacity growth - up to 9100 GB within next three years.

According to (1) the theoretical maximum of IOPS for this drives is evaluated. We have 181,8 IOPS per drive or 40732 IOPS for the whole fast drives group. Now using the expression (3) we obtain the theoretical maximum of IOPS from host side – MW = 32578. This value takes into account the RAID protection and 3:1 relation between read and write operations.

It is very difficult to have such big workflow and achieve these IOPS in practice. Statistics from real storage arrays shows that practically for 146 GB 15k rpm drives there are about 130 IOPS, which means 33600 IOPS for the whole fast group or 26880 IOPS maximum workload from hosts' side.

These values show the maximal number of I/O operations from hosts which storage array could handle. This is only in worst case scenario at maximum workload and it is achieved in very rear situations. Even in this case it is obvious that the proposed configuration of fast group drives will satisfy the performance requirements of the applications and will assure about 2800 IOPS reserve over the desired level. It will comply with the predicted 10% growth in the next three years.

B. Slow group

The 40 TB storage capacities will be used for the slow group. These volumes would handle about 5000 IOPS maximum workload. There is the planned 40% capacity growth and 10% performance increase. In this group not only production volumes will be included but also all volumes needed for replication (local and remote). His group is not so performance critical, so only one director pair will handle all transactions. Having in mind that this group of drives has enormous size - 40TB we will have 16 DAE directly connected to directors' pair and 16 DAE connected into Daisy Chain to the first ones. This scheme will lead to performance degradation about 7-10% for all drives. For the slow group there are also 240 dedicated drives. Protection level for this drives will be RAID 5 (3+1) for Development and Test & Applications and unprotected RAID 0 volumes for data replication. These RAID levels are chosen in order to achieve best price, performance and capacity ratio for the solution. The RAID 5 volumes will be used due to specific application behavior, in this case we lose only 25% of disk capacity for protection, the performance is not a critical issue and volumes workload is not heavy. The RAID 0 groups will be used only for replicas (the source volumes are protected) so they do not need special protection and the performance will be quite satisfactory due to parallel operation over multiple drives.

Following best practice configuration rules [3] for the slow group there is dedicated 8 global hot spare drives and 2 drives for system and service needs. So for user data there are 230 drives available. Due to more effective service and maintenance requirements for the system we propose to use different drives for protected (RAID 5) and unprotected (RAID 0) volumes. Although it is not a default configuration of the storage the main benefit of this approach is that a failed drive in one of groups will not impact the other. For example if on the failed drive there are only one type of volumes - unprotected, it could be replaced immediately without waiting a rebuild process to finish, the information will be lost but it will be renewed immediately after drive change. Capacities of volumes used for replication (BCV) are presented in Table 3:

TABLE III. REPLICATION VOLUMES CAPACITY

Application	Size, TB	BCV	BCV size, TB	Max size within 3 years
Billing	3	Yes, 3 pairs	9	12,6
CRM	1,5	Yes, 3 pairs	4,5	6,3
ERP	2	Yes, 3 pairs	6	8,4
Development	10	Yes, 2T, 20%, 1 pair	2	2,8
Total , TB	16,5	-	21,5	30,1

For the slow group we propose low cost FC drives – 500 GB 7,2k rpm. For better balance of the array we keep the structure of the fast group – the number of drives for hyper and meta volumes should be dividable to 8, and should comply to size requirements, so for unprotected volumes (RAID 0) 64 drives will be needed. This fully answer the requirements for disk space and future growth of 40% within next three years and provide 30100 GB usable capacity.

Having in mind the disk specifications (Table II), we calculate the maximum IOPS per single drive 78,9 IOPS and for the whole unprotected volumes form slow group are 5069 IOPS. In this case we have no loses from protection type (RAID 0) but we have to take into account that back-end is connected in daisy chains, which leads to 7-10% performance degradation. Therefore the maximum theoretical workload from host side should be 4544 IOPS. These values are practically hard to be achieved, in most cases from statistic data the maximal workload which a single 500 GB FC drive handles, is 60 IOPS. This means that the whole unprotected group could receive 3840 IOPS and the maximum host workload should be 3456 IOPS. This numbers fully satisfy the requirements for local and remote replications because they work mainly in background and transparently for host, so after the base synchronization only changed blocks are written to drive tracks.

For the other part of the slow group, which will be RAID 5 (3+1) protected, 166 usable drives remain. The closest number of drives which will meet the capacity and performance need and is dividable to 8 is 160. These 160 drives provide 60 TB maximal usable capacity, which over meet needs for 40% data growth within next three year from currently needed 40 TB to predicted 56 TB. Performance demands for slow group applications show that for Development applications 3300 IOPS would be needed and up to 4950 IOPS during next three years, while Development, Test, Data warehouse application will need 2200 IOPS and up to 2347 IOPS during the next three years. Therefore the whole workload for slow group will be 4950+2347=7297 IOPS. These are theoretical calculations for IOPS from host side. In a real practical situation it couldn't be reached.

A theoretical calculation for 160 drives pointed out that maximum workload is 12624 IOPS. The real number from practical measurements and statistics is 9600 IOPS. Taking into account the specific backend losses due to daisy chain configuration the proposed solution will provide 8640 IOPS in worst case which will assure about 1200 IOPS reserve for these applications.

So finally for the whole slow group will be used 234 drives, from which 64 unprotected for BCV, 160 for production data (RAID 5) and 10 for hot spare and service needs. The summary of the proposed configuration for the primary storage array is presented in Table IV.

TABLE IV. PROPOSED SPACE DISTRIBUTION

Disk group	Drive type	Number of drives	Hot spares	System drives	RAID level	Usable drives
Fast	146 GB 15k rpm	236	8	4	1,0	224
Slow	500 GB 7,2k rpm	234	8	2	0	64
					5 (3+1)	160

In the same manner the capacity in proxy and remote data center have been configured. These configurations lead to significant performance improvement compared to proposed default methods. Using the standard method have to mix protected and unprotected volumes on same disk drive and no performance oriented group are constructed. In this case can obtain more usable capacity but the maximum IOPS which system can handle are about 30% less.

IV. CONCLUSION

The most important requirements for corporate storage systems in Telecom operators have been examined. A new solution for development of reliability, performance and effective use of capacity by choosing proper configuration and design has been proposed. Needs for local and remote replications and configuration of the drives in order to obtain maximum workload of the system have been taken into account.

The storage capacity has been divided into different groups which serve specific application needs. The whole available capacity is divided between the two groups called fast and slow. The result is up to 30% higher utilization IOPS than with standard default configuration recommended by manufacturer.

Proposed solution is based on real statistics data and configurations from Bulgarian telecommunication operator. Although the current research is done for EMC Symmetrix DMX3 storage arrays the same method of configuration could be applied for all other systems due to the fact that the specific Telco application requirements are more or less the same. The described solution is according the latest tendencies, utilize most of the advantages of the enterprise storage systems and follow the best practices and strategies for performance, data protection and reliability.

V. REFERENCES

[1] Raikovska, L., N.Dimchev, M.Saykov, "A modern high performance approach for data storage infrastructure design and data protection", Telecom conference 2004, Varna

[2] Raikovska, L., N.Dimchev, M.Saykov, „Information protection and data transfer methods for disaster recovery solutions", Telecom conference 2004, Varna

[3] EMC Symmetrix DMX-3 Product Guide, Hopkinton, MA, EMC Corporation, November 2006

[4] Economies of Capacity and Speed: Choosing the most cost-effective disc drive size and RPM to meet IT requirements, Seagate Technology Paper, May 2004

Tehnological Aspects of FiWi for Broadband Multimedia Services

Aleksandra Kostić-Ljubisavljević[1], Andreja Samčović[2], Vladanka Aćimović-Raspopović[3]

***Abstract* – Fiber-wireless (FiWi) networks become rapidly mature and give rise to new powerful access network solutions and paradigms. Towards the technical evolution of wireless-optical access networks and the seamless coexistence of both technologies, this paper presents key technical challenges for providing seamless communications in FiWi access networks, and highlights important issues for technologies, architectures, some application areas as well as broadband multimedia services over FiWi networks.**

***Keywords* – Fiber-wireless, technology, architectures, broadband services**

I. INTRODUCTION

In the last decade, wireless and optical communications have experienced rapid growth and commercial success. These great successes in both areas have started fueling the vision of ubiquitous broadband multimedia communications. Building on advances in network infrastructure, low-power integrated circuits, and powerful processing algorithms, the convergence of wireless and optical communications is now underway, paving the way towards widespread acceptance of fiber-wireless (FiWi) access networks in the next decade. FiWi access integrates the high capacity of optical with the flexibility of wireless to enable high-speed communications to mobile users in a transparent and seamless manner.[1] FiWi can be implemented using different architectures. Notable among these, are architectures where optical technologies are used as backhaul network for wireless access points. In particular, architectures which use wireless mesh networks over optical fibers are very attractive, owing to their ease of installation and cost-effectiveness, which allow for wide coverage as well as rapid deployments to support unanticipated needs. Such architecture can yield numerous advantages, such as reduced complexity and flexible resource allocation. Indeed, functions such as modulation/demodulation, up/down-converting, and data multiplexing, can be performed at a central office serving several wireless access points, which in turn reduces operational costs, increases flexibility, and facilitates network scalability. Furthermore, radio resources can be allocated dynamically to wireless sites where they are mostly needed, thus making better usage of the available radio resources, and simplifying mobility handling. Achieving convergence requires the development and evaluation of new architectures and communications techniques that establish the benefits of both technologies while avoiding their shortcomings. In this vein, challenges are involved in all planes, form the network design, planning and dimensioning, to components and business models.

This paper is organized as follows. After the introduction, fiber-wireless networks enabling technologies are presented. Various architectures considering FiWi are given in the next section. The following section is dedicated to radio over fiber application areas. Broadband multimedia services over FiWi conclude this paper.

II. FIWI NETWORKS ENABLING TECHNOLOGIES

Currently, there are two technologies used to implement fiber-wireless (FiWi) networks: free space optical (FSO), also known as optical wireless (OW) and radio over fiber (RoF) [2].

FSO is a type of direct line-of-sight (LOS) optical communications that provides point-to-point connections by modulating visible or infrared (IR) beams. It offers high bandwidth and reliable communications over short distances. The transmission carrier is generated by deploying either a high-power light emitting diode (LED) or a laser diode, while the receiver may deploy a simple photo detector. Current FSO systems operate in full-duplex mode at a transmission rate ranging from 100 Mb/s to 2.5 Gb/s, depending largely on weather conditions. Given a clear LOS between source and destination and enough transmitter power, FSO communications can work over distances of several kilometers. At both source and destination, optical fiber may be used to build high-speed LANs, such as Gigabit Ethernet (GbE).[3]

RoF, on the other hand, allows an analog optical link to transmit a modulated radio frequency (RF) signal. There are different techniques available to realize RoF networks. Typically, an RoF transmitter deploys a Mach-Zehnder intensity modulator in conjunction with an oscillator that generates the required optical carrier frequency, followed by an Erbium doped fiber amplifier (EDFA) in order to increase the transmission range. RoF networks provide both point-to-point and point-to-multipoint connections. Recently, a full-duplex RoF system providing 2.5 Gb/s data transmission over

[1]Aleksandra Kostić-Ljubisavljević is with the Faculty of Transport and Traffic Engineering, Vojvode Stepe 305, 11040 Belgrade, Serbia, E-mail: a.kostic@sf.bg.ac.rs

[2]Andreja Samčović is with the Faculty of Transport and Traffic Engineering, Vojvode Stepe 305, 11040 Belgrade, Serbia, E-mail: andrej@sf.bg.ac.rs

[3]Vladanka Aćimović-Raspopović is with the Faculty of Transport and Traffic Engineering, Vojvode Stepe 305, 11040 Belgrade, Serbia, E-mail: v.acimovic@sf.bg.ac.rs

40 km with less than 2 dB power attenuation was successfully demonstrated using the millimeter-wave band. There are many cost-efficient optical approaches to mixing and upconverting millimeter wave signals.

III. FiWi Architectures

In this section will be presented available architectures for enabling FiWi integration. For instance, the integration of Ethernet passive optical networks (EPON) and WiMAX access networks can be done in several ways; according to [4], the following four architectures can be used.

Independent Architecture — In this approach WiMAX base stations serving mobile client nodes are attached to an optical network unit (ONU) just like any other wired subscriber node, whereby an ONU denotes the EPON customer premises equipment. WiMAX and EPON networks are connected via a common standardized interface (e.g., Ethernet) and operate independent of each other.[5]

Hybrid Architecture — This approach introduces an ONU-base station (ONU-BS) that integrates the EPON ONU and WiMAX BS in both hardware and software. The integrated ONU-BS controls the dynamic bandwidth allocation of both the ONU and BS.

Unified Connection-Oriented Architecture — Similar to the hybrid architecture, this approach deploys an integrated ONU-BS. But instead of carrying Ethernet frames, WiMAX media access control MAC protocol data units (PDUs) containing multiple encapsulated Ethernet frames are used. By carrying WiMAX MAC PDUs, the unified architecture can be run like a WiMAX network with the ability to grant bandwidth finely using WiMAX's connection-oriented rather than EPON's queue-oriented bandwidth allocation.

Microwave-over-Fiber Architecture — In this approach the WiMAX signal is modulated on a wireless carrier frequency, and is then multiplexed and modulated together with the baseband EPON signal onto a common optical frequency (wavelength) at the ONU-BS. The central node consists of a conventional EPON optical line terminal (OLT) and a central WiMAX BS, called a macro-BS. The OLT processes the baseband EPON signal, while the macro-BS processes data packets originating from multiple WiMAX BS units.

Besides the aforementioned generic integration approaches of EPON and WiMAX networks, several other FiWi architectures based on WiFi technology have been studied, as described in the following. The network shown in Figure 1. interconnects the central office (CO) with multiple WiFi-based wireless access points (WAPs) by means of an optical unidirectional fiber ring. The CO is responsible for managing the transmission of information between mobile client nodes (MCNs) and their associated WAPs as well as acting as a gateway to other networks. Each WAP provides wireless access to MCNs within its range. All MCNs take part in the topology discovery, whereby each MCN periodically sends the information about the beacon power received from its neighbors to its associated WAP. In doing so, WAPs are able to estimate the distances between MCNs and compute routes. Multihop relaying is used to extend the range. To enhance the reliability of the wireless link, the CO sends information to two different WAPs (path diversity). The proposed implementation can support advanced path diversity techniques that use a combination of transmission via several WAPs and multihop relaying (e.g., cooperative diversity or multihop diversity). Consequently, the CO must be able to assign channels quickly and efficiently by using one or more wavelength channels on the fiber ring to accommodate multiple services such as wireless LAN (WLAN) and cellular radio network.

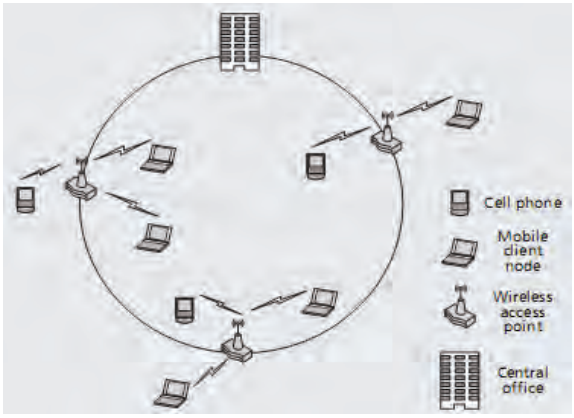


Figure 1. Optical unidirectional fiber ring interconnecting WiFi-based wireless access points [1]

Figure 2 depicts a hybrid FiWi architecture that combines optical star and ring networks. Each fiber ring accommodates several WiFi-based WAPs, and is connected to the CO and two neighboring fiber rings via optical switches. The optical switches have full wavelength conversion capability, and interconnect the WAPs and CO by means of shared point-to-point light-paths. The network is periodically monitored during prespecified intervals.

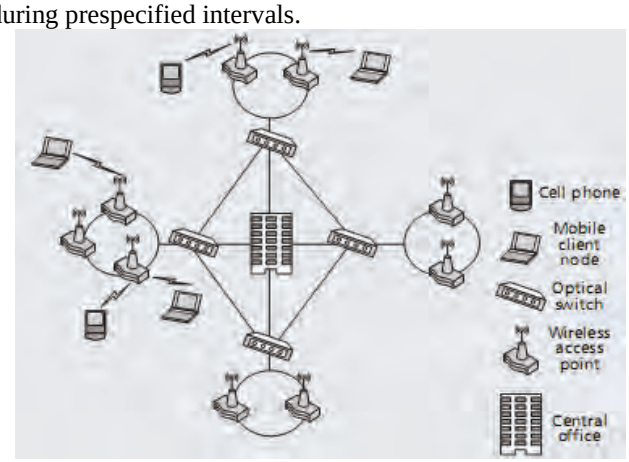


Figure 2 Optical hybrid star-ring network integrated with WiFi-based wireles access points [1]

At the end of each interval, the lightpaths may be dynamically reconfigured in response to varying traffic demands. When traffic increases and the utilization of the established lightpaths is low, the load on the existing lightpaths is increased by means of load balancing. Otherwise, if the established lightpaths are heavily loaded, new lightpaths need to be set up, provided enough capacity is available on the fiber links. In the event of one or more link failures, the affected lightpaths are dynamically reconfigured using the redundant fiber paths of the architecture.

The FiWi network shown in Figure 3 consists of an optical wavelenght division multiplex (WDM) backhaul ring with multiple single-channel or multichannel PONs attached to it.

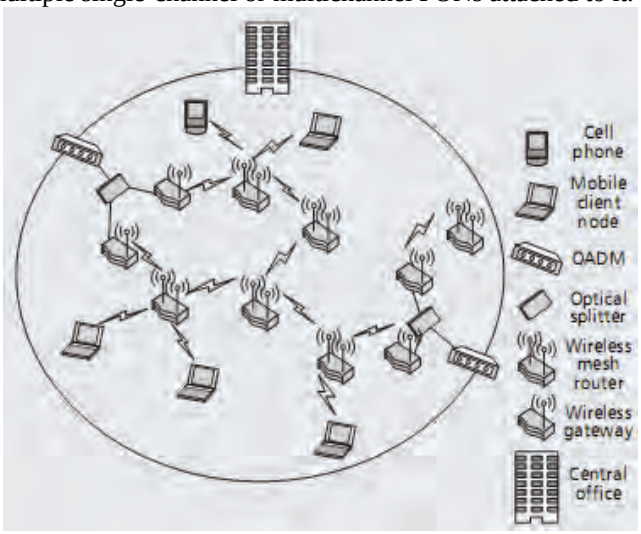


Figure 3 Optical unidirectional WDM ring integrated with a WiFi-based wireless mesh network [1]

More precisely, an optical add-drop multiplexer (OADM) is used to connect the OLT of each PON to the WDM ring. Wireless gateways are used to bridge PONs and WMNs. In the downstream direction, data packets are routed from the CO to the wireless gateways through the optical backhaul and then forwarded to the MCNs by wireless mesh routers. In the upstream direction, wireless mesh routers forward data packets to one of the wireless gateways, where they are then transmitted to the CO on one of the wavelength channels of the optical backhaul WDM ring, as each PON operates on a separate dynamically allocated wavelength channel. Since the optical backhaul and WMN use different technologies, an interface is defined between each ONU and the corresponding wireless gateway in order to monitor the WMN and perform route computation taking the state of wireless links and average traffic rates into account. When the traffic demands surpass the available PON capacity, some of the TDM PONs may be upgraded to WDM PONs. If some PONs are heavily loaded and othershave less traffic, some heavily loaded ONUs may be assigned to a lightly loaded PON by tuning their optical transceivers to the wavelength assigned to the lightly loaded PON. This architecture provides cost effectiveness, bandwidth efficiency, wide coverage, high flexibility, and scalability. In addition, the reconfigurable TDM/WDM optical backhaul helps reduce network congestion and average packet latency by means of load balancing. Moreover, the dynamic allocation of radio resources enables cost-effective and simple handovers.

IV. Radio-Over-Fiber Application Areas

To fully consider the application of RoF technology one must look at the radio technologies that are to be transported over this network. From the perspective of most radio air-interface designers the inclusion of an optical fiber link into the link budget of the interface is an additional challenge. The already tight noise and distortion budgets offer little flexibility and therefore often require high performance from the optical link. In effect, the RoF link designer is being asked to provide a virtually transparent pipe through which a radio signal can be transported.

In this section will be given a brief review of some of the radio standards that have been considered for RoF systems.

Cellular radio systems (GSM and UMTS) have been extensively investigated with commercial offering having been available for some time. Research has typically concentrated on multiple channel operation or considering the extreme performance limits of the systems. For example simulation studies have considered the interaction of second-generation (GSM) and third-generation (W-CDMA) wireless systems orW-CDMA and (WLAN). In these systems distortion due to the linearity of the optical components is a major consideration as discussed previously with solutions such as predistortion being proposed.

As new radio standards appear, different issues arise although most are based on the radio formats' susceptibility to distortion. For example, ultra-wideband (UWB), an impulse radio standard that uses low power and a very wideband bandwidth (typically >500MHz), has been developed for short reach systems. This bandwidth is significantly greater than other radio standards and introduces a number of new challenges relating to linearity and induced distortion, with investigations considering a number of aspects including its co-existence with WLAN. Additionally, optical techniques have also been investigated for the generation of the short pulses that form these signals.

WiFi (IEEE 802.11a/b/g etc.) systems have been researched widely in part due to he easy availability of access points for experimentation. However, due to the low cost of the access points it is hard to see fiber-based system displacing electronic systems unless it forms part of a wider converged, multi-standard network.

Orthogonal frequency division multiplexing (OFDM), which was introduced to the mainstream of communications by the digital video broadcast (DVB) standard and to data communications by the 802.11g standard, forms the mainstay of most of the next-generation communications standards. To date a number of investigations into the transport of OFDM have been made. Theoretical studies have evaluated the performance of OFDM systems in the presence of non-linear distortion from optical components. A number of experimental investigations have been conducted, although most focus on the transmission of OFDM-based standards.

Yhe transport of radio standards to significant distance using optical fiber is possible, at least at the physical layer. However, other effects can also impact on the system performance. In most experiments a commercial vector signal generator and analyzer setup is used for performance measurements. This offers a controllable signal with high purity to allow repeatable experiments. It does, however, have the drawback of only operating at the physical layer. It must be remembered that the addition of optical fiber into the system can add significant delay into what would normally be the air-interface section of the link. In wide area and centrally controlled architectures, such as GSM or WiMAX, this may not cause significant issues. However, in ad hoc and short-

reach architectures, such as 802.11, the addition of 5 μs of delay per kilometer of fiber (10 μs if a round trip is considered) may make a significant contribution to the timeouts of the 802.11 MAC. It is shown that for typical implementations a maximum distance of around 13 km of fiber is permissible before the system fails with complete loss of transmission. It should also be noted that these distances are vendor dependent as the standard only specifies a minimum time-out. The data rate reduces progressively as fiber is added due to the increased delay in the system. One can then see that when the time-out (for either acknowledge of packets in the basic access mechanism or for response to a request to send (RTS) packet in the virtual carrier sense or RTS/CTS mechanism) is exceeded due to the fiber delay, transmission falls sharply. In this section we have reviewed the main, current radio standards and looked at research that has considered their transmission of optical fibers. In general it is seen that transmission is possible, although in some cases linearization either by design in the components or by system techniques is required.

V. Multimedia Services Over FiWi

Access networks, services and driving technologies currently undergo changes that have never been seen before.[6] Differences between telecommunications and media delivery domains blur day by day. Integrated multimedia access is the key for revenue generation, customer satisfaction and ultimately for the survival of providers. There is a tradewar in the home access scenario between the traditional voice telecommunication companies and coaxial cable television providers. Both these sectors are now geared to provide bundled multimedia (video, voice and data) services over their networks, to enter the so-called 'triple play'. More and more households have large-screen, high definition televisions (HDTV); large number of HDTV channels are also available. High quality HDTV needs good signal-to-noise ratio (SNR) and high bitrate. As a result both these players need more bandwidth than ever before.Wireless access technologies are also on the transform. Digital video/audio broadcasts (DVB and DAB) over terrestrial or satellite networks are geared to merge with wireless access networks. Wireless cellular network subscribers are no longer satisfied with just voice; they expect different types of data and streaming video. As a result, the demand for multimedia wireless services is also on the rise. There is some interest on Internet Protocol Television (IPTV) because of the promising interactive nature. IPTV delivers video signals using IP technology over a broadband connection. Currently, most operational IPTV networks are deployed on telecommunication network architectures using various technologies to bridge the last mile. Note that despite the IPTV hype, the majority of IPTV services use MPEG-2 compression that has compromised quality. Without the ability to provide HDTV with interactive services (video on demand), DolbyTM quality surround sound audio and some data (such as closed captioning) no media provider can survive. Although, the uncompressed bit rates for big screen HDTV is very high, typically in the Gbps range, good quality HDTV is provided by 19 or 38 Mb/s with complex video and audio compression technologies. Generally, the video signal could be either radio frequency (RF) video or IP (Internet Protocol) video. The IP video is similar to Internet traffic with additional streaming requirement. However, the RF video signal is analog and subcarrier multiplexing is needed to transmit multiple television channels. Furthermore, the RF video is a proven technology that already dominates existing fiber-coaxial networks and has better quality.

VI. Conclusion

The ultimate goal of the Internet and communication networks in general is to provide access to information when we need it, where we need it, and in whatever format we need it. To achieve this goal, wireless and optical technologies play a key role. Wireless and optical access networks can be thought of as complementary. Indeed, optical fiber does not go everywhere, but where it does go, it provides a huge amount of available bandwidth.

Wireless access networks, on the other hand, potentially go almost everywhere, but provide a highly bandwidth-constrained transmission channel susceptible to a variety of impairments. Clearly, as providers need to satisfy users with continuously-increasing bandwidth demands, future broadband access networks must leverage on both technologies and converge them seamlessly, giving rise to fiber-wireless (FiWi) access networks.

Acknowledgement

This research has been supported by the Ministry of science and technological development of the Republic of Serbia.

References

[1] N.Ghazisiadi, M.Maier, C.Assi, "Fiber-Wireless (FiWi) access networks: a survey", IEEE Communications Magazine, Vol.47, No.2, pp. 160-167, February 2009.

[2] S.Aissa, M.Maier, "Towards seamless fiber-wireless (FiWi) access networks: convergence and challenges", Proc. International Conference on Transparent Optical Networks (ICTON)-'Mediterranean Winter' 2007, Sousse, Tunisia, Dec. 2007

[3] X.Fernando, "Radio over fiber in multimedia access networkss", Proceedings of the 1st International conference on Access networks, Athens, Greece, 4-6. September 2006.

[4] G. Shen, R. S. Tucker, C.-J. Chae, "Fixed Mobile Convergence Architectures for Broadband Access: Integration of EPON and WiMAX," IEEE Communications Magazine, Vol.45, No.8, August 2007, pp. 44–50

[5] A.Shami,·M.Maier,·C.Assi, editors, "Broadband Access Networks: Technologies and Deployments, Springer, 2009

[6] S. Z. Pinter and X. N. Fernando, "Fiber-Wireless solution for broadband multimedia access", IEEE Canadian Review, Summer 2005.

Intra-cell Handover in OFDMA-based Wireless Access Networks

Kiril M. Kassev[1] and Boris P. Tsankov[2]

***Abstract* – OFDMA-based wireless access networks employ Adaptive Modulation and Coding at physical layer that results in different bit rate per radio transmission unit at different location in the cell. The cell is decomposed into regions. The mobile users encounter intra-cell handover at the border between two neighbour regions. The paper considers the intra-cell handover influence over system performance. An analytical method is proposed and numerical results are reported.**

***Keywords* – Adaptive modulation and coding, intracell handover, OFDMA.**

I. INTRODUCTION

In order to satisfy the customer demands for access to a wide variety of multimedia applications over wireless communication networks, significant research efforts are being put in development of next generation mobile networks (NGMN), also referred to as 4G, based on all-IP environment. This can be achieved by employing the OFDMA technique at the air interface, which is immune to intersymbol interference and frequency selective fading, due to the time-varying nature of the wireless channel. As a consequence, OFDMA-based systems have become a popular choice, and have been adopted in several standards towards NGMN [1], [2].

In contrast to the traditional channelized multiple access systems, where each user is assigned a fixed amount of bandwidth during the whole connection time, in wireless networks the channel capacity of a wireless link is time-varying, and thus the quality of service (QoS) requirements may not be satisfied, even though a large amount of resource (i.e. bandwidth) is allocated to a certain connection. This is especially true when a mobile station (MS) is located at the cell edge area. The Adaptive Modulation and Coding (AMC) within the OFDMA based cell means to choose the best Modulation and Coding Scheme (MCS) for given channel quality. AMC is a powerful tool to enhance the overall wireless access network performance [3], [4].

Due to the AMC implementation the OFDMA-based cell is subdivided on concentric rings, each ring characterized by particular modulation and coding scheme depending on the radio channel condition [5]. A MS during its movement might cross the ring border with consequent intra-cell handover (HO). The intra-cell HO leads to a change of resource amount placed to the connection disposal.

The main observation of this paper is the intra-cell HO influence to the overall cell performance. The mobility is mostly related to multimedia services and voice services in particular. The analytical model proposed and the numerical examples considered are based on voce like services with constant bit rate and corresponding QoS parameters.

The rest of the paper is organized as follows: Some related papers are listed in the Section II. The analytical model for teletraffic evaluation of the impact of intra-cell HO is presented in section III, followed by the numerical studies in Section IV. We conclude the paper and present possible suggestions for further study in Section V.

II. STATE OF THE ART

The interaction of analytical framework for evaluation of teletraffic performance characteristics of the system under consideration at the data link layer with AMC at the physical layer provides a base for interesting and realistic design work. This topic is covered by a number of excellent papers describing in depth the essence of the underlying problems. Reference [6] investigates the performance of transmission over wireless links, in case an interaction between a finite-length queuing and different AMC schemes are taken into consideration. The Erlang capacity of WiMAX systems with fixed modulation schemes is calculated in [7], considering two traffic classes – streaming and elastic. An interesting approach for evaluation of Erlang capacity of a multi-class TDMA system with AMC by separating the calculation of blocking and outage probabilities is proposed by the authors of [8].

Interestingly, a similar problem in GSM networks has been considered for full-rate and half-rate transmission, depending on the radio conditions [9-11]. In this case, there only exist two cell areas with different radio channel characteristics.

III. SYSTEM MODEL AND PERFORMANCE ANALYSIS

We will restrict our considerations to voice-like real time traffic with constant bit rate and the following assumptions:

- MSs and their traffic are uniformly distributed over the cell;
- MS's movement direction is uniformly distributed over $[0, 2\pi]$;
- New call arrival follow a Poisson process with rate λ;
- The call holding time is exponentially distributed with mean $1/\mu$.

[1]Kiril M. Kassev is with the Faculty of Telecommunications, Technical University of Sofia, 8 Kliment Ohridski Blvd., 1756 Sofia, Bulgaria, E-mail: kmk@tu-sofia.bg

[2]Boris P. Tsankov is with the Faculty of Telecommunications, Technical University of Sofia, 8 Kliment Ohridski Blvd., 1756 Sofia, Bulgaria, E-mail: bpt@tu-sofia.bg

The employment of the AMC scheme at the physical layer results in splitting-up the cell into non-overlapping cell areas (rings), as shown on Fig. 1. The total cell capacity is C resource units (time slots). A call to (from) a MS in ring l requires exactly d_l resource units in order to be served. The rings are numbered starting with the most inner being number l to the most outer cell ring being number L. Therefore, $d_x > d_y$, if $x > y$.

The system state is defined by the set $N\{n_1, n_2, ..., n_L\}$, where n_l is the number of calls (number of active MSs) in ring l, ($l = 1, L$).

The call duration and ring l dwell time are assumed to be negative exponentially distributed random variables with mean $1/\mu$ and $1/\delta_l$, respectively.

Intracellular HO calling rate from ring l to ring $l+1$ is

$$\lambda_l^{h+} = n_l \delta_l^+.$$

$1/\delta_l^+$ is mean dwell time of calls that handover outward of ring l. By analogy, the intracellular HO calling rate from ring l to ring $l-1$ is

$$\lambda_l^{h-} = n_l \delta_l^-.$$

We use notation $\{n_l\}$ for any state $N_l\{..., n_l, ...\}$, where the number of calls in ring l is equal to n_l.

The possible transition out of any state $\{n_l\}$ and into the same state $\{n_l\}$, due to changes in ring l, are shown on Fig. 2. Therefore, in a condition of statistical equilibrium there is

$$P_{\{n_l\}}[\lambda_l d_l + n_l \mu d_l + n_l \delta_l^+ (d_{l+1} - d_l) + n_l \delta_l^- (d_l - d_{l-1}) +$$

$$+ n_{l-1} \delta_{l-1}^+ (d_l - d_{l-1}) + n_{l+1} \delta_{l+1}^- (d_{l+1} - d_l)] =$$

$$= P_{\{n_l-1\}}[\lambda_l d_l + n_{l-1} \delta_{l-1}^+ (d_l - d_{l-1}) + n_{l+1} \delta_{l+1}^- (d_{l+1} - d_l)] +$$

$$+ P_{\{n_l+1\}}[(n_l + 1)\mu d_l + (n_l + 1)\delta_l^+ (d_{l+1} - d_l) +$$

$$+ (n_l + 1)\delta_l^- (d_l - d_{l-1})], \qquad l = 2, ..., L-1. \qquad (1)$$

It is obvious that

$$\delta_1^- = 0, \qquad (2)$$

and that δ_L^+ is the call handover rate from ring L to another cell.

Therefore, for $l = 1$

$$P_{\{n_1\}}[\lambda_1 d_1 + n_1 \mu d_1 + n_1 \delta_1^+ (d_2 - d_1) + n_2 \delta_2^- (d_2 - d_1)] =$$

$$= P_{\{n_1-1\}}[\lambda_1 d_1 + n_2 \delta_2^- (d_2 - d_1)] +$$

$$+ P_{\{n_1+1\}}[(n_1 + 1)\mu d_1 + (n_1 + 1)\delta_1^+ (d_2 - d_1)]. \qquad (3)$$

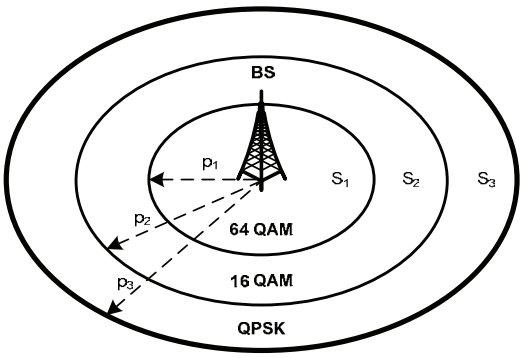


Fig. 1. OFDMA-based cell with AMC scheme employed

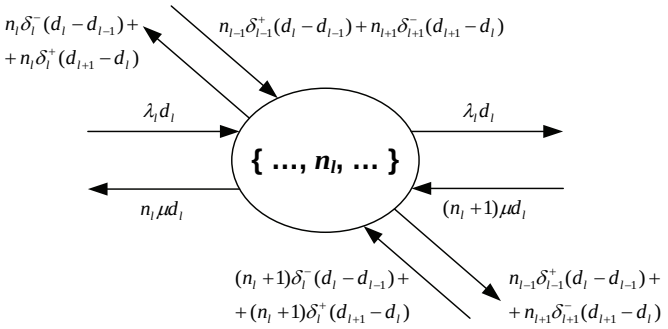


Fig. 2. System state transitions for ring l

For $l = L$

$$P_{\{n_L\}}[\lambda_L d_L + n_L \mu d_L + n_L \delta_L^+ d_L + n_L \delta_L^- (d_L - d_{L-1}) +$$

$$+ n_{L-1} \delta_{L-1}^+ (d_L - d_{L-1}) + \lambda_h d_L] = P_{\{n_L-1\}}[\lambda_L d_L +$$

$$+ n_{L-1} \delta_{L-1}^+ (d_L - d_{L-1}) + \lambda_h d_L] + P_{\{n_L+1\}}[(n_L + 1)\mu d_L +$$

$$+ (n_L + 1)\delta_L^+ d_L + (n_L + 1)\delta_L^- (d_L - d_{L-1})]. \qquad (4)$$

Where λ_h is the incoming inter-cell handover traffic from the other cells. At the same time λ_h is also outgoing inter-cell handover traffic to the outer cells:

$$\lambda_h = \lambda_L^{h+} = n_L \delta_L^+. \qquad (5)$$

For cellular networks [12], [13] the relation between fresh call traffic λ and handover cell traffic λ_h is equal to the relation between mean call holding time $1/\mu$ and mean cell dwell time $1/\delta$, that is

$$\frac{\lambda_h}{\lambda} = \frac{\delta}{\mu}. \qquad (6)$$

The radius r_l is the outer radius of ring l normalized to the cell radius. The ratio p_l of the ring l area to the entire cell area is given by

$$p_l = \frac{r_l^2 - r_{l-1}^2}{r_L^2}. \quad (7)$$

The radius r_l is actually the distance covered by MCS_l.

The value of p_l also represents the probability of an active MS being located in ring l.

The average dwell time in a cell is taken to be in direct proportion to the cell radius and in inverse proportion to the average MS speed [13], [14].

IV. Numerical Results

Following the assumption stated in Section III, the system parameters that were used in the analytical evaluation are as follows. The cell has a capacity of 6 resource units, which are shared among both new voice calls, arriving at each ring, and intra-cell handover calls when MSs move across inner boundaries (rings). We apply a 3-ring cell with the corresponding MCS (Fig. 1). The bandwidth, in terms of a number of resource units, required to maintain a voice call depends on the location of a MS within the cell area, and it is assumed $d_1 = 1$, $d_2 = 2$, $d_3 = 3$. The average service rate is assumed to be $\mu = 0.0167\ s^{-1}$, while the new voice call intensity λ_{MCS_l} in the cell area with the modulation scheme of highest order (64QAM) is varied from 0.001 to 0.01 s^{-1}. The intra-cell HO rates δ_l are mainly related to the MSs average speed, rings area and MSs distribution within the cell [13], [14]. We assume uniformly distributed MSs.

Typical spectral efficiency values for different MCS are shown on Table I [5]. We use the following relations between the distances covered by a particular MCS, depicted on Fig. 3 [15]. As a consequence, the following conclusions can be drawn:

- For constant coding rate (e.g. ¾) the relation between modulation and resources necessary for a given CBR service is

$$d_{64QAM} : d_{16QAM} : d_{QPSK} = 1 : 2 : 3. \quad (8)$$

- In case of the same coding rate, it follows that

$$r_{64QAM} : r_{16QAM} : r_{QPSK} \approx 2r : 3r : 4r. \quad (9)$$

According to (7), the following relation between the ring surfaces can be derived:

$$p_{64QAM} : p_{16QAM} : p_{QPSK} = 4r^2 : 5r^2 : 7r^2. \quad (10)$$

The relation between the ring's new incoming call rates is

$$\lambda_{64QAM} : \lambda_{16QAM} : \lambda_{QPSK} = 4r^2 : 5r^2 : 7r^2.$$

The performance measures under interest are new voice call blocking probability as well as handover dropping probability, which are depicted on Figs. 4 – 6, for different values of users' mobility.

TABLE I
MCS SPECTRAL EFFICIENCY (SE)

Modulation	Coding	SE (bps/Hz)
QPSK	1/2	0.7
QPSK	3/4	1
16 QAM	1/2	1.4
16 QAM	3/4	2.1
64 QAM	2/3	2.8
64 QAM	3/4	3.15

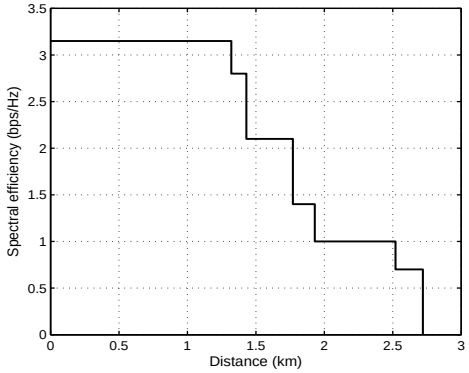


Fig. 3. Distance covered by a specific MCS in an OFDMA-based cell

A comparative study has been carried out for a case when MSs do not cross the inner boundaries - Intra-cell HO does not occur. The numerical results obtained are reasonable, especially for the MSs located in the Ring 3 area (Fig. 6), which are served by the most inefficient MCS (the most resource units per call are required). Thus, more new and HO calls exist in Ring 3, more call blocking/dropping probabilities are increased.

The performance metric studied include the total resource unit utilization under different traffic loads (including both new and HO arrivals) – Fig. 7. Results illustrate that the available resource can be better utilized, but there is a trade-off that the QoS parameters, such as new calls/HO blocking/dropping probabilities approach unacceptable levels.

On the other hand, since the wireless resource is scarce and expensive, a crucial task of further investigations is to guarantee the target QoS metrics while fully utilizing the scarce bandwidth. This is tightly coupled with the necessity of development of an appropriate resource management strategy.

V. Conclusion

An analytical model to study the intra-cell handover influence on system performance in terms of QoS metrics, such as call blocking/dropping probability, and resource unit utilization has been proposed. A comparative study for a case when intra-cell HO does not occur has been carried out. Some drawbacks have been identified and directions for future research have been drawn, which aim at looking for an optimal solution between target QoS measures and scarce resource utilization.

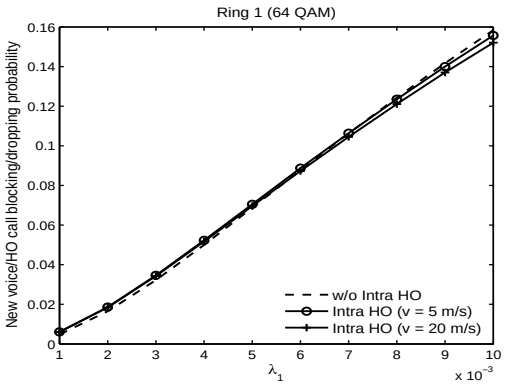


Fig. 4. Performance measures for traffic stream offered to Ring 1

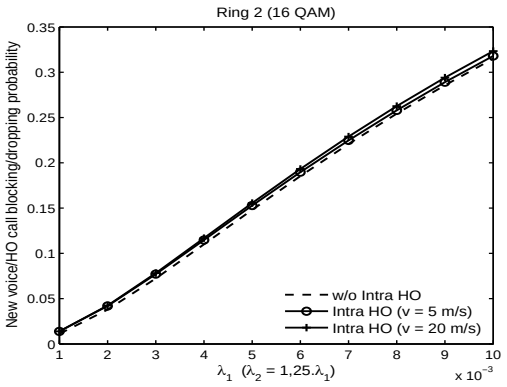


Fig. 5. Performance measures for traffic stream offered to Ring 2

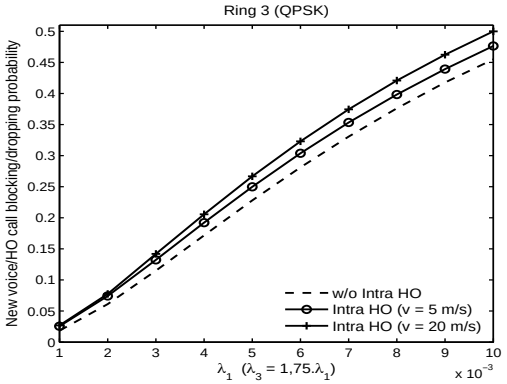


Fig. 6. Performance measures for traffic stream offered to Ring 3

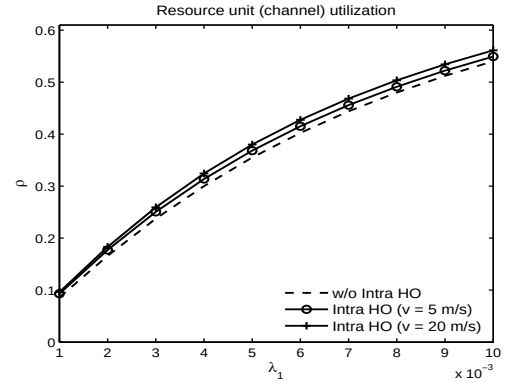


Fig. 7. Average utilization ρ of resource unit vs. traffic load

ACKNOWLEDGEMENT

This paper is a part of investigations, partially funded by the Bulgarian Ministry of Education, Youth and Science under grant DVU01/0109 (DO-02-135/2008), and Research and Development Sector of the Technical University of Sofia under grant 102ni191-7.

REFERENCES

[1] Physical Layer Aspects of UTRA High Speed Downlink Packet Access (Release 4), 3GPP TR 28.848 V4.0.0, 2001.

[2] IEEE Standard 802.16 Working Group, IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface fox Fixed Broadband Wireless Access Systems, 2002.

[3] M. S. Alouini, A. J. Goldsmith, "Adaptive Modulation over Nakagami Fading Channels", J. Wireless Communications, vol. 13, no.1-2, pp.119-143, May 2000.

[4] K. J. Hole, H. Holm, G. E. Oien, "Adaptive Multidimensional Coded Modulation over Flat Fading Channels", IEEE J. Sel. Areas Communications, vol. 18, no. 7, pp.1153-1158, July 2000.

[5] C. Tarhini, T. Chahed, "On capacity of OFDMA-based IEEE802.16 WiMAX including Adaptive Modulation and Coding (AMC) and Inter-cell Interference", 15th IEEE Workshop on Local & Metropolitan Area Networks, pp.139-144. 2007.

[6] L. Qingwen, Z. Shengli, B. Georgious, "Queuing with Adaptive Modulation and Coding over Wireless Links: Cross-layer Analysis and Design", IEEE Trans. Wireless Communications, vol. 4, no. 3, pp.1142-1153, May 2005.

[7] C. Tarhini, T. Chahed, "System Capacity in OFDMA-based WiMAX", ICSNC'06, Conference Proceedings, vol. 4, no. 3, pp.70-74, 2006.

[8] H. Wang, V. Iversen, "Erlang Capacity of Multi-class TDMA Systems with Adaptive Modulation and Coding", ICC'08, Conference Proceedings, pp.115-119, 2008.

[9] F. A. Cruz-Perez, G. Hernandes-Valdez, L. Ortigoza-Guerrero, "Performance Evaluation of Mobile Wireless Communication Systems with Link Adaptation", IEEE Communication Letters, vol. 7, no. 12, pp.587-589, Dec. 2003.

[10] F. Khan, D. Zeglache, "Performance Analysis of Link Adaptation in Wireless Personal Communication Systems", ICC'97, Conference Proceedings, pp.1287-1291, 1997.

[11] K. Samaras, C. Demeterscu, R. Yan, "Capacity of Packet Switched Cellular Network with Link Adaptation", ICC'2000, Conference Proceedings, pp.1568-1572, 2000.

[12] S. Nanda, "Teletraffic Models for Urban and Suburban Microcells: Cell Sizes and Handoff Rates", IEEE Trans. on Vehicular Technology, vol. 42, no. 4, pp.673-682, Nov. 1993.

[13] B. Jabbari, "Teletraffic Aspects of Evolving and Next-generation Wireless Communication Networks", IEEE Personal Communications, pp.4-9, Dec. 1996.

[14] R. A. Guerin, "Channel Occupancy Time Distribution in a Cellular Radio System", IEEE Trans. on Vehicular Technology, vol. 35, no. 3, pp.89-99, Aug. 1987.

[15] A. Kalaydzhieva, "Multi-hop Cell Design for WiMAX", Bachelor's degree diploma project, TU-Sofia, 2009, (in Bulgarian).

Novel Adaptive VoIP QoS Provisioning In Integrated UMTS/WLAN Networks

Tomislav Shuminoski[1] and Toni Janevski[2]

***Abstract* – This paper introduces a novel concept of adaptive QoS provisioning for wireless/mobile multimedia networks. Moreover, simulation results for this adaptive QoS framework in 3G networks (such as UMTS) and IEEE's 802.11 WLANs have been presented. The aim of this novel framework is presenting new module that shall provide the best QoS and lower cost for a given service using one or more wireless technologies. The analysis of this framework has shown good performances and high level of QoS provisioning in a variety of network conditions. The performance of our algorithm is evaluated using ns-miracle augmented with our dual-mode mobile stations.**

***Keywords* – Adaptive QoS module, dual-mode, jitter, PDR, Quality-of-Service (QoS).**

I. INTRODUCTION

With the explosive growth of the wireless and mobile communications and the emergence of the bandwidth-intensive multimedia services (video, data and voice), QoS provisioning for wireless and mobile multimedia networks is becoming increasingly important objective, since it requires great thoughtfulness, scalability and thoroughfull analysis. Since radio bandwidth is one of the most precious resources in wireless systems, an efficient adaptive QoS framework is very important to guarantee QoS and to maximize radio resource utilization simultaneously. Moreover, the most significant QoS parameters in the existing wireless/mobile networks are the throughput, packet delivery ratio, call blocking probability, delay and jitter (especially when we use Constant Bit Ratio (CBR) traffic flows).

The preview is focused on adaptive QoS provisioning for VoIP service over integrated UMTS and WLAN systems, in a tight coupling architecture, using novel dual-mode ME node with adaptive QoS module within. Integration of the WLAN and UMTS networks has been intensively studied recently due to their complementary characteristics. The 3GPP has been continuously evolving to support multimedia services which require high data rates in cellular networks. Today, the UMTS network can support services with maximum data rate of 2Mbps. On the other hand, the IEEE 802.11b can provide data rate up to 11Mb/s in 2.4 GHz. Furthermore, the IEEE 802.11a and IEEE 802.11g can provide up to 54 Mb/s in 5GHz and 2.4GHz bands, respectively. But WLANs have disadvantages of supporting nomadic user movements and of having small coverage (coverage by an access point (AP) is up to several hundred meters in radius) in comparison with a cell covered by a UMTS Node B (usually several kilometers in radius). Such complimentary characteristics of these two popular technologies have stimulated research efforts to integrate UMTS and WLAN networks so that mobile stations can choose the network that has better network quality when they are covered by both networks and have continuous services when they roam in the integrated networks. The hardware requirement for integrating UMTS and WLAN networks is mainly to build dual-mode ME with adaptive QoS module inside, which has the capability of accessing both networks and choose the best connection according to QoS demands for the given service (in our case VoIP).

Moreover, without loss of generality, this adaptive QoS provision can be used in any mobile and wireless IP multimedia networks. Nowadays many mobile equipments (MEs) have also a WLAN and Bluetooth interfaces, and in the near future many MEs will have Long Term Evolution (LTE) interfaces too, besides their UMTS, WLAN, WiMAX, Bluetooth, ZigBee etc. radio interfaces. However, when there are different wireless and mobile networks on one side, and single ME on the other, then consequently the user of that ME should have possibility to use all those technologies in the range using his/her personal settings in the ME, or this user can choose only one from all available technologies. For that purpose the Open Wireless Architecture (OWA) in [1] and in [2] is proposed to provide open baseband processing modules with open interface parameters for supporting different wireless communication standards.

The remainder of this article is structured as follows. In Section II gives an overview of the most relevant research works in this field. Section III briefly presents our ME adaptive QoS module. In Section IV the simulation results have been presented. And finally, the last Section V concludes this research.

II. RELATED WORKS

High level of QoS Provisioning for every given services is a major goal in any wireless and mobile networks, since the fact that the users are the main player from where all business revenues start. The interest for adaptive QoS provisioning is growing together with the tremendous grow of adaptive multimedia services in mobile communication networks, where it is possible to increase or decrease the bandwidth of individual ongoing flows. In [3] is presented bandwidth adaptation algorithm which seeks for high level of QoS

[1]Tomislav Shuminoski is teaching assistant at the Faculty of Electrical Engineering and Information Technologies, University "Sv. Kiril i Metodij", Karpos 2 bb, 1000 Skopje, Republic of Macedonia. E-mail:tomish@feit.ukim.edu.mk

[2]Prof. D-r Toni Janevski is Professor at the Faculty of Electrical Engineering and Information Technologies, University "Sv. Kiril i Metodij", Karpos 2 bb, 1000 Skopje, Republic of Macedonia. E-mail: tonij@feit.ukim.edu.mk

provisioning. Moreover, in [3] the bandwidth of an ongoing multimedia calls can be dynamically adjusted, with the CAC algorithm. Call blocking probability, forced termination probability and call overload probability are the main QoS parameters take in common. But the lack of [3] lays in the fact that only one single class of adaptive multimedia networking has been investigated. Furthermore, in [4] effective QoS provisioning for wireless adaptive multimedia with using a form of discounted reward reinforcement learning known as *Q*-learning is presented. Moreover, the proposed scheme in [4] considered the handoff dropping probability and average allocated bandwidth constraints simultaneously, in order to achieve optimal CAC and BA policies that can maximize network revenue and guarantee QoS constraints. Simulation results in [4] demonstrate that the given scheme is high effective. On the other hand, when we focus on architectures for integrating WLAN/UMTS systems they can be grouped into two categories based on the independence between the two networks [5], tight coupling and loose coupling. Here the loose coupling architecture is used; where those two networks are integrated beyond the Core Network (CN) of UMTS (connected through gateways of the Internet). The loose coupling architecture enables the two networks deployed independently but results in longer delay for signaling and vertical handovers. 3GPP has been working on standardization for integrating cellular and WLAN systems in [6] in which interworking architecture and interworking scenarios are described. Moreover, similar to our dual-mode ME node for UMTS/WLAN interworking networks, have been proposed in [7] and [8], but without emphasized QoS issues. The main motivation that led us to the development of novel adaptive QoS module, which will provide intelligent high level of QoS in any wireless and mobile networks (including integrated UMTS/WLAN networks), using every available technology in a same time, is taken from [2]. In [2] is given the main 5G mobile phone concepts and moreover the needs for creating and implementing adaptive QoS management mechanisms have been introduced. To emphasize that in comparison with other related works, our adaptive QoS module is implemented on IP level.

III. Adaptive QoS System Module

In Fig. 1 is presented our novel ME node, which is dual-mode UMTS/WLAN node, with Adaptive QoS module within on IP layer. According to [1] and [2] physical and OWA define the wireless technology (i.e. OSI Layer 1 and 2), the network layer will be IP, but separation of this layer into two sublayers will be necessary; one subleayer for routing and another will be for each interface (different IPv4 or IPv6 addresses). For more details about all this layers see [2]. Furthermore, briefly is presented our adaptive QoS framework in ME. The core of our work is development of novel adaptive QoS Module; we call it QoS-Cross-IP Module (QXIP), defined separately from each wireless technology (e.g. UMTS, WLAN, WiMAX, 3G-LTE, 4G, etc.), which will be able to provide intelligent QoS management and routing over variety of network technologies.

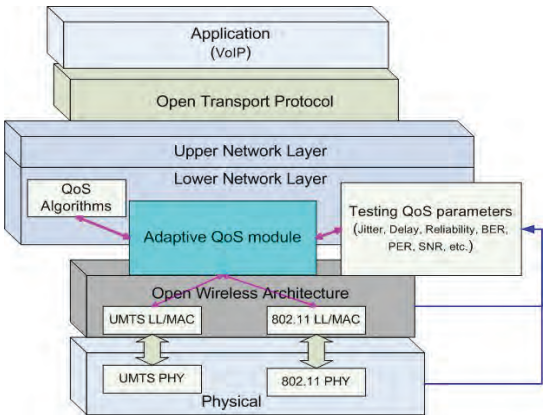


Fig. 1. ME dual-mode Node with Adaptive QoS Module

The QoS parameters (such as delay, jitter, losses, bandwidth, reliability, Packer-Error-Ratio (PER), Signal-to-Noice-Ratio (SNR), Transmission Power (TP), etc.) are continuously, all the time, collecting via cross-layer messages from OSI layer 1 up to Lower Network layer, and stored in a two-dimensional matrix within the QXIP module. The first row of this matrix contains UMTS QoS parameters and second row contains the WLAN QoS parameters, appropriately. Before every transmission of IP packet from QXIP down to UMTS or WLAN LL/MAC modules, the QXIP module is doing service quality testing (testing of collected QoS parameters) in order to choose the best wireless connection upon required QoS. Here, in our current implementation, we are testing only: SNR, PER and TP, which we collected from Physical layer and OWA layer via cross-layer messages. According to the performance analysis in [9], QXIP module always fist try to get admission (in uplink) to the WLAN whenever it is available (i.e. all tested WLAN parameters: SNR, PER and TP, are above their appropriate WLAN thresholds). Second, if QXIP module doesn't get WLAN admission, he try to get admission to UMTS network (all tested UMTS parameters are above their appropriate UMTS thresholds). Finally, QXIP module sent the packet that come from Upper Network Layer down to the chosen LL/MAC module or dropped if he doesn't find admission to any of those two networks. In downlink, all packets from all LL/MAC modules are received and send up from QXIP to Upper Network Layer, without dropping them.

IV. Simulation Results

In Fig.2 the simulation scenario is given, and as can be seen, we create one UMTS Node B and one WLAN Access Point. At the beginning of the simulation, the MEs are randomly scattered within the area of 500x500 m^2. For MEs physical mobility, we adopted the Gauss-Markov Mobility model [10] considering average speeds in the range of 2-21 m/s. The Node B coordinates are: (500,500) which providing coverage for the MEs placed within a distance of about 520 m. On the other hand, WLAN AP is placed at (150,150) which providing coverage for the MEs placed within a distance of about 130 m. The VoIP traffic (CBR traffic) starts at the beginning of the simulation time, and flows between

Internet through the VoIP gateway, which is wired to Node B and AP, to all MEs, until the end of the simulation time. The general parameters used in our simulation are summarized in Table I. At the first case all MEs are dual-mode UMTS/WLAN, and are equipped with QXIP module. We use ns-miracle 1.2.2 [11] for creating our dual-mode ME with two interfaces (one for UMTS, another for WLAN network) and with QXIP module (presented in the previous section). Moreover we create novel cross-layer messaging class, for cross-layer communication between the modules. The performance outline in this case is shown with blue lines in Fig. 3-8, and is compared with the simulation outlines in the cases when we have MEs without QXIP module within, and only with WLAN or with UMTS interface.

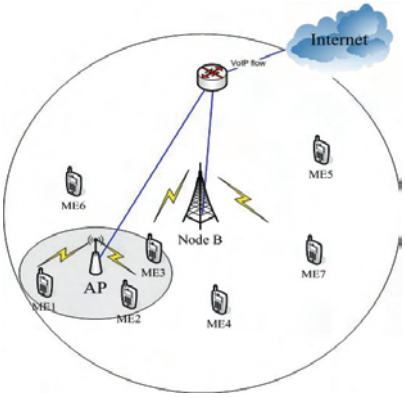


Fig. 2. Simulation scenario

TABLE I
SIMULATION PARAMETERS

Parameters	Values
CBR packet size	160 Bytes
WLAN Data rate	1 Mbps
Phy header	192 bits
MAC header	224 bits
SIFS	10 μs
DIFS	50 μs
Traffic frame interarrivel time	4 seconds
CTS, ACK	112 bits + Phy header
WLAN_PER threshold	$7x10^{-11}$ W
UMTS_PER threshold	10^{-6} W
NodeB spreading factor	32
ME spreading factor	16

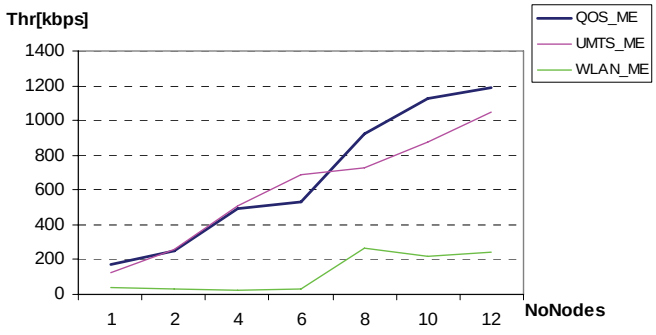


Fig. 3. Average throughput vs number of nodes ($\bar{v}$ =2m/s)

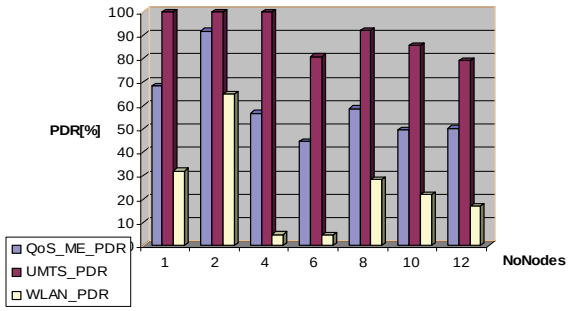


Fig. 4. Average PDR vs number of nodes ($\bar{v}$ =2m/s)

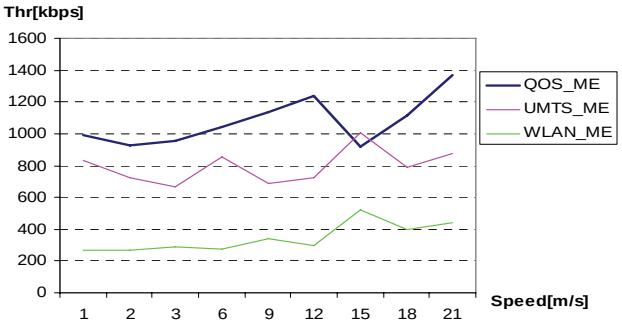


Fig. 5. Average throughput vs velocity (N^0Nodes=8, t_sim=10s)

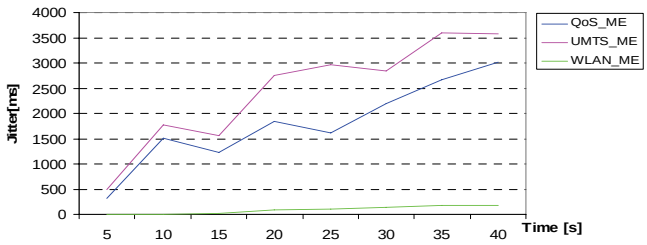


Fig. 6. Average jitter vs simulation time (N^0Nodes=8, $\bar{v}$ =2m/s)

Fig. 7. Average delay vs velocity (N^0Nodes=8, t_sim=10s)

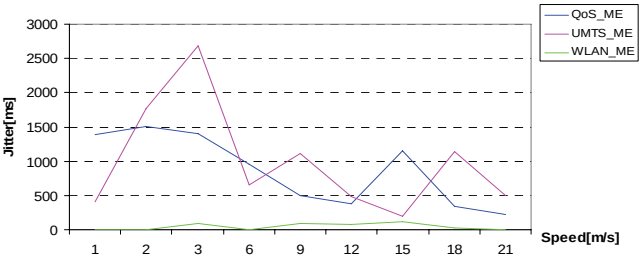


Fig. 8. Average jitter vs velocity (N^0Nodes=8, t_sim=10s)

As we can obtain from Fig. 3, the average throughput for our dual-mode ME with included QXIP module within (QOS_ME), without doubts achieve the greatest values for every number of used nodes, in comparison with the average throughput in the case when we used only WLAN MEs, and equal or greater average throughput values with the case when we use only UMTS MEs. We must to emphasize the fact that we obtain equal (or less) average throughput with the case when we use only UMTS MEs only when we use low number of nodes (less then 7 MEs). This is, more or less, expected due to the fact that when we used just a few MEs with mean speed of $\overline{v}$ =2m/s and the probability some of them to pass through the WLAN area (in simulation time of 10s) is very low, so they achieve only VoIP traffic via Node B.

Moreover, as we see in Fig. 4, the average Packet Delivery Ratio (PDR) values for the first case when we use QXIP dula-mode MEs are very balanced, i.e. our PDR values are somewhere between the PDRs from UMTS and WLAN MEs for every used number of MEs (nodes), and have tendency at some points to reach UMTS ME PDR values. Furthermore, in Fig. 5 is presented average throughput for different mean velocity (form 2 up to 21 m/s) of the MEs, when we have 8 MEs and simulation time of 10 s. First of all, can be clearly obtain that average throughputs of our case (when we used dual-mode QXIP MEs) is much larger then the average throughput from the cases when we used WLAN or UMTS MEs, for any given average velocity ($\overline{v}$). To emphasize that, when we using our dual-mode QXIP MEs, the average throughput curve has ascending trend due to the fact that with higher average speed of the MEs, more frequently they are passing through the WLAN area and consequently achieves greater average throughput.

Furthermore, in Fig. 6 average jitter values as a function of simulation time for any three cases have been presented. In this scenario, the number of MEs is fixed on 8, and the MEs mean velocity is 2 m/s. As we can see, for our case when we used dual-mode QXIP MEs, in comparison with the case when we use UMTS MEs, lover average jitter values have been achieved, with the similar curve trends. On the other hand we have higher average jitter values in comparison with the average jitter values of WLAN MEs, but this is because of the fact that in WLAN technology we have lower processing time and lower area coverage.

Moreover, in Fig. 7 the average delays for all three cases are presented. In this scenario, the number of MEs is fixed on 8, simulation duration is 10s and we change the MEs mean velocity from 2 m/s up to 21 m/s. As we can see, the statistic of our dual-stack QXIP ME delay curve is between the curves for UMTS and WLAN cases, with lower values then the values of UMTS curve (only for lower average velocity is the same), and higher then the values of WLAN curve. Consequently, with rise of the mean speeds of the dual-stack QXIP MEs their delay curve is descending, because in cases with higher mean speed, more frequently the MEs are entering in the WLAN area, and due to the lower latency of WLAN network, all MEs achieve lower latency values.

Finally, in Fig. 8 are shown the average jitters curves for the same scenario as described before. Similar as previously, the statistic of our dual-stack QXIP ME jitter curve is between the jitter's curves of UMTS and of WLAN cases, with lower values then the values of UMTS curve, and higher then the values of WLAN curve. With rise of the mean speeds of the MEs all three curves are descending. The reason is the same we discussed above about the latency.

V. CONCLUSION

In this paper a novel simulation results for the key QoS parameters (throughput, jitter, delay and PDR), using novel adaptive QoS module and dual-mode UMTS/WLAN ME, have been presented.

According to the simulation results, our propose dual-stack UMTS/WLAN ME with adaptive QoS module within perform fairly well, even in different network conditions as: various nodal mobility, various background traffic loads and different number of nodes; achieving really well performances in comparison with the cases when only WLAN or only UMTS MEs have been used.

In our future work we will focus on development advance QXIP module, with more simulation results in various network conditions and with more complex scenarios, follow by profound statistical analysis.

REFERENCES

[1] Willie W. Lu, "An Open Baseband Processing Architecture for Future Mobile Terminals Design", IEEE Wireless Communications, April 2008.

[2] Toni Janevski, "5G Mobile Phone Concept" – CCNC conference in Las Vegas, 2009.

[3] T. Kwon, Y. Choi, C. Bisdikian andM. Naghshineh, "QoS provisioning in wireless/mobile multimedia networks using an adaptive framework", Wireless Networks 9 (2003) 51–59.

[4] Yu Fei, Vincent W.S. Wong, Victor C.M. Leung, "Efficient QoS Provisioning for Adaptive Multimedia in Mobile Communication Networks by Reinforcement Learning", Mobile Networks and Applications 11 (2006) 101–110.

[5] M. Buddhikot, G. Chandranmenon, S. Han, Y. W. Lee, S. Miller and L. Salgarelli, Integration of 802.11 and Third-Generation Wireless Data Networks, Proc. IEEE INFOCOM, Apr. 2003.

[6] 3GPP, Group Services and System Aspects; 3GPP Systems to Wireless Local Area Network (WLAN) Interworking; System Description (Release 6), TS 23.234, v. 6.2.0, Sept. 2004.

[7] Y. Zhou, Y. Rong, H.-A. Choi, J.-H. Kim, J. K. Sohn, and H. I. Choi,"A Dual-Mode Mobile Station Modules for WLAN/UMTS Internetworking Systems," Proc. OPNETWORK 2007, August 27-31, 2007, Washington, DC.

[8] Nicola Baldo, Federico Maguolo, Marco Miozzo, Michele Rossi, Michele Zorzi, "Ns2-MIRACLE:a Modular Framework for Multi-Technology and Cross-Layer Support in Network Simulator 2", *NSTools* '07, October 22, 2007, Nantes, France.

[9] Wei Song, Hai Jiang, Weihua Zhuang "Performance analysis of the WLAN-First scheme in Cellular/WLAN interworking", IEEE transactions on wireless communications, Vol.6, Issue 5, page(s): 1932-1952, May, 2007.

[10] B. Liang and Z. Haas, "Predictive distance-based mobility management for PCS networks", in Proceedings of IEEE Infocom, New York, NY, US, Mar. 1999.

[11] Ns-MIRACLE:Multi InteRfAce Cross Layer Extension for ns-2, http://www.dei.unipd.it/wdyn/?sez_alias=ricerca/signet/tools/nsmiracle

SESSION TST II

Telecommunications Systems and Technology II

Determining Bit Error Rate in Digital Optical Transmission Network Using the Q-Factor

Kiril K. Koitchev[1], Krasen K. Angelov[2], Stanimir S. Sadinov[3]

***Abstract* – Knowledge of the ratio of the signal power to the noise power (signal-to-noise ratio – SNR) is important because it is directly related to the bit error ratio (BER) in digital communication systems, and the BER is a major indicator of the quality of the overall system. The SNR ratio related to the optical communications has its own specifics.**

The purpose of this paper is to show the relationship between the electrical and optical signal-to-noise ratio introducing the Q-factor. Although some of the principles can be applied to the existing coaxial cables, the scope of the paper is limited to binary digital signals over optical fiber. Application of codecs based on Read-Solomon codes will be considered as an example, where using this error correction method can reduce SNR in the receiver keeping the same value of BER.

***Keywords* – Bit Error Rate, Q-factor, Optical Signal-to-Noise Ratio, Probability Density Function, block codes**

I. INTRODUCTION

So far *BER* estimation of characteristics in digital transmission systems, that use quadrature modulations, is important for studying especially in recording *SNR*. This ratio if related to optical communications is relatively sparsely represented and has its specific features [1].

The aim of this paper is to disclose the relationship between electrical and optical signal-to-noise ratio *SNR* by means of the relating parameter Q-factor. Although some of the principles could be applied in the coaxial cables, the particular case of this study will regard the transfer of binary digital signal along an optical fiber because in this way both bit and symbol rates will be equivalent. Application of codecs based on Read-Solomon codes will be considered as an example, since by this method of error correction it is possible to lower *SNR* on the receiving side with unchanged *BER*.

II. CALCULATION OF ERROR PROBABILITY

As is known the power of the optical signal is presented as the product of electric and magnetic field vectors:

$$P_O(t)=\left|\vec{E}(t)\times\vec{H}(t)\right|=\left|\frac{\vec{E}(t)}{\eta}\right|\cdot\left|\vec{E}(t)\right|=\left|\frac{\vec{E}(t)}{\eta}\right|^2 . \quad (1)$$

In practice direct measuring of optical power is rarely the case. The usual practice involves the use of PIN photodiode whose current is proportional to the optical power [5].

A point of interest is the studying of photo-receiver's performance. By synchronizing with the incoming optical signal the photo-receiver performs periodic processing of the received signal at the optimum moment. It registers the optical signal intensity and finally decides, on the basis of a certain threshold value, what type of signal is received: 0 or 1. Quality of the digital communication system performance could be evaluated by way of the so called eye-diagram (Fig. 1).

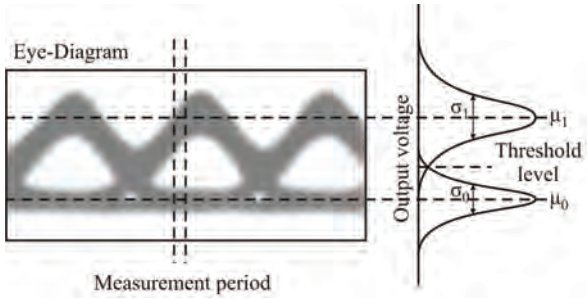


Fig. 1. Typical appearance of eye-diagram while using a code of the type RZ (return to zero) in optical signal receiving

Let us assume that in the optical receiver the converted voltage *v(t)* is compared with one definite value γ, which is referred to as threshold level. When *v(t)* is greater than γ this means that a binary "1" has been sent. When *v(t)* is less than γ – a binary "0" has been sent [1,2].

In presence of white Gaussian noise (AWGN), it is possible to calculate the statistical error probability. The probable density of *v(t)* in presence of AWGN can be mathematically expressed using Gaussian probability density function (*PDF*) [9,10,11]:

$$PROB[v(t),\sigma_x]=\frac{1}{\sqrt{2\pi\sigma_x^2}}e^{-\frac{1}{2}\left(\frac{v(t)-v(s)}{\sigma_x}\right)^2}, \quad (2)$$

where *v(s)* is the mean value of the density function of the receiver, *v(t)* is the moment value of the receiver's voltage, and σ is the standard noise aberration. The graphic interpretation of Eq. 2 is shown in Fig. 2.

[1]Kiril Koitchev is with the Faculty of Electrical Engineering and Electronics, Technical University – Gabrovo, 4 H. Dimitar St., 5300 Gabrovo, Bulgaria, E-mail: koitchev@tugab.bg

[2]Krasen Angelov is with the Faculty of Electrical Engineering and Electronics, Technical University – Gabrovo, 4 H. Dimitar St., 5300 Gabrovo, Bulgaria, E-mail: kkangelov@dir.bg

[3]Stanimir Sadinov is with the Faculty of Electrical Engineering and Electronics, Technical University – Gabrovo, 4 H. Dimitar St., 5300 Gabrovo, Bulgaria, E-mail: murry@tugab.bg

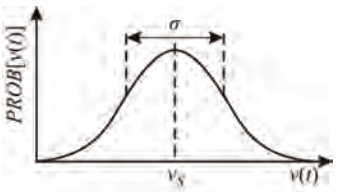


Fig. 2. Gaussian probability density function

If we assume that *v(s)* could take one of the two voltage levels designated as v_L and v_H then the probability for wrong decision making in the receiver is:

$$P[\varepsilon]=P[v(t)>\gamma\,|\,v_S=v_L]P[v_S=v_L]+P[v(t)<\gamma\,|\,v_S=v_H],\quad(3)$$

where $P[\varepsilon]$ is the error probability, $P[x,y]$ is the conditional probability of x for assigned value of y. Suppose that transmission and receiving probability is *50%* then $P[v_S=v_L] = P[v_S=v_H] = 0{,}5$. In this case Eq. 3 is reduced to:

$$P[\varepsilon]=P[v(t)>\gamma\,|\,v_S=v_L]\times 0.5+P[v(t)<\gamma\,|\,v_S=v_H]\times 0.5= \\ =\frac{1}{2}\int_{-\infty}^{\gamma}PROB[v(t),\sigma_L]dt+\frac{1}{2}\int_{\gamma}^{\infty}PROB[v(t),\sigma_H]dt \quad,(4)$$

where $PROB[v(t),\sigma_x]$ is determined by Eq. 2. This interpretation is graphically presented in Fig. 3.

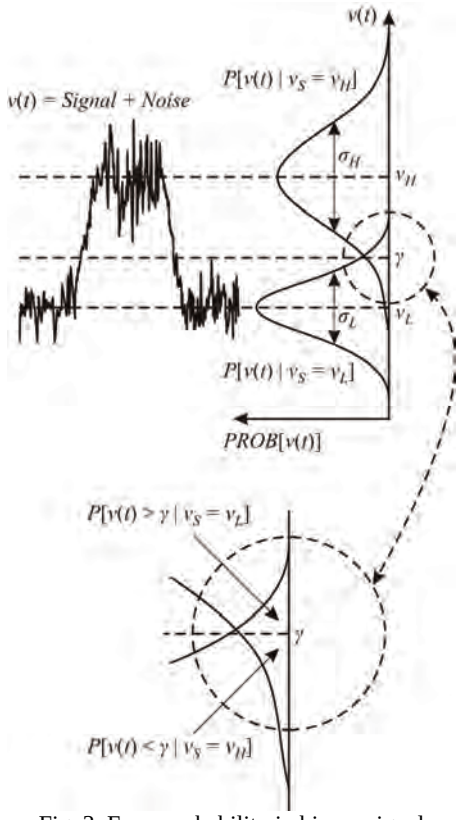


Fig. 3. Error probability in binary signal

Considering Fig. 3 and Eqs. 3 and 4 it is possible to reach the conclusion that the error probability is the area of γ beneath the graphs of the probability density functions. This area determines the magnitude of *BER* which depends on two factors:

- on the standard noise aberrations (σ_L and σ_H);
- on the difference between voltages v_L and v_H.

When $\sigma_L = \sigma_H$, then $\gamma=(v_H-v_L)/2$, i.e. γ is the average value between the high and low level. In most cases $\sigma_L\neq\sigma_H$. Then the optimum value of *BER* will be larger or less than the value of $(v_H-v_L)/2$.

To solve Eq. 4 it is necessary to calculate the result of integrating Gaussian distribution *PDF* $(PROB[v(t),\sigma_x])$ which is defined through Eq. 2. This is possible by employing numerical methods. For this purpose Eq. 2 can be solved by setting it into standard form and applying $z=(x-\mu)/\sigma$ where $x=v(t)$ and $\mu=v_S$.

For the sake of facilitation let us apply this substitution for the error function taking into account Eqs. 2 and 4,

$$\int_{\gamma}^{\infty}PROB[x,\sigma]dx=\int_{\gamma}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx\ ,$$

and by applying $z=(x-\mu)/\sigma$ (where $x=z\sigma-\mu$ and $dx=\sigma\,dz$) we obtain:

$$\frac{1}{\sqrt{2\pi}\sigma}.\int_{x=z\sigma+\mu=\gamma}^{\infty}e^{\left(\frac{-z^2}{2}\right)}\sigma\,dz\ .$$

This result defines the error function $E_r(z)$

$$E_z=\frac{1}{2\pi}\int_{z=\gamma}^{\infty}e^{\left(\frac{-z^2}{2}\right)}dz\ . \quad(5)$$

In this way Eq. 4 can be represent as

$$P[\varepsilon]=\frac{1}{2}E_r\left[\frac{v_H-\gamma}{\sigma_H}\right]+\frac{1}{2}E_r\left[\frac{\gamma-v_L}{\sigma_L}\right]. \quad(6)$$

III. ESTIMATING THE Q- FACTOR BY MEANS OF THE ERROR FUNCTION

The argument in the error function from Eq. 6 is the square root of the signal divided by the square root of the signal level (see Eq. 11). This is equivalent to the optical signal-to-noise ratio and could be written as follows:

$$P[\varepsilon]=\frac{1}{2}E_r\left[SNR_{OH}\right]+\frac{1}{2}E_r\left[SNR_{OL}\right], \quad(7)$$

where SNR_{OH} and SNR_{OL} are optical signal-to-noise ratios for the high and low level respectively.

Optimum initial level γ_{opt} is determined by this initial level which defines the least bit error probability. In addition when determining optimum initial level it is important if high signal levels are transmitted as low level [3,4]. This means that γ_{opt} is obtained when $SNR_{OH} = SNR_{OL}$ which results in the following representation of Q-factor:

$$Q \equiv \frac{v_H - \gamma_{opt}}{\sigma_H} = \frac{\gamma_{opt} - v_L}{\sigma_L}. \quad (8)$$

If Eq. 8 is substituted in Eq. 6 taking in consideration that $\gamma = \gamma_{opt}$

$$P[\varepsilon] = \frac{1}{2} E_r[Q] + \frac{1}{2} E_r[Q] = E_r[Q]. \quad (9)$$

By taking the value of γ_{opt} from Eq. 8 we obtain:

$$\gamma_{opt} = \frac{v_H \sigma_L + v_L \sigma_H}{\sigma_L + \sigma_H}. \quad (10)$$

Considering Eq. 10 from Eq. 8 we obatin:

$$Q = \frac{v_H - v_L}{\sigma_L + \sigma_H}. \quad (11)$$

If both sides of Eq. 11 are multiplied by resistance or impedance then the expression will be converted into equivalent expressions which contain current or optical power:

$$Q = \frac{i_H - i_L}{\sigma_L + \sigma_H} = \frac{P_{OH} - P_{OL}}{\sigma_L + \sigma_H}. \quad (12)$$

Finally, by substituting Eq. 11 in Eq. 9 the outcome will be:

$$P[\varepsilon] = E_r[Q] = E_r\left[\frac{v_H - v_L}{\sigma_L + \sigma_H}\right]. \quad (13)$$

On the other hand *BER* could be logically represented as a function of Q-factor and by using error function.

If the necessary *BER* ratio is assigned, it is possible to obtain the corresponding value for the Q-factor by using the following expression:

$$BER(Q) = \frac{1}{2} erfc\left(\frac{Q}{\sqrt{2}}\right) \approx \frac{\exp\left(-Q^2/2\right)}{Q\sqrt{2\pi}}. \quad (14)$$

For example, for *BER = 1,0*$^{-11}$ it is necessary to ensure a value of *Q = 16,53*; for *BER = 4,15.10*$^{-6}$ is obtained *Q = 13,00*.

IV. Results

The FEC combined with the use of appropriate code will strongly raise the quality of performance of the communication line and, in particular, allows keeping the same value of *BER* at lower value of the *SNR* ratio, i.e longer regeneration section.

The error correction device is usually a basic element of the modern fiber optic data transmission systems [7,8,12]. There are various methods of implementation of FEC-codecs. For example, the application of network standards Ethernet or Gigabit Ethernet in optical interface features increased bit rate by 25% and the concerned encoding circuits are referred to as 4B/5B and 8B/10B. The 10 Gigabit Ethernet standard uses two types of encoding: 64B/66B for transmission along single mode fiber and 8B/10B for transmission along multimode fiber.

In optical communication lines which contain linear optical amplifiers, the codecs based on Reed-Solomon block codes of RS(n,k) type with *s*-bit symbols and in particular RS(255,251), RS(255,239) and RS(255,223) codes with single-byte (*s = 8*) symbols are widely distributed [3,4].

In communication lines, the so called "out-of-band" standard FEC ITU G.975 is the most widely used code which is based on Reed-Solomon code RS(255,239). FEC increases the bit rate from *9,95* to *10,66 Gbps* and allows decreasing of *BER* from *10*$^{-5}$ to *10*$^{-15}$ and improving the *SNR* with *6dB*.

It is a conventional practice to measure the efficiency of FEC correction according to *SNR* change; in other words how much does the error correction method allow to reduce *SNR* on the receiving side keeping the previous value of *BER*.

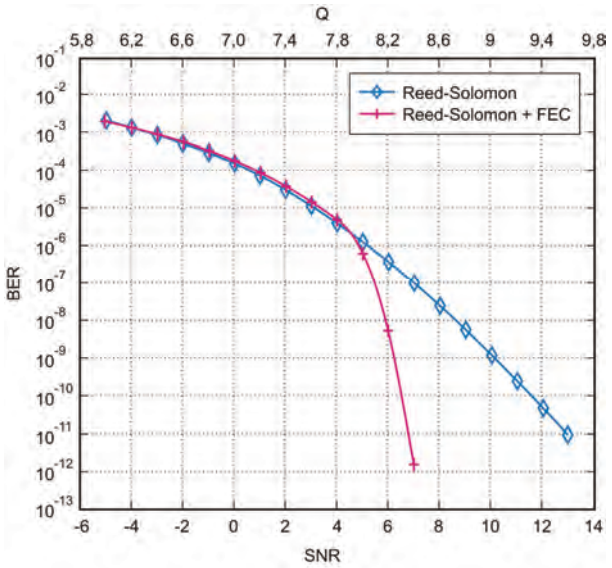


Fig. 4. *BER* parameters for BPSK modulation and addition of FEC to Reed-Solomon code

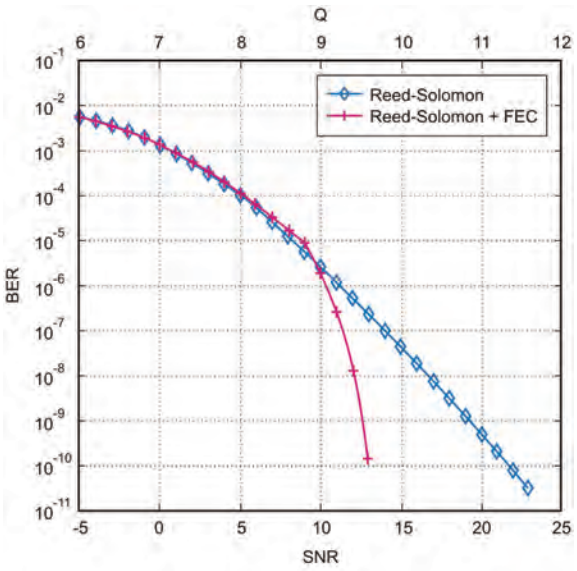


Fig. 5. *BER* parameters for QPSK modulation and addition of FEC to Reed-Solomon code

In Figs. 4 and 5 are presented the results from digital data transfer using Reed-Solomon encoding. It is evident that information transfer is improved when using FEC combined with the Reed-Solomon code while using QPSK and BPSK modulations.

Obtained results indicate that by selection of appropriate Q-factor line's length can be extended, keeping all previous parameters such as the input power, amplification coefficient and amplifiers' noise coefficient and the distance between two contiguous amplifiers.

V. CONCLUSION

The Q-factor estimated by Eqs. 8 and 11 represents optical signal-to-noise ratio for binary optical communication systems. In this way it incorporates individual *SNR* ratios related to high and low levels into a complete system of *SNR* ratios.

The Q-factor which is represented by Eq. 11 facilitates both measurement of *SNR* ratio and the theoretical calculation of *BER* in presence of noise in the system.

It is evident that, *BER* can be improved either by increasing the difference between high and low levels in the nominator or by reducing the noise in the denominator of the Q-factor equation.

In this way, using the Q-factor, it is possible to perform a simplified analysis of the communication system efficiency.

By using appropriate code and error correction it is possible to increase the digital data transfer efficiency along optical communication channels. This efficiency is determined by the margin of reduced *SNR* ratio while keeping the previous value of *BER*.

REFERENCES

[1] K. Vitthaladevuni, M.S. Alouini, BER Computation of 4/M-QAM Hiearchical Constellations. IEEE Transactions of Broadcasting, Vol.47, №3, pp.228-239, 2001.

[2] J. Lu, K. Letaief, C. Chuang. M-PSK and M-QAM BER Computation Using Signal – Space Concepts. IEEE Transaction of Communication, Vol.47, №2, pp 181-184, 1999.

[3] K. Nagayama, T. Saitoh. Ultra Low Loss (10,151 dB/km) Fiber and its Impact on Submarine Transmission Systems. OFC USA, Post-didline Papers FA10, 2002.

[4] M. Trukitani, M. Matsui. Ultra Low Nonlinearity Pure-Silica-Core Fiber with an Effective Area of 211mm^2 and Transmission Loss of 0,159 dB/km. ECOC Copenhagen, SS.6.3, 2002.

[5] B. Sklar, Digital Communications: Fundamentals and Applications, Englewood Cliffs, New Jersey: Prentice Hall, pp. 741-743, 2000.

[6] G. Agrawal, Fiber-Optic Communication Systems, New York, John Wiley & Sons, pp. 172, 1997.

[7] N. Bergano, F. Kerfoot, C. Davidson, Margin Measurements in Optical Amplifier Systems, IEEE Photonics Technology Letters, Vol.5, №3, pp. 304-306, 1993.

[8] G. Frank, Corning Inc., Backbone to the future „Traditionally, long-haul networks served as the proving ground for technologies that migrated to other parts of the network", Lightwave, September 2001.

[9] R. Ramaswami, K. Sivarajan, Optical Networks: A Practical Perspective, Academic Press, 1998.

[10] Optical Fiber Telecommunications IV-B: Systems and Impairments (Optics and Photonics), Academic Press, 2002.

[11] Р. Убайдулаев, Протяженные ВОЛС на основе EDFA, Lightwave Russian edition, №1, стр26-28, 2003.

[12] О. Панагиев, Изследване влиянието на параметрите на HFC комуникационна мрежа върху BER. НКМУ "ТЕЛЕКОМ 2009", Варна, стр. 236-241, 2009.

Estimating the Number of Amplifying Sections in Hybrid Fiber-Coaxial Television Network Using an Iterative Approach

Krasen K. Angelov[1], Kiril K. Koitchev[2],
Nataliya A. Varbanova[3], Stanimir S. Sadinov[4]

***Abstract* – The number of amplifying sections of hybrid fiber-coaxial (HFC) network is causally related to amplitude losses, dispersive expansion of code pulses and nonlinear distortions. Since all parameters are correlated, an iterative approach for estimating the number of amplifying sections in an optical trunk line is proposed. Total signal losses are estimated and through series of iterations a solution is found which for a given number of amplifying sections for the current iteration will yield minimum relative percentage error.**

***Keywords* – optical trunk line, optical amplifier, signal level, nonlinear distortion**

I. INTRODUCTION

The modern hybrid cable TV networks use as backbone network high speed optical trunk-lines. The power potential of the system is used to determine the optimum number of amplifying sections in certain optical trunk-line. It is equal to the difference between the maximum possible signal level and the signal losses in the transmission line and termination devices. These are amplitude losses, the losses due to transient and intersymbol distortions plus additive noise losses. Proposal of a practical approach for determining the number of amplifying sections depending on the physical parameters of the transmission line and used optical equipment has been the aim of this work.

The main factors are the physical parameters of optical amplifiers and optical transmission and receiver module. [2,4]. A transmission line can be rendered as consecutively connected amplifying sections [5] each of which can be modeled as per the block diagram in Fig. 1.

In Fig. 2 and 3 are shown substitution models of optical transmission and receiver modules.

[1]Krasen Angelov is with the Faculty of Electrical Engineering and Electronics, Technical University – Gabrovo, 4 H. Dimitar St., 5300 Gabrovo, Bulgaria, E-mail: kkangelov@dir.bg

[2]Kiril Koitchev is with the Faculty of Electrical Engineering and Electronics, Technical University – Gabrovo, 4 H. Dimitar St., 5300 Gabrovo, Bulgaria, E-mail: koitchev@tugab.bg

[3]Nataliya Varbanova is with the Faculty of Electrical Engineering and Electronics, Technical University – Gabrovo, 4 H. Dimitar St., 5300 Gabrovo, Bulgaria, E-mail: nataliavarbanova@abv.bg

[4]Stanimir Sadinov is with the Faculty of Electrical Engineering and Electronics, Technical University – Gabrovo, 4 H. Dimitar St., 5300 Gabrovo, Bulgaria, E-mail: murry@tugab.bg

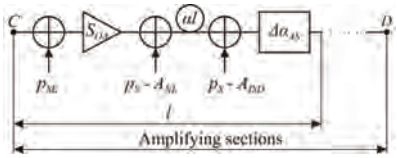


Fig. 1. Structural diagram of amplifying sections

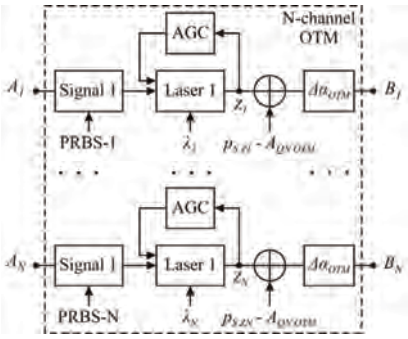


Fig. 2. Structural diagram of optical transmission module

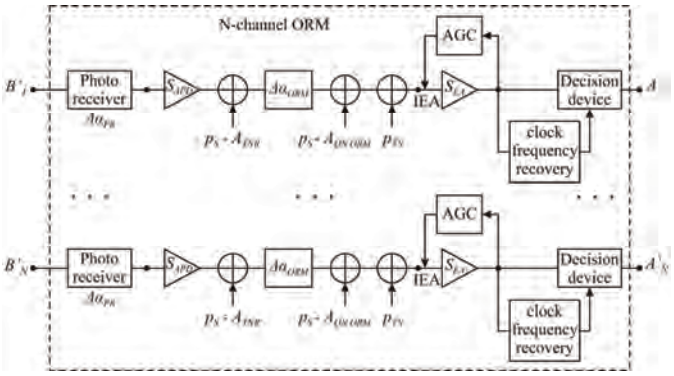


Fig. 3. Structural diagram of optical reception module

The basic parameters characterizing the optical transmission module (OTM) are as follows:

– absolute signal level at the laser output in point Z_i: $p_{S\,Zi}$;
– additional amplitude losses due to inaccuracy of AGC, degradation of laser in time and its thermal instability, the type of linear code and energy losses during its input in the fiber: $\Delta\alpha_{OTM}$;
– signal protectability from laser quantum noise: $A_{QN\,OTM}$.

The basic parameters characterizing optical receiver module (ORM) are:

– attenuation of photo-receiver: α_{PR};
– amplification of avalanche photo-diode S_{APD};

– signal protection from photo-receiver quantum noise: $A_{QN\ ORM}$;
– noise figure of electronic amplifier (EA): *NF*;
– additional signal amplitude losses due to inaccuracy of AGC, instability of the decision device threshold level due to phase distortions of the clock frequency divider, and due to the energy losses at the optical receiver module input plus other factors: $\Delta\alpha_{ORM}$;
– signal protectability from intersymbol distortions generated by the limited frequency band of the signal of the electronic amplifier: $A_{ISD}(0)$.

Optimum number of amplifying sections in the optical trunk lines of HFC network can be determined by means of iterative approach. By applying iterative cycle it is possible to search for such a solution which, given the number of amplifying sections m_j for the current iteration, will yield relative percentage error that satisfies the inequality

$$\delta_{Cj} = \left|\Delta A_E(m_j) / A_{N_E}(m_j)\right| \le 0{,}1, \qquad (1)$$

where $\Delta A_E(m_j)$ is the difference between the maximum possible and the total real signal losses given for the input of the electronic amplifier (IEA) (see Eq. 2); $A_{N_E}(m_j)$ – is the minimum permissible signal level (see Eq. 18). Consequently, this is a strict criterion which demands that the minimum permissible signal level at the IAE should be 10 times higher than the level of the relative signal percentage losses.

The difference between maximum possible losses and the total real signal losses given at the IEA is shown in [1]:

$$\begin{aligned}\Delta A_E(m) = A_{\max_E}(0) - \Delta A_{S_E}(m) - \Delta A_{TD_E}(m) \\ - \Delta A_{ISD_E}(m) - \Delta A_{AN_E}(m) - A_{N_E}(m) \ge 0\end{aligned}. \qquad (2)$$

A complete analysis of Eq. 2 requires determination of all input components in the expression. In [1] has been shown the determination of two of the components in Eq. 2 – i.e the losses due to transient $A_{TD}(m)$ and inter-symbol distortions $\Delta A_{ISD}(m)$ in transmission line. Here will be considered the determination of the other terms: $A_{\max_E}(0)$ – the energy potential of the system, $\Delta A_{S_E}(m)$ – signal amplitude losses, and $\Delta A_{AN_E}(m_j)$ – and additive noise losses.

Calculations for signal losses are to be performed for the electronic amplifier input (Fig. 1), therefore the index "E" is introduced.

The solution of the problem is reduced to determining the number of amplifying sections *m* according to permissible signal losses due to transient and intersymbol distortions. This solution is used as boundary condition at the second more complex iterative solution for determining the number of amplifying sections according to total losses.

II. DETERMINING THE NUMBER OF AMPLIFYING SECTIONS m_0 ACCORDING TO THE PERMISSIBLE SIGNAL LOSSES

To ensure reliable performance of the decision device, the signal protectability from its pertaining transient and inter-symbol distortions should be greater than or equal to 6*dB* [3]. Accordingly, the sum of total losses in this particular case should not exceed *6dB*:

$$\Delta A_{TD_E}(m_0) + \Delta A_{TD_E}(0) + \Delta A_{ISD_E}(m_0) + \Delta A_{ISD_E}(0) \le 6dB\,. \qquad (3)$$

To simplify the solution in Eq. 3 it is possible to apply two equal conditions taking into account its monotony:

$$\Delta A_{TD_E}(m_0) \le 3 - \Delta A_{TD_E}(0) = \Delta A_{TD\,AV}\,, \qquad (4)$$

$$\Delta A_{ISD_E}(m_0) \le 3 - \Delta A_{ISD_E}(0) = \Delta A_{ISD\,AV}\,, \qquad (5)$$

where $\Delta A_{TD\ AV}$ are permissible threshold values of signal protectability losses generated by the transient distortions in the transmission line, $\Delta A_{ISD\ AV}$ is the permissible value of signal losses due to the intersymbol distortions [1].

Drawing upon these relationships and those found in [1] it is possible to formulate the following analytical expressions for estimating the number of amplifying sections, depending on the transient and intersymbol distortions:

$$m'_0 \le INT\left[\frac{-5\lg\left(\left(1-10^{0{,}1(A_{NL}+1{,}25)}\right)\left(1-10^{-0{,}2\Delta\alpha}\right)\left(1-10^{-0{,}05\Delta A_{TD\,AV}}\right)^2\right)}{\Delta\alpha}\right] \qquad (6)$$

$$m''_0 \le INT\left[\frac{\left(10^7\dfrac{41}{\left(-20\lg\left(1-10^{-0{,}05\Delta A_{ISD\,AV}}\right)\right)-1}\right)^{0{,}5}}{l}\right]. \qquad (7)$$

where $A_{NL} = 10\lg\left(\left(NC_1P_S\right)/\left(N^3C_3P_S^3\right)\right)$ is the group signal protectability resulting from the total power of nonlinear transient distortions (*N* – number of channels at the optical amplifier output, NC_1P_S и $N^3C_3P_S^3$ – are the power rates of the transient distortion signal at the optical amplifier output, C_1 and C_3 – coefficients of decomposition of optical amplifier output power in the order determined by the levels of input power), $\Delta\alpha$ – difference between the amplitude losses in the amplifying section of length equal to *1km* and the amplification of the optical amplifier. If in Eq. 6 the value under the logarithm sign is negative then it follows that $m'_0 \to \infty$.

After determining the number of amplifying sections according Eqs. 6 and **7**, the value m_0 should be selected from the lower one between m'_0 and m''_0. In the second part of the solution m_0 is used as initial limit value at first iteration.

III. DETERMINING THE NUMBER OF AMPLIFYING SECTIONS m_j BY THE TOTAL SIGNAL LOSSES

Initially it is necessary to determine the unknown terms from Eq. 2.

A. Estimation of the energy potential of the system $A_{\max_E}(0)$

$A_{\max_E}(0)$ characterizes the energy potential of the system which is necessary to calculate the length of the amplifying section. He is equal to the difference between the maximum possible signal protectability $A_{\max E}$ (in each channel) of the optical system and the sum of signal losses generated in the termination devices i.e. he does not depend on the number of amplifiers *m*. Consequently the maximum possible signal level $A_{\max_E}(0)$ will be determined by

$$A_{\max_E}(0) = A_{\max_E} - \left[\Delta A_{S_E}(0) + \Delta A_{TD_E}(0) + \Delta A_{ISD_E}(0)\right], \quad (8)$$

The first term in Eq. 8 depends on the physical characteristics of the transmitter and the receiver module

$$A_{\max_E} = 2p_{SZ1} + S_{APD} - \alpha_{PR} - p_{TN}, \quad (9)$$

where $p_{TN} = 10\lg(kT\Delta fNF/1mW)$ is the level of thermal noise at the input of the electronic amplifier ($k = 1{,}38.10^{-23}$ is the constant of Boltzmann, T – absolute temperature, $\Delta f = f_{T1}$ – frequency band equal to the signal clock frequency for a single channel, NF – amplifier noise figure).

The second term in Eq. 8 reflects signal losses due to atenuation in termination optical devices (optical transmission module and multiplexer/demultiplexer) and is determined by

$$\Delta A_{S_E}(0) = 2(\Delta\alpha_{OTM} + \alpha_M + \Delta\alpha_M + \alpha_D + \Delta\alpha_D) + \Delta\alpha_{ORM}, \quad (10)$$

where α_M, α_D – attenuation in multiplexer/demultiplexer, $\Delta\alpha_{OTM}$, $\Delta\alpha_M$, $\Delta\alpha_D$, Δa_{ORM} – variation in attenuation of optical transmission module, multiplexer/demultiplexer and the optical receiver module. It is evident that all parameters depend on the physical characteristics of the optical equipment.

The third and fourth terms in Eq. 8 are determined by Eqs. 4 and 5 [1].

Therefore, by substituting Eqs. 4, 5, 9 and 10 in Eq. 8 can be determined $A_{\max_E}(0)$.

B. Calculation of losses generated by signal attenuation

Signal losses due to attenuation in the amplifying section $\Delta A_{S_E}(m_j)$ are determined by the physical characteristics of the used fiber and the amplification of the optical amplifiers

$$\Delta A_{S_E}(m_j) = 2\Delta A_{S_O}(m_j) = 2m_j\Delta\alpha, \quad (11)$$

where $\Delta\alpha$ is the difference between the attenuation in the amplifying section of length equal to *1km* and the amplification of the optical amplifier [1]. It is evident from Eq. 11 that these losses will have linear increment with the rise of number of amplification sections.

C. Calculation of losses generated by additive noises

The real signal protectability losses $\Delta A_{AN_E}(m_j)$ generated by additive noises consist of two components

$$\Delta A_{AN_E}(m_j) = A_{TNR}(m_j) - A_{AN_E}(m_j), \quad (12)$$

where $A_{TNR}(m_j)$ is the real value of signal protectability from thermal noises (which accounts and signal attenuation in the transmission line), $A_{AN_E}(m_j)$ – equivalent signal protectability from additive noises at the IEA:

$$A_{TNR}(m_j) = A_{\max_E}(0) - \Delta A_{S_E}(m_j), \quad (13)$$

$$A_{AN_E}(m_j) = -10\lg\left(10^{-0,1A_{QN\,OTM}} + 10^{-0,1A_{QN\,ORM}} + 10^{-0,1A_{SE}(m_j)} + 10^{-0,1A_{TNR}(m_j)}\right), \quad (14)$$

Eq. 14 accounts the influence of $A_{QN\,OTM}$ and $A_{QN\,ORM}$, which stand for signal protectability from laser quantum noise in the transmission optical module and the photo-receiver in the optical receiver module as well as the signal protectability from spontaneous emissions $A_{SE}(m_j)$.

$A_{QN\,OTM}$ and $A_{QN\,ORM}$ are determined by the physical characteristics of the optical transmitter and receiver. $A_{SE}(m_j)$ at the input of the first optical amplifier is determined by

$$A_{SE} = p_{SZi} - \Delta\alpha_{OTM} - \Delta\alpha_M - \alpha_M - p_{SE}, \quad (15)$$

The level of spontaneous emission at the input of the first optical amplifier for wavelength $\lambda = 1550nm$ in the frequency band from *0* to f_{T1} is the same and equal to

$$p_{SE} = 10\lg\frac{h.v.f_{T1}.K_{SE}}{1mW} = -158 + 10\lg f_{T1} + 10\lg K_{SE}, \quad (16)$$

where f_{T1} is the clock frequency, $h = 6{,}62.10^{-34}$ *J.s* – the constant of Plank, $v \approx 2.10^{14}$ *Hz* – spontaneous emission frequency, K_{SE} – spontaneous emission coefficient, *–158,9dB* is the absolute level of spontaneous emission for *1Hz* at $K_{SE} = 1$.

The level of spontaneous emission at the input of every other optical amplifier decreases by $\Delta\alpha$. At the input of the second optical amplifier it has the value $A_{SE} - \Delta\alpha$ and at the input of the *m*-th optical amplifier: $A_{SE} - (m-1)\Delta\alpha$. Consequently by accounting the addition of the spontaneous emission [3] from every optical amplifier, the equivalent signal protectability will be of the kind

$$\begin{aligned} A_{SE}(m) &= -10\lg(10^{-0,1A_{SE}} + 10^{-0,1(A_{SE}-\Delta\alpha)} + 10^{-0,1(A_{SE}-2\Delta\alpha} + \ldots \\ &+ 10^{-0,1(A_{SE}-(m-1)\Delta\alpha)}) = \\ &= A_{SE} - 10\lg(1 + 10^{0,1\Delta\alpha} + 10^{0,1.2\Delta\alpha} + \ldots + 10^{-0,1(m-1)\Delta\alpha}) = \\ &= A_{SE} - 4{,}34\ln S'_G \end{aligned}, \quad (17)$$

with a sum of the terms of the geometrical progression $S'_G = \frac{10^{0,1m\Delta\alpha} - 1}{10^{0,1.\Delta\alpha} - 1}$.

D. Calculation of $A_{N_E}(m_j)$

Minimum permissible signal level at the IEA (used in Eq. 1) is determined by

$$A_{N_E}(m_j) = 9 + 10\lg H^2 = 20\lg(U_S/\sigma_N), \qquad (18)$$

where $H = 0{,}25\sqrt{2}(U_S/\sigma_N)$ (U_S – voltage at the IEA, σ_N – root mean square voltage of the additive noise in relation to the IEA, therefore U_S/σ_N is signal-to-additive noise ratio at the IEA).

The value of $A_{N_E}(m_j)$ is connected with the requirement for error probability in the electronic decision device $P_e \le mlP_{km}$, where P_{km} and l are kilometric error probability (usually 10^{-13}) and length of amplifying sections and are known from the input parameters.

Based on the relationships drawn in [1] and relationships in Eqs. 12 to 17 it is possible to draw a graph of the variation in losses generated by transient and intersymbol distortions plus the losses from additive noises as function of the number of amplifying sections. In Fig. 4 is shown a graphic relationship with the following physical parameters of the optical equipment: $\Delta\alpha = 0{,}3dB$, eight-channel system and optical amplification $20dB$.

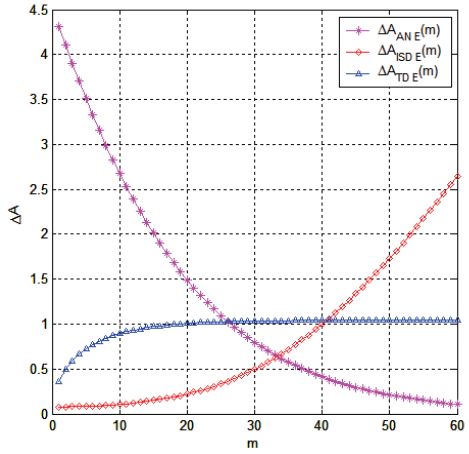


Fig. 4. Losses due to transient, intersymbol distortions and additive noises as function of the number of amplifying sections

From Fig. 4 it is evident that with small number of amplifying sections there is a substantial influence of additive noise losses whereas with the increase of the number of amplifiers – of the losses due to intersymbol distortions. The area closed by the three graphs determines the number of amplifying sections by the criteria of transient and inter-symbol distortion losses and additive noise losses. Proportionally increasing losses due to signal attenuation along the fiber should also be added and the other terms in Eq. 2 are to be accounted for as well.

E. Evaluation of δ_{Cj} *in determining* m_j *by total signal losses*

After all terms have been determined in Eq. 2 and in order to determine Eq. 1, it is necessary to calculate $\Delta A_E(m_j)$ according to Eq. 2.

By substituting Eqs. 2 and 18 in Eq. 1, there should be checked whether *j*-th step of iteration from the calculations for m_j gratifies Eq. 1. Provided $\Delta A_E(m_j) \le 0$ then in the next *(j+1)*-th iteration *m* should be decreased (i.e. $m_{j+1} < m_j$) and all calculations repeated in order to determine δ_{Cj+1}. If $\Delta A_E(m_j) > 0$, but Eq. 1 is not gratified then in the next *(j+1)*-th iteration *m* should be increased (т.е. $m_{j+1} > m_j$).

Sequential iterations are applied unless the condition of Eq. 1 is fulfilled. Consequently the final result for the number of amplifiers (amplifying sections) will gratify limitations imposed by the total signal losses in assuring respective quality factors of the transmitted signals.

IV. Conclusion

The following conclusions could be drawn:

- The characteristics of the transmission line depend on the type of fiber whilst the signal amplification depends on the characteristics of the optical transmission and receiver module plus the multiplexer/demultiplexer. These characteristics have their impact while determining the number of amplifying sections and, accordingly, are included in the content of the analytical expressions.
- The iterative approach allows optimal determination of the number of amplifying sections based on two principal criteria: the losses generated by transient and intersymbol distortions and the total losses.
- The proposed approach is applicable i.e. it is based on the physical characteristics of the used optical equipment.

References

[1] K. Angelov, K. Koitchev, N. Varbanova, Estimating losses from transient and intersymbol distortions in hybrid fiber-coaxial television network, ICEST 2009, Proc. of Papers Vol. 1, pp.113-116, Sofia, Bulgaria, 2009.

[2] K. Angelov, K. Koitchev, S. Sadinov, An Investigation of Noise Influences in Optical Transmitters and Receivers in Cable TV Networks, ICEST 2006, Proc. of Papers, pp.102-105, Sofia, Bulgaria, 2006.

[3] R. Freeman, Fiber-Optic Systems for Telecommunications, John Wiley & Sons, New York, 2002.

[4] Л. Йорданова, В. Топчиев, Оптимизиране на параметрите на усилватели, използвани в оптичния канал на CATV системи, TELECOM 2008, стр. 206-211, Варна, 2008.

[5] О. Панагиев, Определяне на оптималния брой на последователно свързани усилватели в широколентова кабелна комуникационна мрежа. TELECOM 2008, стр. 255-260, Варна, 2008.

Adaptive Cross-layer Optimization in Wireless Fading Channel and Limited Buffer Capacity

Zoran S. Veličković[1] and Milojko Jevtović[2]

***Abstract* – Difficulties in the realization set Quality of Service (QoS) in wireless multimedia communications can be overcome by jointly optimizing the network parameters over ISO/OSI layers. This paper presents optimization algorithm that maximizes the network throughput of the wireless communication fading channel and limited capacity of transmit buffer and the available transmission power of network nodes. The applied algorithm is based on the Markov Decision Process (MDP), which requires modelling the entire communication system with Markov chain; determine the set action, transition probabilities and functions of the cost. The structures of optimal policies for different models of network traffic at the constant BER are analyzed. Numerical results obtained in the simulation environment show the reasonableness of applying MDP.**

***Keywords* – Adaptive cross-layer adaptation, Markov Decision Process, Optimal transmission policy.**

I. INTRODUCTION

Adaptive network techniques have great importance in the implementation of modern wireless network systems. They provide a dynamic adaptation of wireless network like as reaction on the changes in the quality of the channel or the type of network traffic. Adaptation of network parameters is intended to improve network performance, and increasing the quality of services [1]. In addition to standard network services, modern wireless mobile communications provide attractive new services. In order to define the quality of services, ITU-T G.1010 specification standardizes its key network parameters from the perspective of users [2]. According to G.1010 parameters which affect Quality of Service (QoS) are: *delay*, *delay variation* and *information loss*. Depending on the application, according to this standard, certain levels of delay, delay variation, or loss of information can be tolerated. Thus, delay up to 150ms and delay variation up to 1ms at conversational speech can be tolerated. On the other hand, downloading content from the Web can tolerate a delay up to 10s; a variation of the delay is not specified. It is evident that different applications require a wide range of QoS, from intolerant of delay to those more tolerant of packet loss. QoS may be required in different forms, but occurs most frequently in the form of the required rate or a defined BER. The inability to guarantee the delivered QoS is a fundamental problem in wireless packet communications. This problem is particularly pronounced in mobile multimedia applications. In order to guarantee QoS and multimedia applications in wireless computer networks has been developed 802.11e standard [3]. Under this standard, depending on the type of application is given priority before the packets delivery to MAC layer. In this way, better system performance can be got in low traffic, but the higher traffic they will degrade. This leads to the conclusion that the solution to this problem can not be required to optimize network parameters only one ISO/OSI layer. Generally, to improve the quality of services in a wireless communication system is necessary to discuss the joint optimization of parameters with multiple ISO/OSI layers. Jointly optimization of network parameters of several ISO/OSI layers according to [4] is called *cross-layer* (CL) design. Jointly consideration of PHY and MAC-LLC layers in order to satisfy QoS is presented in [5]. In [6] cross-layer model for multimedia multicast/broadcast services to efficiently support the diverse QoS over mobile wireless networks were shown. Power/delay tradeoffs of a single user system with a finite buffer are analyzed in [7]. Optimal packet scheduling over correlated fading channels which tradeoff between minimization of three goals: average transmission power, average delay and average packet dropping probability are considered [8].

In this paper, CL is considered an optimization problem related to the maximization of flow in the wireless communication system with limited resources. First of all, restrictions are related to the level of engaged transmission power and the available buffer capacity in the communication node. Set QoS request relates to the provision of the required bit error (BER) as required by some multimedia applications. Packet traffic characteristics also significantly affect the way to achieve the required QoS. In this paper packet traffic is modeled by Poisson's distribution and by packed constant rate (CPR). The optimal transmission policy should take into account the type of packet traffic, limited buffer capacity and condition of the communication channel. This requires an optimization protocol can access the network parameters from the two layers. At the PHY layer contains information about the quality of wireless channel, until the MAC-LLC layer contains information about the state of the communication buffer in the function of the type packet traffic. Markov decision process (MDP) [9] has been applied as tools for optimization of network parameters. If MDP is used, transition probability matrices depend on the system state and actions which "the software agents" can perform in every discrete time segment in which decisions are being made (decision epoch). The software agent can achieve *rewards* or *costs* depending on the performed action and the state of the system at that decision epoch. The aim of the MDP is found

[1]Zoran S. Veličković is with the High Technical School, A. Medvedeva 20, 18000 Niš, Serbia, E-mail: zoran.velickovic@vtsnis.edu.rs.

[2]Milojko Jevtović is with the Yugoslav Engineer Academy, L.Ribara 7, 11000 Beograd, Serbia, E-mail: vlamijev@ptt.rs

optimal policies for defining the series of actions which should be taken in every state of the system, the utility functions that could be maximized. Adaptive techniques to PHY and MAC layer as well as the needs of their joint optimization will be presented in the following work.

Further in this paper the optimization problem of wireless communication system with one user for a constant and Poisson model of traffic has been defined. It has been developed as composite communication system model based on Markov chains of PHY and MAC layer LLC. The system model and solution of the optimization problem by using MDP is presented in section III. The obtained results and the analysis of the optimal structure policies in a simulated environment are presented in section IV. Main conclusions are summarized in the final section.

II. ADAPTIVE TECHNIQUES

A. *Adaptation of the PHY layer*

Adaptation of the code and modulation technique is the primary form of network parameters adjusted to new conditions in the wireless communication channel. Frame is a unit of data to the PHY layer to which the applicable code and modulation scheme. Modern wireless communication protocols at the PHY layer have a limited choice set of code and modulation scheme. However, the present protocol do not specify the way to use this scheme, they are only offer a choice. An important parameter in the choice of code and modulation scheme is the level of fading in communication channel which is determined by the SNR at the destination [10]. The quality of the communication channel can also be characterized as probability of discarded packets P_p. The level of SNR can be affected by a variation of transmitting power and adaptation of modulation parameters, which will achieve different optimization objectives. Thus, the basic mechanism of adaptation of network parameters on the PHY layer is a variation of transmission power and choice of code and modulation scheme, which adapts to flow in the wireless communication system. In real conditions they can not achieve the maximum network flows defined by the selected network parameters given the fact that there is not always data available to send. Temporary delay in sending can be optimized transmission policy in order to save transmission power, as shown in the following work.

B. *Adaptation of MAC-LLC layer*

The basic data unit at the MAC-LLC layer is packet, so that the maximum capacity is determined by the number of packet which can be put into it. In addition to the quality of the channel PHY layer, the size and occupancy transmitting buffer significantly affect the number of discarded packets. Since the buffer is organized on the principle of first-in-first-out, the last arrived packet is placed on top of this structure. However, if the packet arrives in the buffer when it is completed, there is a rejection of the packet.

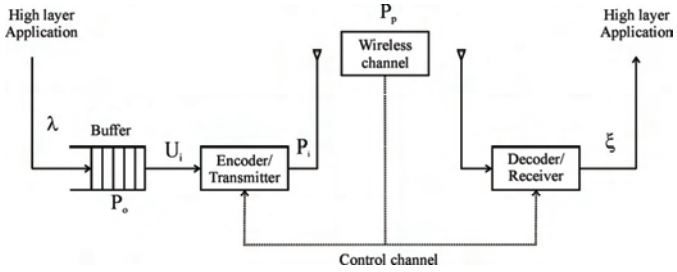


Fig. 1. Wireless communication system with one user and limited buffer capacity.

The dynamics (loading/emptying) and the current state of the buffer will depend on the number of discarded packets. The probability of discarded packets P_o due to buffer overflow depends on the dynamics of the system and the size of the buffer. Thus, the probability of rejection of the packet is reduced if the buffer capacity increases, or, if you increase the speed of emptying the buffer. As already mentioned, the adaptation of code and modulation scheme may affect the likelihood of rejection of the packets. Depending on the type of application, loading transmitting buffers with higher ISO/OSI layers can be modeled as a random process or a process with a constant rate of packets. To analyze the structure of transmission policy, a buffer loading is modeled as Poisson process, i.e. the CPR process.

C. *Cross-layer optimization of PHY and MAC-LLC layer*

In order to meet different QoS, it is necessary to consider optimization problem in several ISO/OSI layers. Adaptation of transmitting policy depends on the state of communication channels and buffers state in the receiving node. Taking into account the probabilities of rejection packets due to the occurrence factor in the communication channel and the probabilities of rejection packets due to overflow, packet loss rate can be expressed as:

$$\xi = 1 - \left(1 - P_o\right) \cdot \left(1 - P_p\right), \tag{1}$$

where P_o is the buffer overflow, P_p is the packet error probability. If the application requires a constant bit error rate (BER), the maximum throughput obtained by minimization of expression (1) is equal to minimization the parameter P_o.

III. SYSTEM MODEL

The structural block diagram of the considered wireless communication system with one user is presented in Fig. 1. This communication system consists of a transmitter, which incorporates a buffer of limited capacity and encoder-modulator and of a receiver. Loading of the buffer on the transmitter side is done with packets from the higher ISO/OSI layers and is modeled by constant packed rate (CPR) and Poisson distribution. The connection between the transmitter and the receiver is realized in Rayleigh fading channel, while information about the quality of the channel and the buffer state are being interchanged through the control channel as it

has been presented in Fig. 1. The analysis of the work of the presented system has been considered through a series of successive time frames *i* whose duration is T_f. The mean value of the number of packets coming into the buffer while the frame lasts, λ, is calculated as, $\lambda = \mathrm{E}\{A_i\}$ where $\mathrm{E}\{\bullet\}$ is the operator of the mathematical expectation. It has been supposed that all the packets are of equal length and that the buffer is of limited capacity and that it can receive only *B* packets. All the packet that arrive at the moment when the buffer is completely loaded will be rejected and regarded as lost. Another source of the rejected packets in the considered system is the wireless communication channel with correlated fading. PHY [10] and MAC-LLC models layer are based on Markov chains. The space state of the communication channels (*g*) and buffer (*b*) defines the space state of complete communication system. In order to complete Markov model, the composite transition probability matrix is determined. The resulting model of the entire communication system is complete so that you are able to join set of actions which the system can translate from one state to another. Transition probabilities of the composite communication system p_S in realization of the action *u* can be determined:

$$p_S(s,s') = P_r\{s_{i+1} = s' | s_i = s, U_i = u\}$$
$$= p_G(g,g') \cdot p_B(b,b',u), \quad (2)$$

where p_G and p_B are transition probabilities of the PHY and the MAC-LLC subsystem respectively. To complete the MDP model of the entire communication system need to determine the cost function that is implemented at each transition:

$$R = P(u,g,\overline{P_b}) + \beta \cdot L_o(b,u). \quad (3)$$

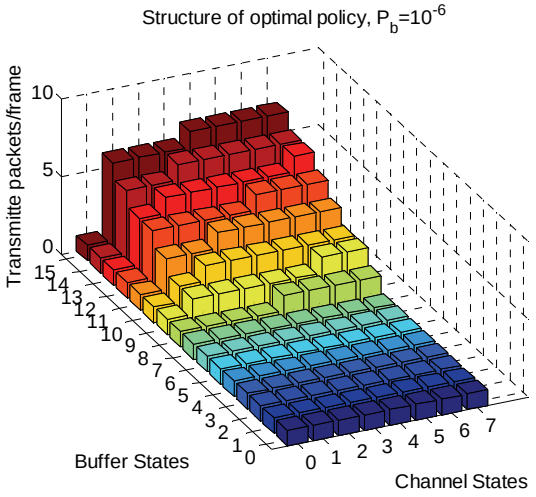
The first addendum in the weight sum of costs (3) relates to the level of the engaged power (*P*), while the second addendum relates to the number of rejected packets (L_o). By the parameter β the mutual relation between the addenda is determined and the tradeoff between the set of criteria is provided. Stationary optimal policy π^* should provide a minimum of the function (4) along with meeting the criterion of power set. The aim of MDP is minimization of the mean value of the expected discount sum:

$$\pi^* = \arg\min_{\pi}\left\{\limsup_{T\to\infty}\frac{1}{T}\mathrm{E}\left[\sum_{i=0}^{T-1}\alpha^i R(S_i,U_i)\right]\right\},\ 0<\alpha<1. \quad (4)$$

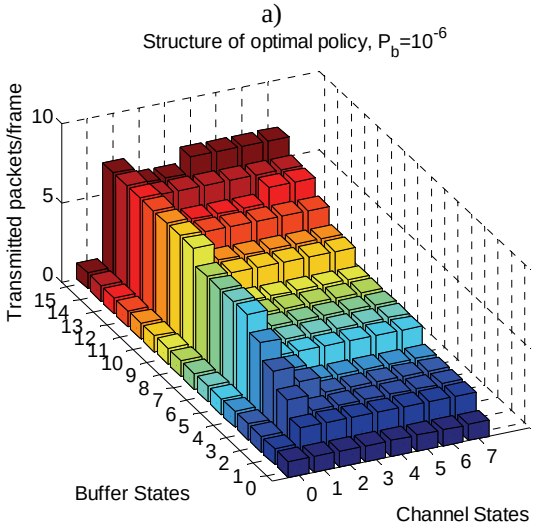
Solution of the optimization problem can be presented by following notation, $U_i = \pi^*(S_i)$ where S_i represents state of the communication system and α^i discount factor of the i-th frame. Therefore, an optimal policy relates to every frame i on the base of the state of the communication system Si and is realized with emptying the transmitting buffer in optimal number of packets U_i. By solving the optimization problem (4) for individual values of the parameter β, Pareto-optimal values $\overline{P}$ and L_o are determined.

IV. Simulation Results

This section presents the results of the optimization of flow with limitation average transmission power and the satisfaction of a required BER. Structure of optimal transmission policy is determined for two models of loading communication buffer. The first model (CPR) involves loading the buffer at a constant rate, λ, while the second model of loading buffer is realized by Poisson distribution with mean value λ. MDP was used for the optimal solution of the problems, while the simulation parameters have the following values: $\lambda = [1100 \quad 1200 \quad 1300]$ packets/s, *L=100* bits in a packet, the buffer capacity *B=15* packets, BER $P_b = 10^{-6}$, bandwidth *W=100* kHz, noise power density $N_o = 2x10^{-5}$ W/Hz, duration of the symbol is fixed to $T_s = 1/W$ and the frame duration is determined with 100 symbols. In MQAM systems this parameter arrangement allows for the buffer to be emptied with the *u* packets/frame with a signal constellation $M = 2^u$. The channel with the time correlated Rayleigh fading is modeled with eight states as in [10, Table I]. For different values of the weight parameter, β, the solution of the optimization problem is the vector of the rates U_i in which the communication buffer should be emptied for every state of the communication system. On the base of the value of the vector U_i the number of rejected packets L_o is determined and the value of the engaged power $\overline{P}$. By averaging the values obtained for the engaged power and the number of rejected packets for all frames, Pareto-optimal values have been obtained. In Fig. 2 the structure of the optimal policies is shown for a) CPR and b) the Poisson model of traffic in the channel fading. For the same parameter value, λ, one can notice different structure of the optimal policy. The delay in sending packets in low occupancy of the buffer regardless of the quality of the channel is notably for the CPR traffic model, Fig. 2a. On the other hand, when it reaches certain buffer occupancy, improving quality of the channel and increasing occupancy buffer increase the speed of emptying the buffer. Figure 2b shows the structure of the optimal policy for the case Poisson's model traffic conditions in case of the same values of fading. Similar to the CPR model, withdrawal of sending in the case of the worst quality of communication channels is recommended, but we recommend sending the packets to all other conditions. Maximum transmission rate is recommended for the worst quality of the communication channel that allows you to send yet. Of course, if the buffer occupancy increases, speed of emptying the buffer increases too. However, unlike water-filing algorithm, this transmitting rate does not necessarily increase with the increase quality of the channel. With the increase of buffer occupancy, the rate of empting Poisson's model increases too. Based on the presented results, the structure of the optimal policies depends on the model of the communication channel, i.e. transition matrix of the communication channel. Fig. 3 shows Pareto-optimal values of the average engaged power and the average number of rejected packets. These parameters allow satisfying the wide palette of compromise demands.



a)



b)

Fig. 2. Structure of optimal policy: a) CPR b) Poisson's traffic model for constant BER.

From Fig. 3 it can be concluded that the CPR model of traffic can be realized by policies with lower transmission power compared to the Poisson model of traffic. Regardless of the traffic model, with an increase in the number of packets which buffer is loading, a minimum average of discarded packets is also increasing. In order to realize the optimal policies information from two layers (PHY, MAC-LLC) has been used, while realization was done only by adjusting parameters on one layer (PHY). The results obtained in simulated environment of Matlab show that the application of cross-layer design in modern wireless multimedia applications is justified.

V. Conclusion

Jointly optimization of network parameters with multiple ISO/OSI layers provides the satisfaction of QoS requirements set out in multimedia applications. Type of application determines the characteristics of traffic and it has a significant impact on the optimization algorithms. This paper demonstrates that the CPR traffic model in the channel with

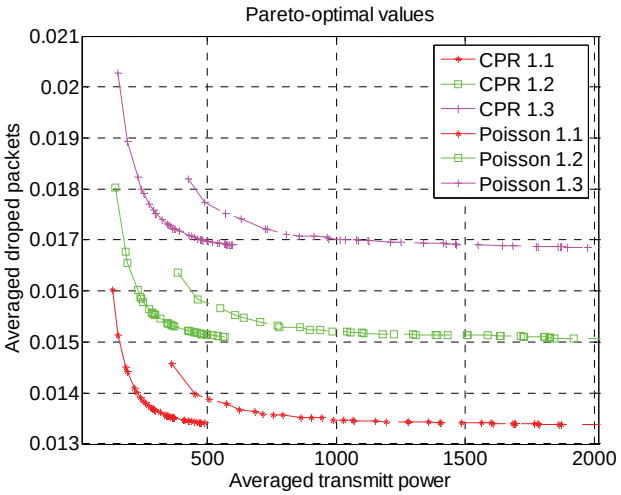


Fig. 3. Pareto-optimal values of averaged dropped packets vs. averaged transmit power for CPR and Poisson's traffic model.

Rayleigh fading requires less transmission power compared to Poisson's traffic model. Cross-layer design provides joint optimization of application layer with the rest of the protocol stack, thereby improving network performance. For the appropriate model of network traffic parameters an optimal transmission policy can be stored in a table, where it can be used later. Optimization algorithm based on the MDP is demonstrated through simulation examples and its effectiveness is illustrated.

References

[1] M. Ibnkahla, *Adaptation and cross layer design in wireless networks*, CRC Press, 2009.

[2] ITU-T Recommendation G. 1010: End-user multimedia QoS categories – ITU-T, 2001.

[3] IEEE 802.11e WG. 2005. Wireless LAN MAC and PHY specifications amendment 8: MAC quality of service enhancements, http://standards.ieee.org.

[4] V. Srivastana, M. Motani, "Cross-layer design: a survey and the road ahead", IEEE Commun. Mag., vol. 43, no. 12, pp. 112–119, Dec. 2005.

[5] P. Popovski, M. A. Ingram, C. B. Peel, S. Hara, S. Toumpis, Editors, "Cross-Layer Design for the Physical, MAC, and Link Layer in Wireless Systems", EURASIP Jour. On Adv. In Sig. Proc., 2009.

[6] X. Zhanh, Q. Du, "Cross-layer Modeling for QoS-Driven Multimedia Multicast/Broadcast over Fading Channel in Mobile Wireless Networks", IEEE Commun. Mag., vol. 43, no. 12, pp. 62–70, Avg. 2007.

[7] A. J. Goldsmith and S. G. Chua, "Variable-rate variable-power mqam for fading channels," IEEE Trans. Commun., vol. 45, no. 10, pp. 1218–1230, Oct. 1997.

[8] A. K. Karmokar, D. V. Djonin, and V. K. Bhargava, "Optimal and suboptimal packet scheduling over correlated time varying flat fading channels," IEEE Trans. Wireless Commun., vol. 5, no. 2, pp. 446–457, Feb. 2006.

[9] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, New York, NY: John Wiley & Sons, 1994.

[10] H. S. Wang and N. Moayeri, "Finite-state markov channel–a useful model for radio communication channels," IEEE Trans. Veh. Technol., vol. 44, pp. 473–479, Feb. 1995.

Functional Model of Algorithms for Optimal Data Transmission Over Air Interface

Vasil Shterev[1] and Georgi Iliev[2]

***Abstract* – This paper proposes a model of architecture for optimal transmission in a wireless channel. The nature of this type of channels is random and stochastic; hence algorithms and methods of control are more difficult and complicated. Our goal is to achieve optimal performance and maximum throughput for data transfer over this type of channels. The model is implement in an UML (Unified Modeling Language) environment. We investigate the wireless channel and mitigate the negative effects in it. In our model, we take into account most important parameters that influence the characteristic of the channel. On the other hand these parameters must be easy and fast to estimate in order to achieve effective and feasible management of our system.**

***Keywords* – Teletraffic, Network Control, Data Communication.**

I. INTRODUCTION

The main problem in a wireless communication system is the stochastic channel. Many existing systems [2] try to mitigate the negative effects when we exploit a radio channel to transfer information (data, voice and video). Popular well-known examples of these systems are IEEE 802.11 (Wi-Fi) and IEEE 802.16 (WiMAX). They play a scenario in which most reconfigurable parameters are separately tune due to changes in the environment. In the IEEE 802.16 system, designers have include three basic values from the OSI MAC layer. These parameters are: output power level; adaptive modulations scheme and control of coding rate [2].

Further, we develop a model in the UML environment, where an exact sequence of actions is taken. The reason to select UML as tool is because: it is demonstrative and easy to understand; it is a standardize general-purpose modeling language and it combines the best techniques from data modeling [1].

II. A BRIEF OVERVIEW OF UML

UML has many versions (from ver. 1.0 to 2.2 nowadays). The latest version has 14 types of diagrams divide into two categories [4]. Seven diagram types represent the structural information, and the other seven represent general types of behavior, including four that represent different aspects of interactions.

[1]Vasil Shterev is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: vas@tu-sofia.bg

[2]Georgi Iliev is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: gli@tu-sofia.bg

UML does not restrict UML element types to a certain diagram type. In general, every UML element may appear on almost all types of diagrams. This flexibility is very important for developers.

We will concentrate on three types of diagrams: one of structural information and two of behavioral category. From the first category, we chose a class diagram, because it depicts how different components are connect together to form larger components.

Our second choice is to use a use case diagram, because it presents a graphical overview of the functionality provide by a system in terms of actors, their goals (represent as use cases), and any dependencies between those use cases. The actors in the use case diagram represent functionality which have a relation to the functionality present in use case and this functionality is not depict in detail [3]. These diagrams are very important when we model the behavior of a system.

The third type is a sequence diagram. It is shown as parallel vertical lines (lifelines), different processes or objects that live simultaneously, and, as horizontal arrows, the messages exchang between them and in the order in which they occur. This allows the specification of simple runtime scenarios in a graphical manner.

III. THE UML DIAGRAMS

In this section we propose three types of UML diagrams. We will start with use case, then follow class diagram and finally is sequence one.

A. Use Case Diagram

Use case diagram is depict in Fig.1. This figure depicts two main types of actors: operator and customer. The Operator has connections (also call communication association) with different use cases. In our case, they are: system startup, effective usage of system, maximization of profit, illegal usage and user account. Operator has also connection to a billing system, which in turn is connect with session use case. The last one has communication association with second actor – the customer. A Session is communication with connection, which in turn has a connection with QoS requirements use case and allocation channel. Power management and free resource unit are connect to an allocation channel. In turn the free resource unit has a communication with the free time slot and the free bandwidth use case. This is an overview of interaction between different use cases.

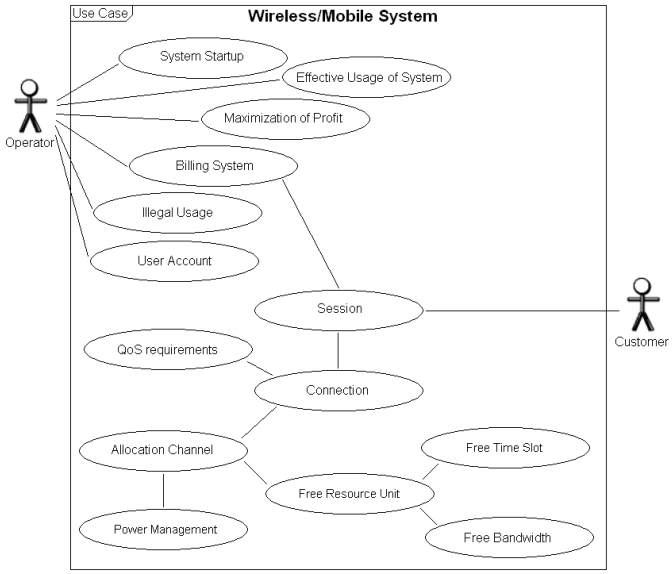


Fig. 1. Use case diagram of wireless/mobile system.

B. Class Diagram

Next, we show a class diagram which depicts a relation between separate classes. The reason, why in Fig. 2 we miss to write attributes and methods, is clear - all attributes and methods are describe below Fig. 2.

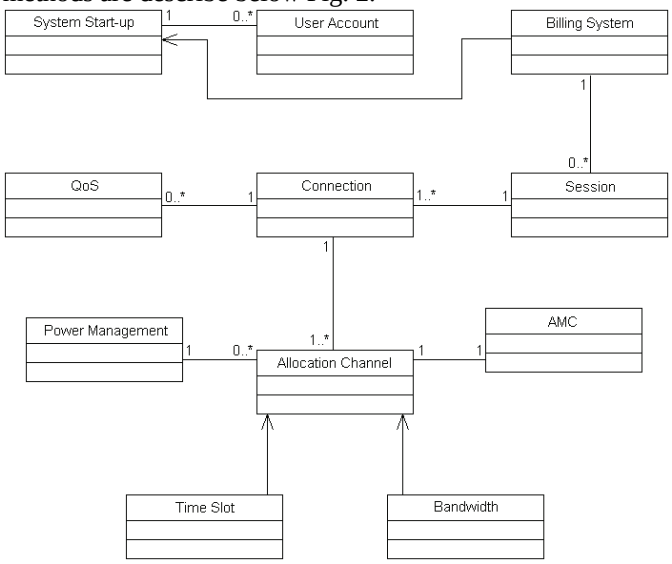


Fig. 2. Class diagram of wireless/mobile system. In second and third row, we intentionally miss to put the attributes and methods.

The first abstract class is System Start-up. It has a relationship with two other classes, which are User Account and Billing System. The association between User Account (UA) and System Start-up (SS) is such that exactly one SS can interact with 0, 1 or more UA. The SS has a dependency relationship with abstract class Billing System (BS). In turn, BS has an association with Session abstract class, with multiplicities of type one BS and multiple (including zero) Session.

QoS and Connection have the same multiplicities as previous ones, because it is possible that we do not have any QoS requirement for our connection. The opposite situation, where we need nearly all requirements is also possible. The relation between the Connection and the Session is little bit different because one Session must have at least one or more Connections. The situation is the same in the relationship between the Connection and the Allocation Channel (AC). Here one Connection must have at least one AC. The situation between Power Management (PM) and AC is as follows: one PM has the possibility to serve zero or more AC, because PM functions independently.

The abstract class AMC (adaptive modulation and coding) and AC have a one to one type of relationship, because we use a particular AMC scheme for each channel. Finally, the relationship between Time Slot (TS) and AC is the same as that of a Bandwidth (BW) and an AC. They are dependent on each other, because the AC cannot exist without TS and BW. Further we describe the attributes and methods of each class.

The SS is very important, because it is responsible for the start of the whole system and checks different components, so attributes assign to this abstract class are: system test, memory test and component test, while the methods and operations which correspond to this class are: check memory; check system block and check peripheral component. Next class is the UA, which is characterize by identification, user name and password as attributes; and check for illegal usage, check password and validation account for operations.

BS is also very important, especially for the operator of the system. This class has an attribute call identification and a method call update account record. The next class is QoS, which has the following attributes: delay, jitter, and bandwidth and packet loss ratio. The operations that correspond to this attributes are: watch delay, watch jitter parameter, look up for bandwidth and look up packet loss ratio. Further we examine class PM. It possesses the following attributes: transmitting power and maximal power. The operation here is call power control. The next abstract class is the AC, which has two attributes: network status and scheduling. Operations that correspond to this abstract class are: network status estimation and actual schedule scheme. The Session class has four methods: start session time, end session time, requirements and maximal profit. Methods are: set session start time and set session end time. Next abstract class is Connection. It has five attributes: start connection time, end connection time, bit rate, bit error rate (BER) and delay. Corresponding operations are: set connection start time, set connection end time, watch bit rate, watch BER and look up delay. The next class is call AMC and it has two attributes and methods. They are: modulation scheme, convolutional code rate and respectively adaptive modulation and code rate adjust. The last two abstract classes are exceptionally important for the correct operation of wireless / mobile systems. The first one is TS, which has the following attributes: time and free time slot. Methods are: system time,

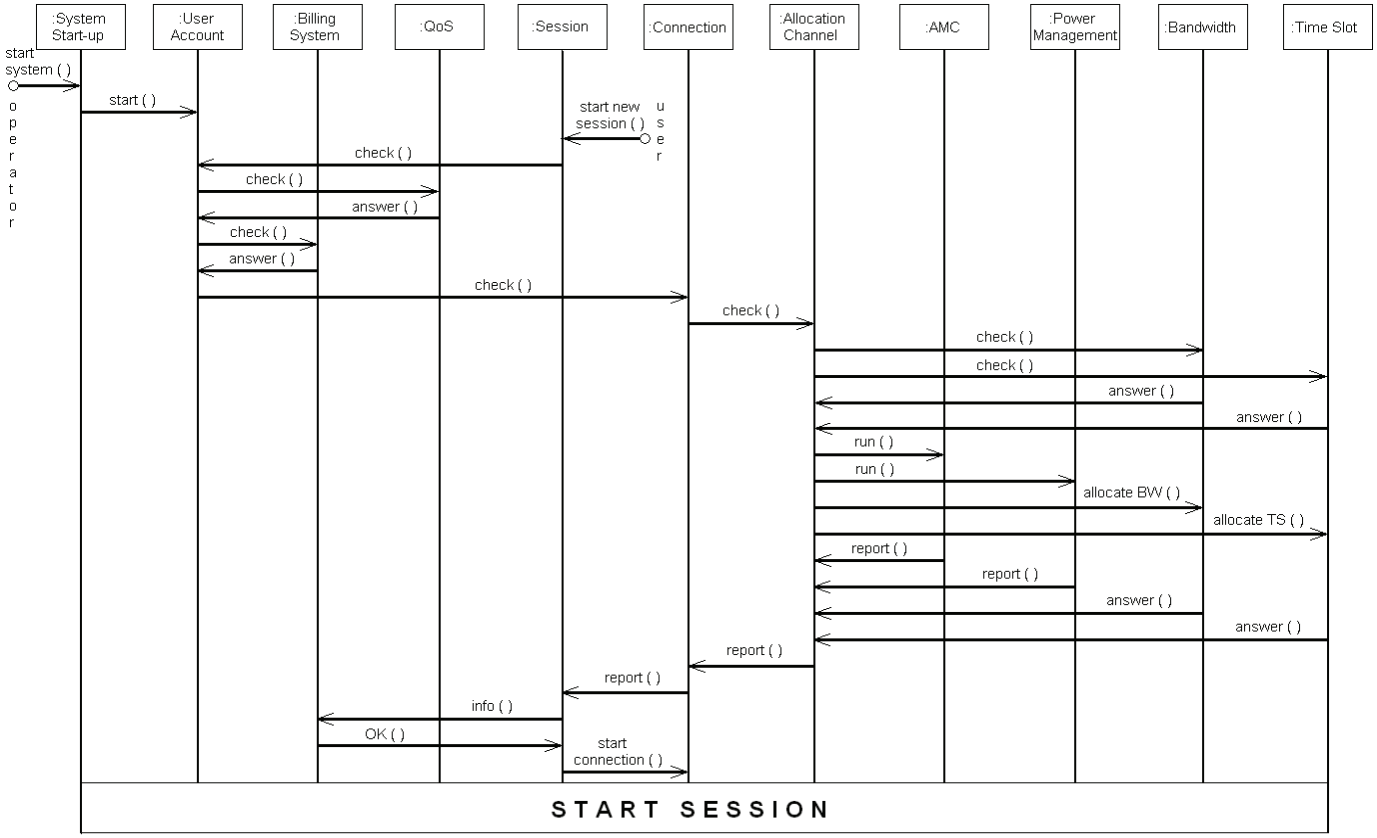


Fig.3 Sequence diagram, which depicts the system star-up and the start of a new session initiate by a customer. The diagram ends when session start is successful.

time slot estimation and time slot validation. The second class is BW. It has three attributes: frequency, free bandwidth, signal to noise ratio (SNR) and two methods, which are: frequency estimation and SNR estimation.

C. Sequence Diagram

Next, we will describe message flow from the initial moment, when the operator turns on the system and user (costumer) who wants to begin a new session, with parameters that are correspond to his requirements. We will discontinue the message flow when the session starts. Each message has various specific parameters, which are denote with brackets.

In Fig. 3 we show the message sequence. The First message comes from a wireless / mobile network operator. Its purpose is to start the system when reach SS. When SS receives this, it performs the actions that have been programmed in it. Actions are logically connect with attributes and methods in class diagrams. Further, the SS sends a start message to UA to activate its functionality.

From some arbitrary moment on, our customer wants to begin a new session. The message representing this action is call *start new session*, and it is receive from the Session object. This message contains information about customer requirements (video call, video on demand, file transfer etc.). The next step is to send a message *check* from Session to UA to check for illegal usage. If the user has a valid account, the system will send a *check* message to the QoS object to obtain the parameters that characterize the specific session. Further, the QoS sends back an *answer* containing the specific characteristics of this customer session. Then UA sends the *check* to BS to verify whether the customer has enough credit to perform the session. If he has enough credits, the *answer* message between BS and UA is positive and then the UA sends the *check* message to Connection to inspect, whether the wireless / mobile system can perform the requirement. Further, Connection sends again the *check* message to the AC object. The aim is to check-up whether the system has enough capacity to accept one more session. This is perform by sending the *check* message from AC to BW and to the TS objects. They, on their part, send back an *answer* to AC. The answer is either yes or no, and it depends on whether the system has or has not enough free time slot and bandwidth. If the answer is positive, the AC sends a *run* message to AMC and PM; *allocates BW* and *allocates TS* to BW and TS objects. The next step is to send the corresponding reply messages to AC. They are: *report* from PM and *answer* from BW and TS. The aim is AC to verify that AMC and PM are in good condition. Moreover, when BW and TS return the *answer* message, they report that they are allocate in this slot and are ready to use it. Further, AC sends a *report* message to the Connection, and the Connection also sends a *report* to the Session. Then the Session sends an *info* message to BS with the parameters of session such as session start time (to start the bill) and others that relate to the user profile. The answer

message from BS to Session is *OK*. The last message is from the Session to the Connection and it is call *start connection*, thus the session is starting. Logically, termination of a connection is doing reversely, but it has many variations depending on the type of session.

IV. The Algorithm for Power Control, Code Rate and Modulation

In this section we propose an algorithm, which deals with power control, convolutional code rate and modulation. It is done as block scheme, which is illustrate in Fig. 4.

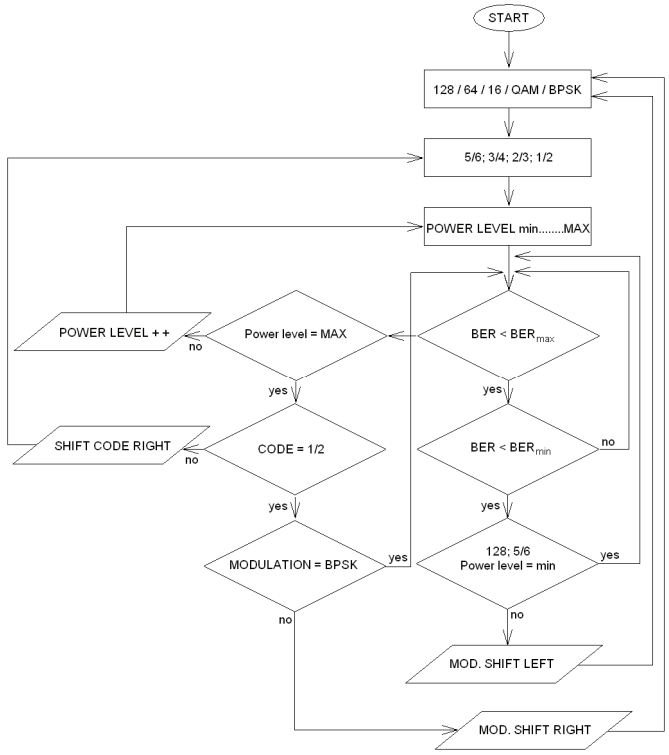


Fig.4 Block scheme of algorithm that controls power level, modulation and convolutional code rate.

The algorithm begins at start block. Next is a block where the system changes the modulation scheme from among five types (128QAM, 64QAM, 16QAM, QAM and BPSK). They are arrange from more effective to most robust. The next block is the convolutional code rate. It has four schemes which are 5/6, 3/4, 2/3 and 1/2. Here, we again arrange it from low information excess to higher. Further, comes the block call power level. If we have to interpret upper logic, from the left side is that of minimum power level, which gradually increase when we move right. The final level, of course, is maximum permit transmitting power. For these three blocks, we have to notice, that if the arrow comes from above, blocks choose the value most on left. If the arrow comes sideways, the block selects the value which is indicate by the algorithm. Next, the algorithm has a combination of decision blocks, where the algorithm is branch depending on the channel condition. Finally, the system has four processing step blocks, which provoke action (move left or right) to change power level, convolutional code rate and modulation scheme. Parameters BER_{min} and BER_{max}, can be choose specifically depend from application and customer necessary. Typical values for BER_{min} and BER_{max} are 10^{-8} and 10^{-4}. If BER is less than BER_{min}, we have to change parameters of system because we waste resource. If BER is bigger than BER_{max}, system is trying to mitigate this.

The logic behind this is as follows: when the channel quality is good, the MAC layer chooses the highest modulation rate, (128 QAM), giving the system the highest throughput. When the channel quality degrades, the MAC layer reduces the modulation rate (64 QAM), reducing the throughput. In practice, adaptive modulation and coding rate are use in conjunction with power control. In point-to-multipoint network deployments with multiple users in a cell service by a base station, when a link degradation arises for a user, the base station first increases the transmit power of the user to provide extra link budget gain, until it reaches the maximum permitted. If the receive signal quality does not improve, then the coding rate is reduced. Extra redundancy is add to provide more coding gain for better error correction performance. If the receive signal quality still does not improve, then the modulation rate is reduce as a last resort (as this significantly affects the throughput then others). Similar (reverse) process is also follow when link quality improves.

V. Conclusion

In this paper we propose three types of UML diagrams for wireless / mobile system. The first (use case) is more general and illustrates the main subsystems and topics in the system. The second one (class) depicts separate systems with their attributes and methods. It also shows the types of dependencies and relations between classes. The third diagram (sequence) depicts the sequence of messages between different objects. Parameters of messages are explain briefly.

Finally, we propose an effective algorithm for power control, code rate adjustment and modulation management. The algorithm is present as a block scheme, and it is so easy to implement in hardware. Further, we consider developing the algorithm and include more important parameters such as buffer control and interaction between customers.

Acknowledgment

This work is supported by the Technical University of Sofia Science Fund – Grant No. 102ni191-7/2010 „Investigation of adaptive methods for QoS in communication networks”.

References

[1] G. Booch, I. Jacobson, J. Rumbaugh, “UML Distilled Second Edition”, A Brief Guide to the Standard Object Modeling Language, 1999.

[2] Y. Zhang, “WiMAX Network Planning and Optimization - Wireless Networks and Mobile Communications” CRC Press Taylor & Francis Group an Auerbach Book, 2009.

[3] Т. Николов, А. Ценов, „Моделиране на Телекомуникационни Процеси и Системи” София 2007.

[4] http://en.wikipedia.org/wiki/Sequence_diagram

Functional Architecture of a Service Level Management System

Todor I. Georgiev[1], Aleksandar K. Tsenov[2]

***Abstract* – The goal of the proposed work is to introduce the functional architecture of Service Level Management system. The system should be implemented in an enterprise that delivers a wide range of IT – services. The architecture includes the main functional blocks, the interfaces between them and the interfaces to external functions as well. The internal structure of each functional block will be modelled and the information structure for the interactions will be presented as well. The work includes also recommendations for feasible and effective tools and platform for the implementation of the proposed functional architecture.**

***Keywords* – Service Level Management (SLM), Quality of Services (QoS), Information Technology Infrastructure Library (ITIL).**

I. INTRODUCTION

Increasingly, the relationship between IT (Information Technology) departments and their internal customers is that of client–supplier, based on the mechanisms of marketing and competition. IT departments are losing their internal company monopoly and have to compete with services offered on external markets. In this scenario a consistent customer focus of the IT management is of pivotal significance [1; 4; 10]. IT departments are facing the challenge of emerging from a technology-oriented applications developer and infrastructure operator to a client-oriented IT service provider.

Reference models help to reduce the costs and risks inherent in the transformation of organizational processes [2; 3]. This explains IT management's increasing interest in reference models for service oriented IT processes. Within the framework of change mentioned above they ensure systematic structuring of customer focused IT management processes at reduced cost and risk [11]. Within the last few years the IT Infrastructure Library (ITIL) has developed into a de facto standard for IT service management [14]. This is corroborated by the rapid increase in membership of the IT Service Management Forum, which is an interest group enhancing and propagating the ITIL principles [9]. Also the large number of practice-oriented ITIL conferences, publications and training opportunities [7; 8; 9; 14] indicate the growing relevance of ITIL. Recent studies substantiate that the ITIL holds a position of high relevance as well as being utilized extensively in the everyday running of German companies [12]. In spite of its relevance, its wide distribution and a large number of publications, a critical analysis of the ITIL reference model from a formal point of view is lacking. On the one hand existing literature is content to simply describe the areas of IT management as documented in the ITIL [3; 6; 8; 9; 11] and on the other it makes suppositions about the general usefulness of the ITIL in practice [7; 13]. The authors know of only a few publications on ITIL in scientific journals [11; 14]. This results in uncertainty in the execution of ITIL projects and misunderstandings regarding the attainable advantages of adapting ITIL.

In the present work we introduce the full concept of a Service Level Management System that should be built according to the ITIL standards, requirements and principals. We will start with the global Service Level Management Concept, presented earlier. Then we will make our proposal for the architecture of the SLM – System and describe its main activities. A deeper look of a measurement tool located at the customer premises will be presented.

II. THE SLM CONCEPT

An important advantage of the proposed SLM concept [16] is that it provides tight integration across all solutions, permitting them to work together to benefit from - and build upon - the specific capabilities of each (Figure 1).

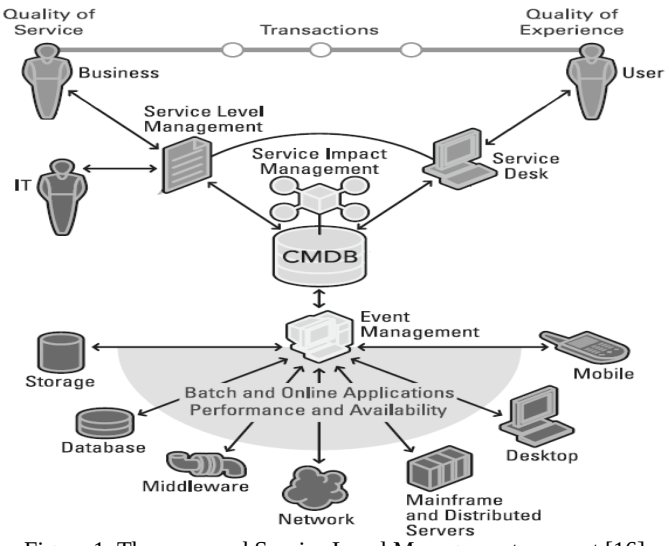


Figure 1. The proposed Service Level Management concept [16]

In the table below the synergies between the Standard Management elements and the functions of the proposed SLM concept are shown.

[1]Todor Georgiev is with the TELELINK EAD, Business Park, Building 13, Sofia E-mail: todor.georgiev@telelink.com

[2]Aleksandar Tsenov is with Telecom Department at Technical University of Sofia, "Kliment Ohridsky" Blvd 8, 1756 Sofia, Bulgaria, E-mail: akz@tu-sofia.bg

Standard Management Elements	Proposed SLM Functions
Incident and Problem Management	▪ Tracks help desk response and resolution times and compares them with SLA (Service Level Agreement) commitments. ▪ Generates alerts and notifications to the support staff when support SLA's are in danger of being missed. ▪ Provides the help desk with service impact information to assist in determining incident priorities and to facilitate root cause analysis.
Change and Configuration Management	▪ Tracks SLA availability targets to ensure that change tasks and requests are performed in order and on time.
Service Impact and Event Management	▪ Delivers business – aware information about the real-time state of services.
Capacity Management and Provisioning	▪ Enables the analysis and predic-tive modeling of potential IT configu-ration changes and their effect on ser-vice levels.
Asset Management and Discovery	▪ Measures availability targets for specific assets and ser-vices, and shows the latest calculated availability metrics for the specified items in the CMDB (Common Management Database) to help ensure that critical business assets or services maintain committed levels of availability.
Infrastructure and Application Management	▪ Uses infrastructure and application data for service level measuring of both infrastructure and applications.

III. Proposed Functional Architecture of a Service Level Management System

The proposed functional Architecture of a Service Level Management System is shown on fig. 2.

The following section introduces the main elements of the system and their functionality. The system is named NOC (Network Operation Centre). It is build upon the proposed SLM concept and the proposed measurement tool, both described above.

1) ***What's Up Gold MSP Edition***, monitoring system provided by IPSwitch, with managed services based platform allows monitoring of multiple customers and helping one of the most important processes of the remote management.

It has a distributed architecture with a clustered Central servers and remote probe servers located at customer premises. The remote probe is polling all customer devices through ICMP (Internet Control Message Protocol), SNMP (Simple Network Management Protocol), WMI (Windows Management Instrumentation), Syslog etc. respectively all events and alarms are sent to the Central server in the NOC, through secure SSL (Secure Sockets Layer) connection where it triggers the Incident Management procedure. Depending on alarm priority the event can automatically open a trouble ticket in the Incident tracking system – Cerberus, as well as send notifications through different medias e-mail, sms, pager etc.

More than 150 customizable reports are available for the MSP (Managed Services Provider) and Customer so different statistical data, SLA, and problem areas are easily tracked and identified.

Monitoring system looks for different attributes in the network and sends notifications and alarms when disrupted:

- Availability
- Performance (CPU, Memory, Interface)
- Environment (Temperature, Humidity – depending on sensors)
- User experience (delay, jitter, packets loss)

2) ***Cerberus,*** An Incident tracking system located at the NOC, has the purpose of register all incidents related to supported customers and keep track of updates and progress on the incident management process. It provides flexible web access for NOC engineers and customer IT personnel as well as notification through e-mails. The system helps identifying different SLA KPI's (Key Performance Indicator) like MTTF (Mean Time To Failure), MTTR (Mean Time To Restore) and etc.

Key features of the system ensuring the Incident and Problem Management Processes are:

- Technical escalation
- Administrative escalation
- Multiple queues
- Role Based Account access
- Notifications

3) ***Rancid*** is an open source tool for automatic configuration collection from Cisco based devices.

Located at the NOC the application is configured to login to fetch device configurations 4 times per 24 hours over secure SSH (Secure Shell) connection. It then differs and keeps the configurations in a "configuration tree" where it's easy accessed by NOC engineers through web access. Every change is registered with information of the user account performed the change.

The tool ensures Configuration and Change management processes and is also very useful in recovering lost configurations.

4) ***Share-point Portal***, is the corporate server where customer information is stored in a secure way. Access is controlled by the embedded RMS (Remote Monitoring System) and is only allowed for authorized personnel. Specific information like customer assets, inventory, technical documentation etc. is available at any stage of the Managed services process. It helps all procedures including Configuration and Change Management, Capacity planning, Security management.

IV. Remote Monitoring Probe

The distributed Probe-Central architecture is very useful for large networks where access to parts of the network is difficult due to topology restrictions, overlapping address spaces, non-routable addresses and etc.

It's also very useful measurement tool, the SNMP polling would allow measuring different types of utilization, like

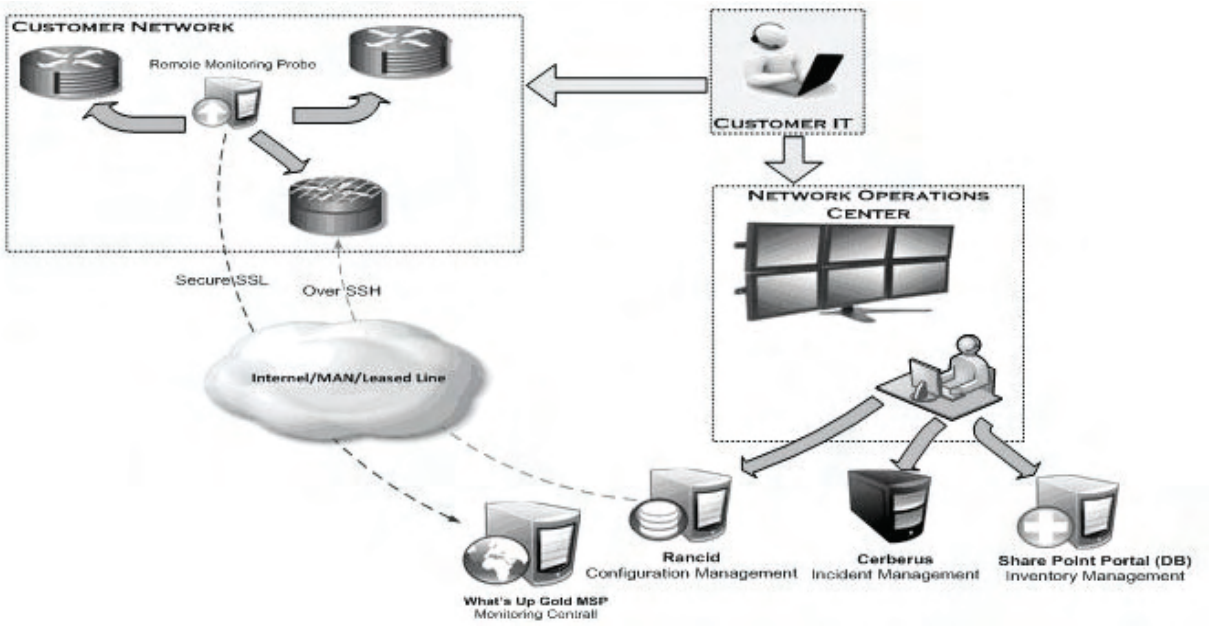


Figure 2. Functional Architecture of a Service Level Management System

bandwidth consumption with maximum, minimum, average and current values. Keeping these values in a SQL database would allow statistical data for a long period and consequent capacity analysis and planning, an important part of the ITIL process run in the NOC (Capacity Planning).

In this section an overview of the Remote Monitoring Probe function will be presented. This function is the most significant function in the customer premises. It has the goal to provide the measurement procedures near to the customers, according the customer network architecture and functionality. After the measurements are taken, the information shouldbe transferred over secure SSH channels through Internet to the Network Operation Centre.

In previous work [15] we proposed an appropriate sampling strategy and the structure of the measurement tool for completing the measurements.

A. Main Usages of Internet Measurements

The main usages of Internet measurements are Internet topology measurement, workload measurement, performance monitoring and routing measurement.

All of the mentioned above measurements can be performed in two ways – passive and active measurements.

The IETF's IPPM (IP Performance Metrics) has developed series of standards called Requests For Comments (RFC) on network performance measurements. The standard metrics for measurements are as listed below:

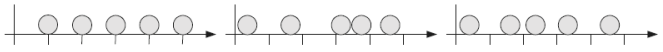
- Metric for Measuring Connectivity;
- A One-way Delay Metric;
- A One-way Packet Loss Metric;
- A Round-trip Delay Metric;
- One-way Loss Pattern Sample Metric;
- IP Packet Delay Variation Metric.

B. Sampling Techniques

In this chapter, three conventional sampling techniques, i.e., systematic sampling, random sampling and stratified sampling, and their characteristics are introduced.

Then a new sampling technique called "adaptive sampling" is presented.

Figure 3 illustrates these three sampling techniques.



a) systematic sampling b) random sampling c) stratified sampling

Figure 3. Sampling techniques

Systematic sampling generates sampling traffic according to a deterministic function. Random sampling employs a random distribution function to determine when a sample should be generated. Typically the samples are generated according to a Poisson process. With random sampling, an unbiased estimate of the QoS metric can be achieved. However, the entirely random nature of the sampling process may also cause the undesirable effect that sampling intervals are not uniformly distributed, and therefore the network may not be sampled for a rather long time. Stratified random sampling combines the fixed time interval used in systematic sampling with random sampling.

The proposed sampling method for the measurements inside the Remote Monitoring Probe Function is called adaptive sampling. In conventional sampling, the sample selection procedure does not depend on the observations made during the sampling, so that the entire samples may be selected prior to the start of the sampling process. In adaptive

sampling, the procedure for selecting samples may depend on the values of the variable of interest observed during the

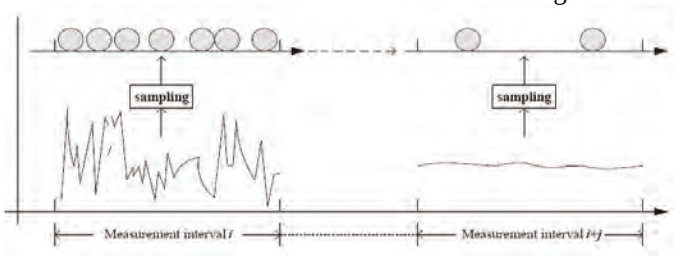


Figure 4. Example of adaptive sampling

sampling process. The primary purpose of adaptive sampling design is to take advantage of population characteristics to obtain more precise estimates, for a given sample size or cost, than is possible with conventional designs. Figure 4 shows the adaptive sampling in two measurement intervals.

Architecture of the Measurement Tool - MT System

MT should consists of two kinds of systems: (a) Control System (CS) and (b) Measurement System (MS). Fig. 5 describes the architecture of MT.

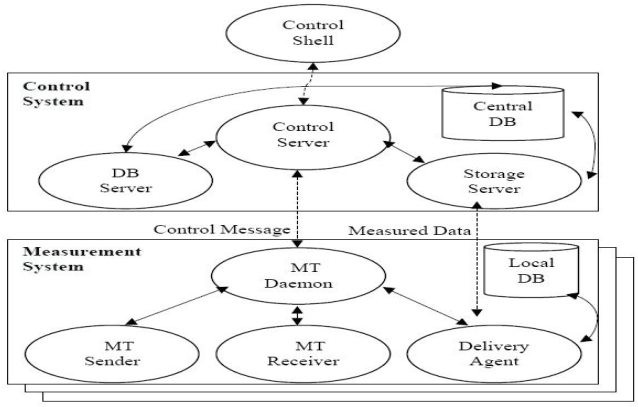


Figure 5. Architecture of the proposed MT

Control System (CS): CS, main system of MT, receives commands sent from Control Shell (CSH), with which operator controls and manages AMT.

Measurement System (MS): MS has four processes: (a) MT Daemon (MTD), (b) MT Sender (MTS), (c) MT Receiver (MTR) and (d) Delivery Agent (DA).

Procedure of Measurement

Step 1. Initialization of MTD for measurement
Step 2. Fork of measurement processes
Step 3. Establishment of control channel
Step 4. Confirmation about readiness from MTD
Step 5. Start of measurement
Step 6. Start of actual measurement
Step 7. Injection of measurement packets
Step 8. Storing of measurement records

In order to implement the proposed measurement tool into the Customer Network some changes and additions should be made. The changes and the additions concern not only the architecture of the measurement tool, but the measurement procedure too.

First: the MS should be located at the routing equipment of the customer network; second: an interface to the CS for transfer over customer LAN should be developed; third: a new interface in the Control Shell that supports SSL connection to the Network Operation Centre should be developed as well.

New procedure step – filtering and sending the information to the NOC should be also performed.

V. Conclusion

In this paper a full architecture of a Service Level Monitoring system is introduced. The main functionality of the system is defined. A closer look to the measurement tool and some of its functions are described as well. The future work in this direction will include the realization of the mentioned above necessary functions and interfaces in order to fulfill the whole functionality of the measurement process at the customer premises.

References

[1] The Official Introduction to the ITIL Service Lifecycle, 2007.
[2] Guldentops, E., G. Hardy, J. Heschl, R. Stroud, Aligning COBIT®, ITIL® and ISO 17799 for Business Benefit. A Management Briefing from ITGI and OGC, 2005
[3] Hoekstra, A., N. Conradie, CobiT, ITIL and ISO17799. How to use them in conjunction, 2002
[4] Curtis, B., Integrating CMMI® with COBIT® and ITIL®, 2005
[5] Defining, Modeling & Costing IT Services. Integrating Service Level, Configuration & Financial Management Processes, September 2004
[6] Computer Associates' Delivering. Best Practice Support ITIL, ISO 17799, BS15000, White Paper
[7] Symons, C., IT Strategy Maps: A Tool For Strategic Alignment by BEST PRACTICES November 21, 2005
[8] Delivery of ITIL Best Practices using Computer Associates management solutions
[9] Nabiollahi A., Considering Service Strategy in ITIL V3 as a Framework for IT Governance, The IT Service Management Forum, UK , 2007
[10] Brown A.B., A. Keller, A Best Practice Approach for Automating IT Management Processes, IEEE Publishing, 2006
[11] Jantti1M., A. Eerolal, A Conceptual Model of IT Service Problem Management, IEEE Publishing, 2006
[12] "ITIL Das Munich Institute for IT Service Management – mITSM", available on http://www.mitsm.de/
[13] "ITIL Forum 2008 - Neupositionierung und Wertorientierung", available on http://www.itil-kongress-iir.de/
[14] "ITIL process procedure templates", available on http://www.metocube.com
[15] Tsenov Al., T. Georgiev, "SLA Monitoring System of QoS Parameters to Network Performance Metrics Mapping", International Scientific Conference ICEST 2008, pp. 36 - 39, Nish, Serbia, June – 25-27, 2008
[16] Georgiev T., Al.Tsenov „Modelling Service Level Monitoring Processes for QoS Guarantee of Managed Services", International Scientific Conference ICEST 2009, Vol. 1, pp. 135 - 138, V. Tarnovo, Bulgaria, June – 23-25, 2009
[17] Tsenov Al., Todor Georgiev, Development Of Platform And Approach For Analysis And Estimation Of Events In Telecommunication Networks, TELECOM' 2009 08-09 October, St. Constantine, Varna, Bulgaria (in Bulgarian)

A Modified Cascade Model of a GSM Stream Cipher

Elitsa Gospodinova

Abstract **- This paper presents an optimized method for improving the A5/1 stream cipher encrypting the line from the mobile terminal to the base station in the GSM standard. The method realization is master-slave-based on standardized algorithms, with the realized alternative ciphers being serially applied in rotation with a large non-linearity period. The A5/1 stream cipher modification is a convenient and practical method, minimizing the modification on the software of the end terminals.**

Keywords: security; GSM; stream cipher; A5/1;

I. INTRODUCTION

The advances in computation technology increase the efficiency of the existing kinds of cryptoattacks, and create a flourishing base for developing new kinds. Thus, the contemporary cryptographic algorithms become weaker and less reliable for ensuring the security and confidentiality of the transmitted information. From the cryptoattack viewpoint the radio air is a well protected zone, but from the undesired access viewpoint, it is the weakest component.

The modern broadband and the future more attractive services increasingly attract the interest of persons and organizations wanting to violate the confidentiality of the transmitted information. The security of the wirelessly transmitted data in cellular networks is achieved through encrypting the digital streams to and from the mobile terminal. In GSM communications the encrypting algorithm is A5. It processes all the messages – voice, text and video stream. The A5 algorithm has to be common for all GSM operators, end terminals and base stations, in order the roaming service to be applicable. The A5 realization is twofold – in the end terminal, as well as in the base station.

The GSM specifications allow up to 7 algorithms based on the A5, as well as a non-encryption mode. The earliest realizations A5/1 and A5/2 are widely implemented to date.[1]

- A5/1 – reliable algorithm used in Europe and the Americas;
- A5/2 – "artificially weakened" algorithm used in Eastern Europe and Asia;
- A5/3 – the newest version of the algorithm. It is a stream cipher generating two 114-bit keys and is based on the block cipher KASUMI.

The development of A5/3 is aimed at its application in 3G. It is implemented and standardized [6],[7],[8],[9] for usage in existing cellular networks as well but is not used to the moment.

Elitsa Gospodinova is with the Technical College at Asen Zlatarov University, 8010 Burgas, Bulgaria, E-mail: ed_gospodinova@abv.bg

The weaknesses of the A5 implementations and the many successful attacks [2],[5] require improvement of the security level but the established position of the acting operators, as well as the necessity of large investments make the modification of the infrastructure equipment very unlike, since over 100 million users worldwide would be affected.

Of a primary importance in the cellular networks is the delivery of the offered services without excessive time delay, with offering newer and attractive services demanding large computational resources. This makes it necessary to look for alternative approaches to security improvement, demanding the fewest possible intrusions in the software of the cellular network. In principle, the multiple encryption of the data stream or the series interconnections of ciphers are logical ways to security improvement of a given wireless network since they can be based on standardized and specified algorithms. Their drawbacks such as the time delay for the multiple encryption and need for a specified number of implemented algorithms for the series encryption, make them inapplicable for the cellular networks.

The paper is organized as follows. It begins with a description of a model for generating of ciphers, alternative to A5/1, applied on a rotational principle. Next, new algorithm controlling the sequence of the applied alternative ciphers is presented. After that, conditions for the method realization are noted, as well as its advantages compared to the standard A5/1 are given.

II. CASCADE MODIFIED MODEL OF THE A5/1 STREAM CIPHER

A5/1 is a stream cipher composed of three linear feedback shift registers (LFSR) having 19, 22 and 23 bits, respectively [3]. The most significant bits (MSB) of the LFSRs are modulo 2 summed and determine the output of the generator. The feedback of each register is obtained as a modulo 2 summation of specific bits called tap bits, placed in positions:

$$13, 16, 17, 18 \Rightarrow R1;$$
$$20, 21 \Rightarrow R2;$$
$$7, 20, 21, 22 \Rightarrow R3$$

The cascade will be obtained by changing the positions of the tap bits which will lead to a given number of alternative ciphers of A5/1, with their sequence controlled by the generated by A8 session key (MASTER - SLAVE). For the model description a cascade model of eight modifications of the A5/1 cipher is chosen.

The time-division approach is applied, resulting in obtaining separate time slots. Within each of them the encryption/decryption operation will use different modification of the A5/1 cipher. The different modifications used in the separate time slots are not applied in series but on

a rotational principle. Their sequence is determined according to the state table (Table 1). It visualizes which cipher for which binary combination is applied.

TABLE I
STATE TABLE

SESSION COMBINATIONS	ALTERNATIVE CIPHER	TAP bits									
		R_1				R_2		R_3			
		X_1	X_2	X_3	X_4	Y_1	Y_2	Z_1	Z_2	Z_3	Z_4
000	A5/1/1	13	16	17	18	20	21	7	20	21	22
001	A5/1/2	12	16	15	18	19	21	6	19	21	22
010	A5/1/3	13	15	16	18	18	20	8	19	20	21
011	A5/1/4	11	14	15	17	18	19	6	18	20	22
100	A5/1/5	13	15	17	18	19	20	7	18	19	21
101	A5/1/6	14	13	17	18	17	21	9	18	20	21
110	A5/1/7	12	15	16	17	18	21	8	17	21	22
111	A5/1/8	11	14	16	18	17	20	6	17	20	22

Since a cascade model of eight A5/1 modifications is chosen, the binary code in the state table is required to contain three information bits – session combinations (S_c). The session combinations are fetched serially from the generated by A8 session key (S_k). The S_k has a standard 64-bit length corresponding to 21 whole bit strings. Upon entering the next time slot (in a working state of the system) the current combination is compared with S_c in the state table. This process is repeated until running out of the contained three-bit strings from S_k (Fig. 1).

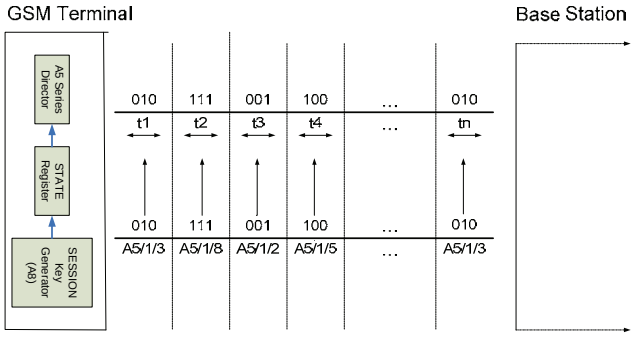


Fig. 1 Cascade modified model of the A5/1 stream cipher

The set of combinations (21) correspond to one whole rotational cycle. It is convenient for the generated by A8 session key to be stored in a specific state register. Upon call initialization between a GSM terminal and a base station the session key is copied into the state register before sending. The three-bit strings corresponding to the sequence of the applied ciphers are fetched from the state register (RS) and have the following sequence:

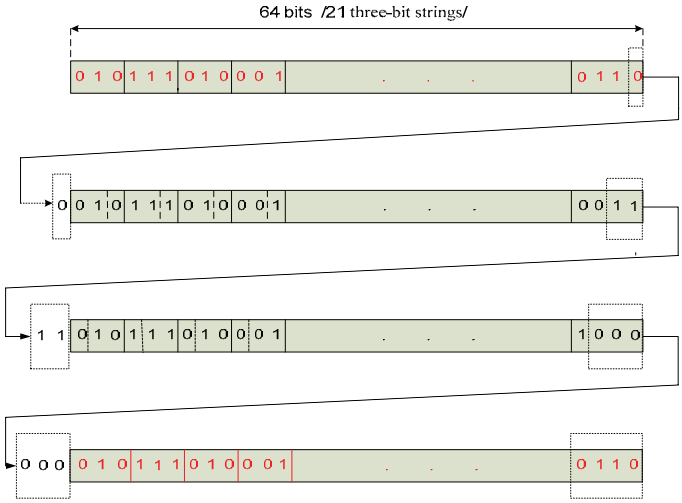


Fig. 2. Algorithm for controlling a rotational sequence of modified ciphers

The bit sequences fetched by the session key and controlling the sequence of rotation of the encryption algorithms are visualized in Fig. 2, based on which the following claim may be constructed:

A series "x" has an equal bit sequence with another series "y" when the following condition is met:

$$M_x - M_y = Z \qquad (1)$$

where:
M – number of bits carried over to the next cycle;
Z – number of information bits in the string, corresponding to one algorithm in the state table;

The expression in Eq.(1) shows same bit sequences generated by the A8 session key in its standard form.

In the particular case Z=3 which leads to a high degree of repeatability of the bit sequences, hence the repeatability of the rotational sequences. This drawback can be overcome by modifying the algorithm as follows. The total length of 21 three-bit strings is equal to 63 bits, and S_k is 64 bits long (Fig. 2). The remainder – 1 bit is the most significant bit (MSB) in the string during the first cycle. It is carried over as the least significant bit (LSB) in the register during the next cycle. It follows that the sequence is 1+64 bits long. In order to keep the same cycle length, the 65th bit is not processed (it is eliminated), and the preceding bit is processed according to the described procedure. The carried over MSB in each next cycle determines the non-repeatability of the rotational sequence of the ciphers (Fig. 3).

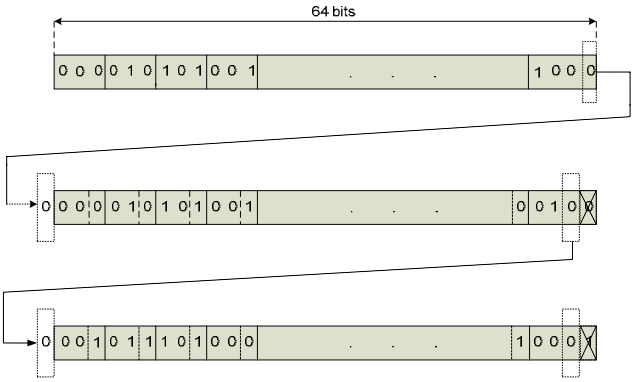


Fig.3 Improved algorithm for controlling a rotational sequence of modified ciphers

$$P_k = S_k^2 + S_k \tag{2}$$

$$N = \frac{S_k - M}{Z} \tag{3}$$

The algorithm for controlling a rotational sequence of modified ciphers may be expressed recursively as:

$$f^0(x_n) = (x_n) \tag{4}$$

$$f^1(x_n) = x_n div 2 + 2^{n-1}(x_n \bmod 2) \tag{5}$$

$$f^m(x_n) = f^{(1)}\left(f^{(m-1)}(x_n)\right) \tag{6}$$

where:

P_k – repeatability of the bit sequences, $[bit]$;

S_k – length of the session key, $[bit]$;

N – number of bit combinations for one rotational cycle;

x_n – key sequence, decomposed into three-bit strings;

n – number of bits in one sequence;

m – number of sequences;

The described algorithm for $2^3 = 8$ ciphers might be applied for 2^7 and 2^9 ciphers as well, with the sequence of the modified ciphers controlled by the MASTER-SLAVE based algorithm presented in Fig. 4.

III. Software realization of the method

For maximum efficiency of the method, the software realization should be optimized with assembly language code. The A5/1 stream cipher is extensively used; the method does not require implementations of new algorithms, but slight changes in the A5/1 software code, with the tap bits being input into the code not as constant values, but as variables, respectively: X_1, X_2, X_3, X_4; Y_1, Y_2; Z_1, Z_2, Z_3, Z_4; During each next cycle their values corresponding to the particular combination generated by the A8 algorithm would be read from the state table. This process guarantees that the data streams are processed with different ciphers for each cycle. Although the state table is static, the control of the sequence of the ciphers is generated by the A8 algorithm which is unique for each session. In practice this guarantees a non-predictable sequence of rotation of the ciphers.

IV. Analysis of the advantages of the method

The many analyses of the A5/1 cipher lead to the conclusion that it has weaknesses determining it as vulnerable to different kinds of attacks [2],[4].Its major weaknesses are the small number of inner states, short key, as well as the linear initialization of the key and the frame counter. The methodology of the present approach aims at avoiding the particular weaknesses. By introducing variable values of the tap bits, the number of inner states of the cipher is increased, that is, it is obtained of eight modified ciphers, with the pseudorandom sequence of rotation of the ciphers being 65 series × 21 three-bit strings = 1365 strings. In addition, the standardized A5/1 consists of three LFSRs, with a total of 10 tap bits, with modulo two summations, forming the feedback of each register. With the MASTER–SLAVE method the tap bits are fetched from the state table and are different for each of the eight alternative ciphers. The complex length of the key with a rotational sequence of the ciphers controlling the A8 algorithm is 87360 bits. A basic advantage of the MASTER–SLAVE method compared to the standardized A5/1 is also the fact that the order of the alternative ciphers is controlled by the A8 algorithm, that is, when an unauthorized access is attempted, data are necessary for the standardized A5/1, the alternative ciphers along with their corresponding tap bits, as well as the A8 algorithm with its corresponding bit sequence for the particular session, which increases the resistance of the method against attacks and overcomes its weaknesses. The slight software intrusions necessary for the realization of the method gives the opportunity the computing resources to be harnessed mainly for providing high-quality broadband services with increased security level of the transmitted data.

V. Conclusion

The described method is very convenient, does not require significant software intrusions into the build infrastructure worldwide. It is based on standardized and certified algorithms and its deployment does not require changes in the authorization and encryption procedures. From the viewpoint of data stream security level the method is comparable to the methods for multiple encryption and use of cascaded ciphers, avoiding the resulting time delay and the need of more powerful energy resources. The method provides an opportunity for a step-by-step deployment, a possibility for optimization for next generation networks. In the future, the method shall be analyzed based on all kinds of attacks for which the standardized A5/1 is known to be vulnerable.

ACKNOWLEDGEMENT

This paper was supported in part by project 213/2010 in the Department of Computer Systems and Technologies in the Faculty of Technical Sciences at Asen Zlatarov University, Burgas.

REFERENCES

[1] Aissi, S., Dabbous, N,. Prasad A., Security for Mobile Networks and Platforms, Artech House Universal Personal Communications, 2006

[2] Birynkov, A., Shamir, A., Wagner D., Real Time Cryptanalysis of A5/1 on a PC, Proc. of Fast Software Encryption –FSE 2000. pp. 1–18

[3] Boudriga, N., Security of Mobile Communications, CRC Press, 2010

[4] Barkan, E., Biham, E., Keller, N., Instant Ciphertext – Only Cryptanalysis of GSM Encrypted Communication (Technical Report CS – 2006 – 07 - 2006)

[5] Ekdahl, P., Johansson, T., Another Attack on A5/1, Abstract, Proceedings of International Symposium on Information Theory (ISIT), Washington, 2001 (http://www.it.lth.se/patrik/publications.html).

[6] Vijaya, C., Authentication and Access Control for Mobile Communications, (http://www.ittc.ku.edu/rvc/documents/865/865_securityreport.pdf.)

[7] European Telecommunications Standards Institute (ETSI), European Digital Cellular Telecommunications System (GSM), General description of a GSM Public Land Mobile Network (PLMN), 1997, (http://www.etsi.org)

[8] GSM 02.17 (ETS 300 509): European Telecommunications Standards Institute (ETSI), European Digital Cellular Telecommunications System, Subscriber identity modules (SIM), Functional Characteristics, (http://www.etsi.org)

[9] GSM 04.07 (TS 100 929): European Telecommunications Standards Institute (ETSI), European Digital Cellular Telecommunications System, Security Related Network Functions, (http://www.etsi.org)

Customer Experience Hierarchy Model

Aleksandar K. Tsenov

***Abstract* – Customer experience management (CEM) is a very powerful tool to improve customer value and maintain customer loyalty in the experience economy, while the identification for customer's experience requirements is the precondition of CEM. Following analyzing the hierarchy of CEM idea, the paper argues that customer's experience requirements can be summed up to some valuable dimensions based on which the model of customer experience dimensions can be set up. Two different CEM hierarchy models are being elevated and as result a possible approach to achieve and estimate the customer experience in the mobile telecommunications services is proposed. Some advice for CEM in telecommunication industry has been given based on the proposition for customer experience dimensions**

***Keywords* – Experience Hierarchy Model; Customer Value Hierarchy, Mobile Telecom Services.**

I. INTRODUCTION

With the evolution of telecommunications services and the improvement of the consumer's requirement level, experience and its value are getting to be realized gradually. At the same time, there are still some problems existing in the practice of Customer Experience Management (CEM).

Firstly, people don't have a systematic and consistent understanding on the concept of CEM, which influences the practice of CEM and its system info. CEM is mainly to manage the customer's total experience for enterprise and its products strategically. And it is a kind of customer relationship management pattern that emphasizes customer's participating and is dynamic and systemic. Identifying the different experience needs of the different customers is the precondition of CEM. Creating the total and consistent customer experience by using enterprise's resource such as product, service, surroundings, brand and employees on the customer contact points is the guarantee of CEM. But the previous research on CEM is mainly about experience marketing or branded experience design. This is a progress from the view of development, but limits the research range of CEM and neglects the effective identifying for customer experience needs and customer's experience feelings. All of these reduce the difficulties of research and practice on customer experience, but it makes the potential value of CEM unachievable. And even more, some enterprises that are anxious to pursue value and profit think that to set up "experience hall" or add the word of "experience" to advertisement is equal to the implement of CEM.

Aleksandar Tsenov is with Telecom Department at Technical University of Sofia, "Kliment Ohridsky" Blvd 8, 1756 Sofia, Bulgaria, E-mail: akz@tu-sofia.bg

Secondly, some people feel very pessimistic and are worrying about whether to manage customer experience. As the core content of CEM, experience is a psychological concept in the first. How to apply this psychological concept into business administration seems to be a question. If the enterprises would really take the experience of designing and transferring as their administration core, will they benefit from it? After all, all businesses take profit as their start point and final goal.

As far as the reason is concerned, the problem lies in the confusion on definition of CEM. The conception is the foundation stone which contributes to academic development. If CEM hasn't a clear definition, system info construction will be baffled. CEM is a neonatal conception, so it hasn't been recognized comprehensively. Beginning with analysis on customer value hierarchy model, the paper does some research on the CEM conception and its connotation deeply, and makes it much more concrete and structural with the hope of helping the implementation CEM.

II. REVIEW OF CEM CONCEPT

As the summarization of human behavior and mental feeling, experience came into being a long time ago. From the view of research development and its origin, experience firstly was used in psychology and philosophy, and subsequently got attention in economics and management science. Looking at development of customer experience conception, the research process and view can be concluded in the figure 1.

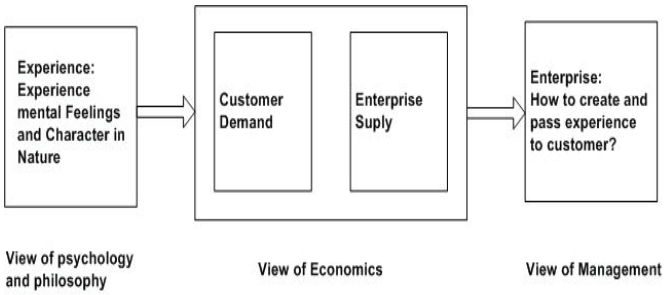


Figure 1. Research view on Customer Experience [1]

A. View of Psychology and Philosophy

Most of the research on experience is about peak experience now, and flow is one kind of it. Flow experiences offer absorption, personal control, joy, values, spontaneity, and a newness of perception and process [1, 2]. Based on the current research, experience conception in the view of psychology points to customer inner directly, and represents some essential characters. For example, experience refers to the human being's happy feeling when they reach a given

level in emotion, physical force, intelligence and even spirit [3]. Experience refers to the human being's inherent response to some stimulus during the process of watching or taking part in some activity, whatsoever it is true, dreamy or dummy [4].

B. View of Economics

There has been only culture meaning for experience in the long history which emphasizes its essence character. Just because added to economy activity and became the main source of economic return, experience was located in the important place in economic field and experience economy came into being. As an economic term, experience was firstly brought forward by Toffler 30 years ago [5]. He took experience as economic value and thought it was a psychological outcome produced from products or service. Pine II and Gilmore declared the experience economy was coming, and discussed how experience was produced [4].

C. View of Management Science

In the field of management science, researches on experience are mainly about experience supply, namely, how enterprises create experience for customers? One important kind of the research is about how participation of customers influences experience [6]. From the view of marketing [7] the experience is divided into sense, feel, think, act and relation by different ways of customer participation based on the theory of physiology, psychology and sociology. LaSalle and Britton [8] think customer experience as a kind of intercommunion between customer and product, company or its service agent, by which customer reaction can be motivated. If the reaction is positive, customer will feel the value of product or service. Pullman [9] argues that experience is an aggregate in which enterprise can communicate sense stimulus, information and emotion with customers. This aggregate emphasizes both the supplier and the demander, and the customer may be individual or another company.

III. THEORY OF CUSTOMER VALUE HIERARCHY

The enterprise that is willing to create experience for its customer should firstly understand fully the experience themes and relative creation activities. Customer value hierarchy which think consumer value achievement during the process of purchasing product or service comprise three layers, such as product attributes, outcome and final aim help to solve this problem [10]. The customer value hierarchy (CVH) is shown in figure 2.

Firstly, the layer of attributes is the lowest layer during the process of value achievement. On this layer the customers describe product with attributes. At first, customers contact and feel the attributes of product and service which describe the serviceability to customer.

Secondly, the layer of outcome lies in the middle of customer value hierarchy. In this layer the customers express the subjective judgment after using the product. The layer of outcome being helps enterprise to consider how to satisfy customer from the view of them or what outcomes will brought about to customers.

Thirdly, the layer of final aim is the basic driver force of consumer including person, family and enterprise. This layer expresses the core value of consumer. For example, the final purpose may include innermost calmness, comfort of life, sense of achievement, and so on. For enterprises, the layer of aim is the final object achieved by the means of product or service.

In a word, customers always define the achieved value from three layers, such as product attributes, outcome and final aim. It is the necessary demand for enterprises to improve customer value, especially the value of final layer in the new economic model that pay more attention to creating and transfer-ring of experience. At the same time, experience being as a particular economic supply substance, the realization of final aim also depends on the creation of customer experience value.

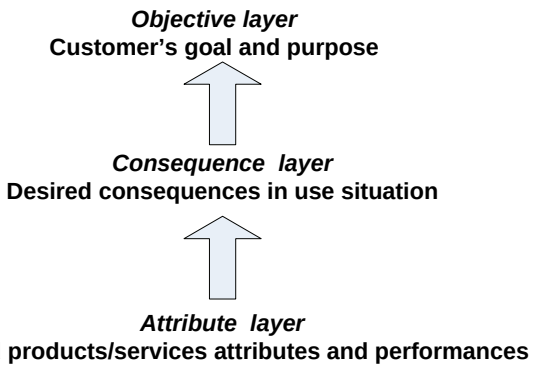


Figure 2. Customer Value Hierarchy [10]

IV. CUSTOMER EXPERIENCE HIERARCHY MODEL FOR MOBILE TELECOMMUNICATION SERVICES

Based on the theory of customer value hierarchy, the conception of experience can be divided into three layers, such as customer experience attributes, customer experience themes and customer experience effect. These three layers relate to the layers according the CVH. **The CVH Attribute layer** specifies the usage of mobile services. Customer experience attributes provide resource for achieving customer experience effect and transfers value experience in the layer of customer experience attributes to customers. **The CVH Consequence layer** represents the customer experience of mobile services. The corresponding customer experience themes are a bridge between customer experience effect and custommer experience attributes and transfers value experience in the layer of customer experience themes to customers. **The CVH Objective layer** includes the ultimate motivations of customers engaging in mobile telecommunication services. Customers may have multiple motivations in the objective layer. Customer experience effect is a kind of beautiful and happy feeling in consciousness when customer achieves a given level in emotion, physical force, intelligence and spirit during the interaction between enterprise and them, and transfers the highest layer of value experience to customers.

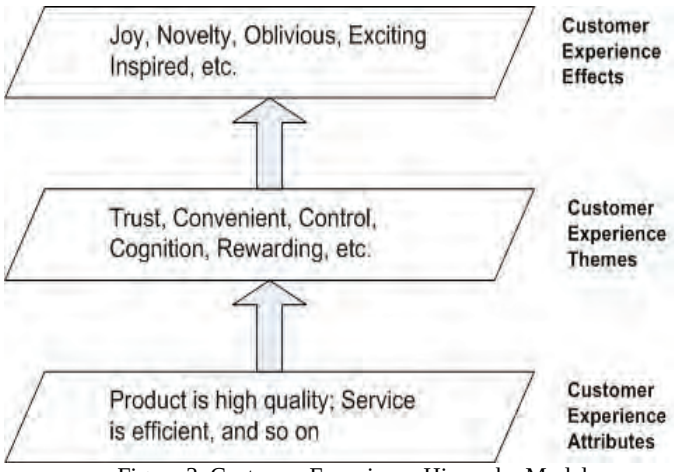


Figure 3. Customer Experience Hierarchy Model

A. *Layer of Customer Experience Attributes*

Relational context refers to two important types of interaction: (1) between the guest and service provider and (2) between the guest and the other guests. Firstly, as the representative of enterprise and the executants of business management, employees take the most important context and play a crucial role in the creation of customer experience. Secondly, the meaning of brand is not limited tangible name, logo and banner, and even more important, it symbolizes customer's value and life style which contributes to the perception of oneness with or belongingness to a collective, so influence the interaction between customer and the other customers. In a word, employees and brand must be considered in relational context.

Customer experience attributes should take experience design context as the core to feel and evaluate, and define experience according to the quality of them too. Experience attributes may be expressed with "the use of product of convenient", "the brand is full of rallying point", "the employees are kind", and so on. In a conclusion, customer experience attributes establish the detailed rules to implement customer experience based on the existing customer service resource, and ensure the needed resource to satisfy customer experience need and attain the ultimate effect.

B. *Layer of Customer Experience Themes*

Based on the previous research, we put forward some rules to identify customer experience by which customer experience requirement can be reduced to ten themes strategically. That is trust, convenience, realization of promise, respect, control, various choice, knowledge, cognition, rewarding life style, status. Every theme should represent a kind of mental intent of customer. Although different theme isn't irrelevant at all, they have a definite limit between them correspondingly.

Customer experience themes go beyond the product-centered. As the expression form as considered, the expression, such as "I feel the channel (that is the content) is convenient (that is customer's feeling)", is the basic form. The content to be felt is the common characters extracted from the experience attributes, such as trust, convenience, control, status, benefit to body and mind, and so on. The relation between customer experience themes and attributes can be described in figure 4.

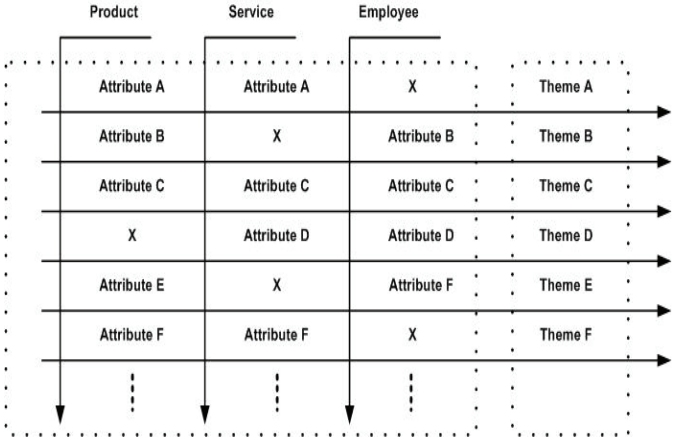


Figure 4. Relation between experience themes and experience attributes

Experience attributes coming from different experience design context can be used to create the same experience theme, and vice versa.

C. *Layer of Customer Experience Effects*

Customer experience effects are corresponding to custommer's mental feeling and they refer to a kind of beautiful and happy feeling in consciousness when customers achieve a given level in emotion, physical force, intelligence and spirit during the interaction with enterprises. This kind of happy feeling go beyond satisfaction and bring about pleasant surprise to customers. The layer of experience pays much more attention in psychological meaning of experience and emotional response of customers.

The peak experience above should belong to the layer of customer experience effects. Through transcending customers' satisfaction and providing much more pleasant surprise, customers' loyalty can be improved.

D. *Relationship among Three Layers*

The forming of total experience includes three layers: customer experience effects, customer experience themes, and customer experience attributes. Customer experience effects are the final goals of CEM, customer experience themes build a connection between customer experience effects and customer experience attributes, and customer experience attributes provide resource for achieving customer experience effects. The lower the layer is, the more concrete and more closer it is with enterprise; the higher the layer is, the more abstract and more closer it is with customers.

The three layers are interactional and interdependent and they will transfer different experience. Based on the three layers in customer experience hierarchy model, "themes" is consistent with the viewpoint of effects and themes above, while "activity" can be interpreted as transfer of experience attri-

butes based on experience context. Firstly, enterprises should ascertain experience effects and embody experience themes. Secondly, experience attributes transferring experience to customers should be listed according to enterprise's resource, and related attributes to the objective themes should be found. Finally, taking experience design context as the tool, all contact points must be integrated so as to make sure the realization of experience effects.

E. Determination of Mobile Customer Value Hierarchy Model

Based on the complete chain of customer value layers, the next step in the procedure is to shift the layers from the individual perspective to the aggregate perspective of a group of customers. We can accomplish this step by using association methods to find the association rules among attributes of different layers or by cumulating the "connection" times of two adjacent layer's attributes. Based on the mobile customer investigation, this paper constructs the mobile customer value hierarchy presented in figure 5.

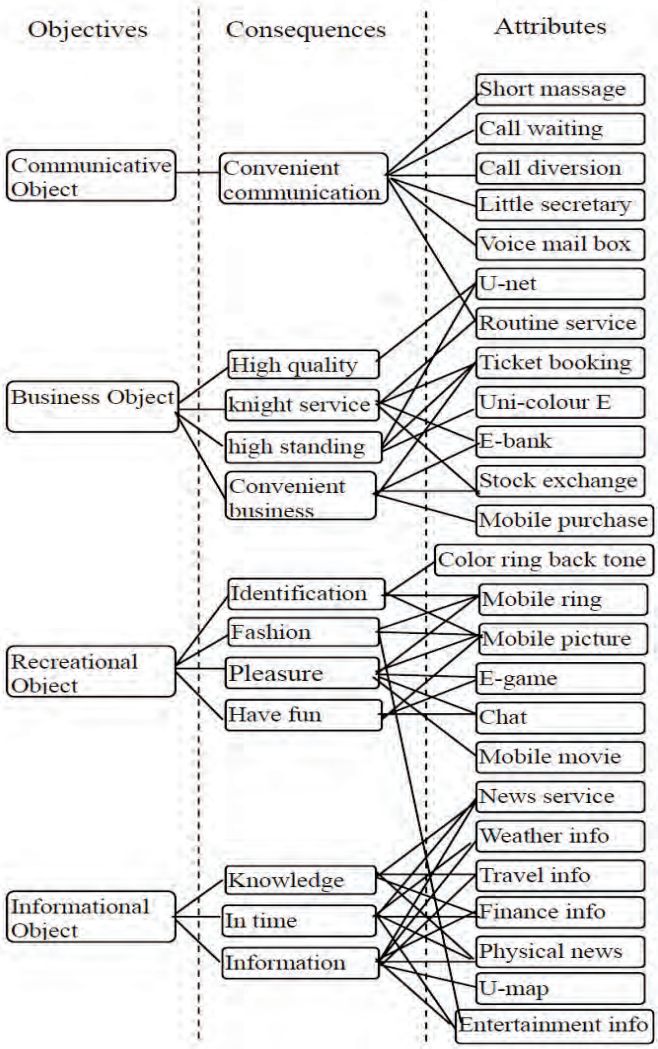


Figure 5: Mobile customer value hierarchy

The next steps in order to estimate the customer level of experience in the mobile communications are:

- Mobile customer demand analysis and the knowledge capture: thus includes constructing of an attribute-objective map and performing a Significant Attributes Analysis of the Customer Value Hierarchy. The significant attributes of customer value hierarchy are the key attribute variables of the attribute layer which distinctly correlate to the objective layer.
- Mobile Customer Demand Discrimination Modeling – this may be achieved by adopting a methodology based on adaptive computational architecture – for example neuronal network.
- Customer Demand Knowledge Capture.

These tasks can be performed in cooperation with the mobile operators in order to collect the necessary information about the customer's experience level.

V. CONCLUSION

Based on the background of the mobile industry, this paper proposes the framework and process of a mobile service recommender system that identifies potential customer's demands of unsought services/products by using customer demand analysis. This paper proposes a mobile customer demand analysis model and develops a mobile customer value hierarchy to capture customer demand knowledge.

A well-formed model should identify the customer demand objectives dynamically from their engagement record; then a personalized product recommendation based on the customer value hierarchy is to be done.

REFERENCES

[1] Csikzentmihalyi, M., *Flow: the psychology of optimal experience*. New York: Harper Perennial, 1991.

[2] Csikzentmihalyi, M., *Finding flow: the psychology of engagement with everyday life*. New York: Basic Books, 1997.

[3] Pine II, B. J., and Gilmore, J. H., *Welcome to the experience economy*, Harvard Business Review, 1998, 76(7-8): 97-105.

[4] Schmitt, Bernd H., *Customer experience management: a revolutionary approach to connecting with your customers*. Hoboken, New Jersey, John Wiley & Sons, Inc., 2003.

[5] Toffler, A., *Future shock*. New York, Bantam Books, 1970.

[6] Larsson, R., and Bowen, D. E., *Organization and customer: managing design and coordination of services*, The Academy of Management Review, 1989, 14(2): 213-233.

[7] Schmitt, Bernd H., *Customer experience management: a revolutionary approach to connecting with your customers*. Hoboken, New Jersey: John Wiley & Sons, Inc., 2003.

[8] LaSalle, D., and Britton, Terry A., *Priceless: turning ordinary products into extraordinary experiences*. Harvard Business School Press, 2003.

[9] Pullman, Madeleine E., and Gross, Michael A., *Ability of experience design elements to elicit emotions and loyalty behaviours*, Decision Sciences, 2004, 35(3): 551-578.

[10] Woodruff, Robert B., and Gardial, Sarah F., *Know your customer: new approaches to understanding customer value and satisfaction*, Blackwell Publishing Ltd, 2002.

Frequency Offset Influence on OFDM/MDPSK System Performance

Slavimir N. Stošović[1], Bojan R. Dimitrijević[2], Nenad D. Milošević[3] and Zorica B. Nikolić[4]

Abstract **- In this paper, basic characteristics of OFDM systems are presented. Characteristics of OFDM systems with MDPSK modulation in every channel are simulated. For simulation and analysis of OFDM systems performance we used simulation environment designed for this purpose. We stressed OFDM frequency offset. We analyzed influence of OFDM system parameters on system performance for various values of frequency offset, parameter *M* and channels number.**

Keywords **- orthogonal frequency-division multiplexing (OFDM), M-ary differential phase-shift keying (MDPSK), frequency offset, frequency synchronization.**

I. INTRODUCTION

Orthogonal Frequency Division Multiplexing (OFDM) is the special kind of multicarrier modulation, where we simultaneously transmit a block of data symbols on a group of subcarriers with frequency-division multiplexing. Within one OFDM symbol duration, each subcarrier is modulated with a data symbol using any conventional method, such as quadrature amplitude modulation (QAM), M-ary phase-shift keying (MPSK), M-ary differential phase-shift keying (MDPSK), as in a single-carrier system. The spacing between adjacent subcarriers is carefully selected so that each subcarrier is located on all the others' spectral nulls, and all the subcarriers are also packed as closely as possible. Because of this spectral orthogonality, the modulation symbols on all the subcarriers can be ideally recovered by sampling the received baseband signal at a rate which is the reciprocal of the intercarrier spacing followed by a fast Fourier transform (FFT).

In the OFDM system, the ISI problem caused by delay spread of dispersive fading channels can be overcome by inserting a guard interval between adjacent OFDM symbols. Doppler spread due to the motion of a terminal destroys orthogonality of OFDM signal and introduces interchannel interference (ICI) on the demodulated OFDM signal. ICI degrades the system performance and results in the irreducible error rate. In frequency-nonselective fading channels, equalization combined with channel estimation has been proposed to combat the ICI problem. In frequency-selective fading channels, a channel estimation technique is used to compensate for multiplicative distortions on the demodulated OFDM signals. At slow fading, MDPSK modulation in the temporal direction has good performance as phases of multiplicative distortions are assumed to be constant during consecutive OFDM symbol intervals.

The error performance of MDPSK/OFDM is degraded and the irreducible error rate is introduced at fast fading.

Data transfer with OFDM system has many advantages:

-High spectral efficiency due to nearly rectangular frequency spectrum for high numbers of subcarriers.

-Simple digital realization by using the FFT operation.

-Low complex receivers due to the avoidance of ISI and ICI with a sufficiently long guard interval.

-Flexible spectrum adaptation can be realized.

-Different modulation schemes can be used on individual subcarriers which are adapted to the transmission conditions on each subcarrier.

There are also certain disadvantages:

-Multi-carrier signals with high peak-to-average power ratio (PAPR) require high linear amplifiers. Otherwise, performance degradations occur and the out-of-band power will be enhanced.

-Loss in spectral efficiency due to guard interval.

-More sensitive to Doppler spreads then single-carrier modulated systems.

-Phase noise caused by these imperfections of the transmitter and receiver oscillators influence the system performance.

-Accurate frequency and time synchronization is required.

The popularity of OFDM systems is rising due to its ability to support high-data-rate transmission over time-variant multipath fading channels. OFDM transmission techniques have found applications in the two digital terrestrial broadcasting services - digital audio broadcasting (DAB) and digital terrestrial video broadcasting (DTVB) [1], [2]. OFDM is used in the standards for wireless 5-GHz local area networks (IEEE 802.11a in US and HIPERLAN in Europe) [3], [4]. Asymmetric digital subscriber lines (ADSL) based on OFDM technology are used to deliver high-rate digital data over existing plain old telephone lines (pots) [5]. OFDM can also serve as an alternative transmission method to digital European cordless telephone (DECT)-like digital cordless systems [6].

In this paper, we presented the performance of OFDM system with MDPSK modulation in every channel. For this purpose we designed a special simulation platform and analyzed the influence of OFDM system parameters on

[1]Slavimir N. Stošović is with the High Technical School of Professional Studies, A.Medvedeva 20, 18000 Niš, Serbia, E-mail: slavimir.stosovic@vtsnis.edu.rs

[2]Bojan R. Dimitrijević is with the Faculty of Electronic Engeneering, A.Medvedeva 14, 18000 Niš, Serbia, E-mail: bojan.dimitrijevic@elfak.ni.ac.rs

[3]Nenad D. Milošević is with the Faculty of Electronic Engeneering, A.Medvedeva 14, 18000 Niš, Serbia, E-mail: nenad.milosevic@elfak.ni.ac.rs

[4]Zorica B. Nikolić is with the Faculty of Electronic Engeneering, A.Medvedeva 14, 18000 Niš, Serbia, E-mail: zorica.nikolic@elfak.ni.ac.rs

system performance for various kinds of MDPSK mudulations. The performance of OFDM/MDPSK system in the presence of frequency offset is also showed.

II. System Model

Models of OFDM transmitter and receiver are shown in Figures 1. and 2., respectively.

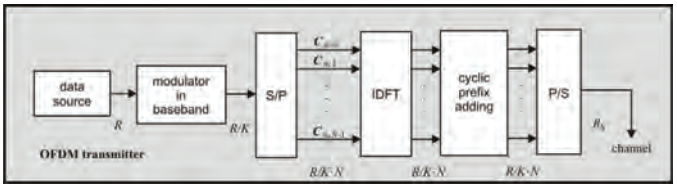


Fig 1. OFDM Transmitter

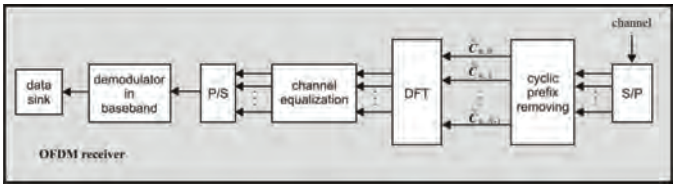


Fig 2. OFDM receiver

A. OFDM Transmitter

Typical OFDM transmitter consists of the following components:

Data source - Output is the sequence of uncorrelated random binary symbols ("0" i "1") with bitrate *R*.

Modulator in the baseband – Input is the sequence of binary symbols, and modulated symbols are at the output, chosen from MPSK, MDPSK or 16-QAM constellation, and the symbol rate at the output *R/K*. Modulator function is to map *K* binary input symbols to a modulated symbol according the signalization scheme.

S/P (serial to parallel converter) – output symbol rate is *R/K·N* and the frame of *N* symbols is formed.

IDFT (Inverse DFT) - *N* modulated input symbols are transmitted over a OFDM subcarrier. OFDM data frame of length *N* is at the output. Output symbol rate is *R/K·N*. OFDM modulator with *N* subcarriers is implemented using IDFT. Input of this block is in the frequency, and output in the time domain.

Cyclic prefix adding: OFDM data frame is at the input, and cyclic extended OFDM data frame with prefix consisting of the last *L* frame symbols (*L* is maximum delay in the channel). This is efficient against the multipath. Output symbol rate is *R/K·N*.

P/S (parallel to serial converter) – transforms parallel input data stream consisting of *N* symbols into serial stream. Output rate is $R_s = R(1+L/N)/K$.

B. OFDM Receiver

DFT: received and reconstructed OFDM data frame is at the input, and transmitted modulated symbols influenced by frequency channel response are at the output. In this case we use OFDM demodulator with *N* subcarriers and descrete Fourier transform. Input is in time, and output in the frequency domain.

Channel equalization: received symbols and frequency channel estimations on OFDM subcarrier frequency are at the input, and reconstructed modulated symbols are at the output. *Zero-forcing equalization* of OFDM subcarrier is performed here in the frequency domain. Only the subcarriers above the certain predefined threshold are equalized, in other cases they are unreliable (very low signal to noise ratio (SNR)).

Data Sink: reconstructed binary information is on the input. Here is performed processing typical for the received information.

III. OFDM System In The Presence Of Frequency Offset

In the case of ideal synchronization between transmitter and receiver, OFDM signal in the receiver, after FFT (Fast Fourier Transformation) block, can be represented as [7]:

$$Y_{i,k} = X_{i,k} H_{i,k} + N_{i,k} \tag{3}$$

where $X_{i,k}$ is a complex data symbol in baseband, $H_{i,k}$ is transfer function of *i*-th channel in *k*-th signalization interval and $N_{i,k}$ is noise in *i*-th subchannel in *k*-th signalization interval. Equation (3) shows that ideal synchronized OFDM system can be represented as a set of parallel Gaussian channels, meaning that OFDM enables that frequency selected channel is equivalent to the set of mutually orthogonal subchannels with frequency flat fading. However, in the presence of frequency offset, subcarriers orthogonality can be destroyed.

In the indoor case, Doppler offset can be neglected, but there is a frequency offset between transmitter and receiver oscillators. Frequency offset is mathematically described with frequency offset δf and phase offset θ in the lowpass equivalent signal. In the presence of frequency offset, signal in the receiver, after FFT block, can be represented as [7]:

$$\begin{aligned} Y_{i,k} &= X_{i,k} H_{i,k} \operatorname{sinc}(\delta f T) \cdot \\ &\exp\{j[\theta + 2\pi\delta f (kT_s + T/2]\} + N'_{i,k} \end{aligned} \tag{4}$$

where T is effective part of symbol duration (within guard interval),

T_s is full OFDM symbol duration, $N'_{i,k}$ is the overall noise of *i*-th channel in *k*-th signalization interval, including additive noise, due to ICI.

From equation (4) it can be seen, that frequency offset influences only phase change between two OFDM symbols, if frequency offset is constant. So, in the slowly varying flat fading channel, where channel and frequency offset estimation are periodically done, cumulative effect of frequency offset becomes very important, meaning that for every following OFDM symbol in the frame estimation error increases.

Frequency offset can appear because of the mismatch between transmitter and receiver oscillator or as the consequence of Doppler offset. Interference among subcarriers (ICI – Intercarrier Interference), due to frequency offset, degrades system performances, [7]. Therefore,

numerous methods for estimation and correction of frequency offset are proposed. Some of them use redundancy inherently built in every OFDM symbol, because of cyclic prefix usage, [8], [9]. Second group of estimation methods is based on use of special pilot sequences for frequency estimation offset [10], [11].

IV. Performance Analysis

Simulations of OFDM system were performed using specially developed environment at 2.4 GHz, without protective coding. Three different cases were simulated. In the first case, number of subcarriers (*N*) is 16, number of virtual channels (*VC*) is 2, number of channels used for cyclic prefix is 2 and number of channels used for guard interval (*GI*) is 2. In second case number of subcarriers is 32, number of virtual channels is 4, number of channels used for cyclic prefix is 4 and number of channels used for guard interval is 4. And finally, in the third case, number of subcarriers is 64, number of virtual channels is 8, number of channels used for cyclic prefix is 8 and number of channels used for guard interval is 8.

We used MDPSK modulation for each channel and tested OFDM system performances as a function of parameter *M* (level of DPSK modulation).

Figs. 3, 4 and 5 show OFDM system symbol error rate (*SER*) as a function of the energy per bit to noise power spectral density ratio (E_b/N_0), for different values of parameter *M*, in the presence of frequency offset, Δf=150 kHz (dashed lines) and without frequency offset (solid lines). The frequency offset is not estimated and corrected in simulations.

Fig. 3 shows OFDM system performances, in the first case (*N*=16, *VC*=*GI*=*CP*=2), versus E_b/N_0. From the Fig. 3, we can see that system performances depend of parameter *M* and decrease with increase of parameter *M*. Also, symbol error rate difference in the presence of the frequency offset compared to the symbol error rate without frequency offset increases with parameter *M* increase.

Similarly, Figs. 5 and 6 show performance of system in the second and third case versus E_b/N_0 with *M* as a parameter, respectively.

From the Figs. 3, 4 and 5, we can conclude that the frequency offset causes a performance drop compared to the ideal synchronizating system, and the performance drop rises with the increase of DPSK modulation level (*M*) and the number of subcarriers (*N*).

When the number of subcarriers and parameter *M* are greater, system performances decrease and the system stops operating, that correspondents to OFDM characteristics.

Fig. 6 shows symbol error rate versus frequency offset Δf, for system in the first case and *M* as a parameter. For smaller values of parameter *M*, frequency offset has less influence on the system performances. It means that, the band within it is possible to achieve satisfying transmission quality is the widest. With increase of parameter *M*, the influence of frequency offset on transmission quality also increases.

In the Figs. 7 and 8 we presented symbol error rate versus frequency offset for system in second and third case, respectively. Parameter *M* influences the symbol error rate in the same way as in Fig. 6.

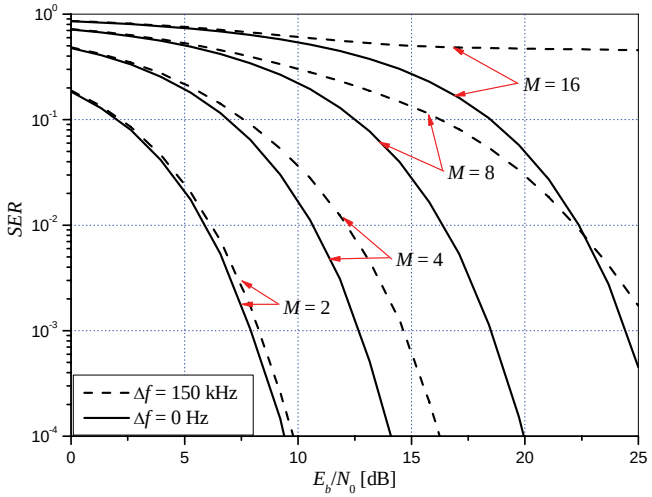


Fig. 3. Symbol error rate versus E_b/N_0 for *N*=16 and *VC*=*GI*=*CP*=2

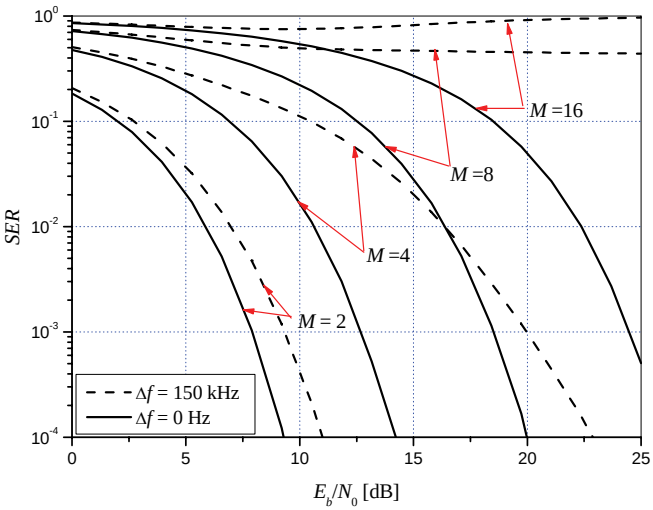


Fig. 4. Symbol error rate versus E_b/N_0 for *N*=32 and *VC*=*GI*=*CP*=4

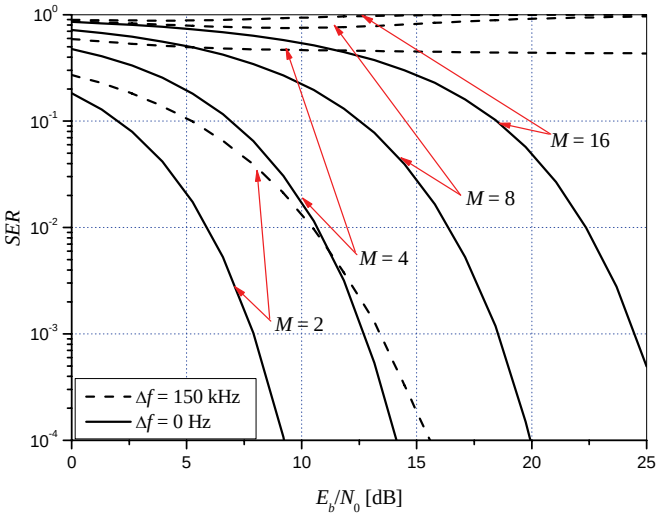


Fig. 5. Symbol error rate versus E_b/N_0 for *N*=64 and *VC*=*GI*=*CP*=8

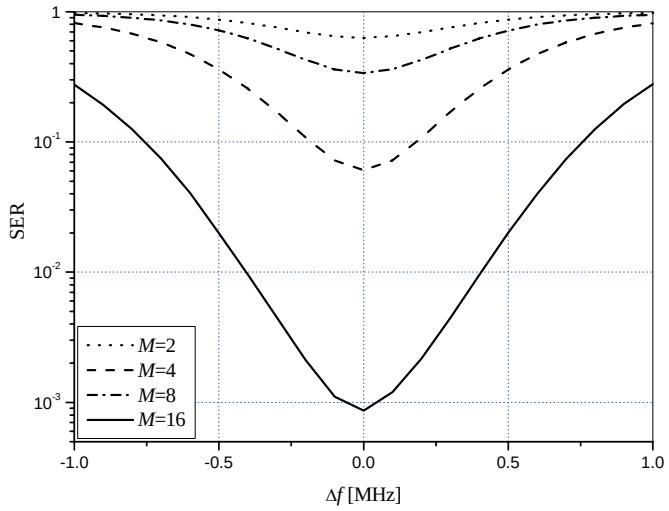


Fig. 6. Symbol error rate versus frequency offset Δ*f* for *N*=16 and *VC*=*GI*=*CP*=2

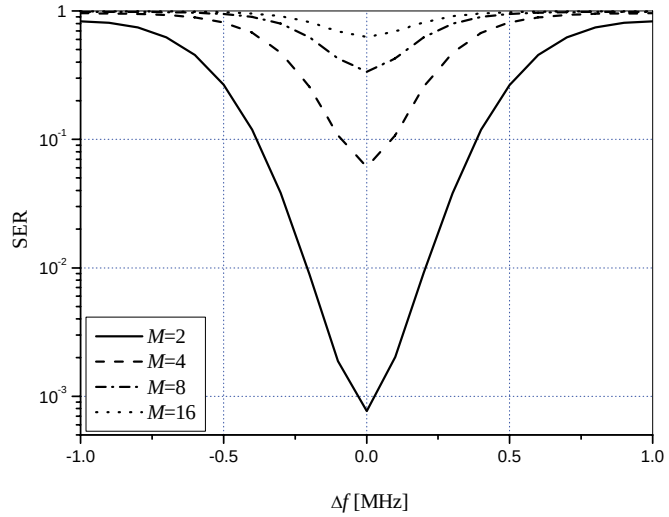


Fig. 7. Symbol error rate versus frequency offset Δ*f* for *N*=32 and *VC*=*GI*=*CP*=4

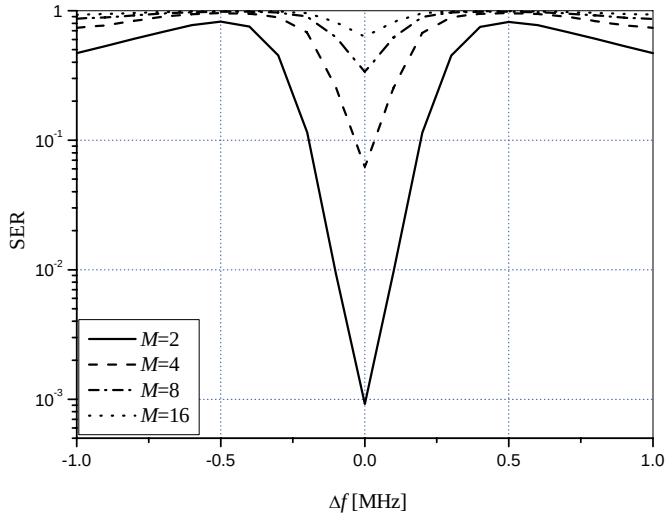


Fig. 8. Symbol error rate versus frequency offset Δ*f* for *N*=64 and *VC*=*GI*=*CP*=8

Parameter E_b/N_0 for the Figs. 6, 7 and 8 is 8dB. From the Figs. 6, 7 and 8 we can conclude that number of subcarriers influences the OFDM system performances with MDPSK modulation. Frequency offsets range where there is a satisfying transmission quality becomes narrower with the incerase of the number of subcarriers and the number of modulation levels.

CONCLUSION

Own simulation environment is developed and used for presenting the basic characteristics of OFDM system. Analysis of OFDM parameters influence to system performances for different values of frequency offset is performed. We also analyzed the influence of the modulation level of MDPSK modulation on the system performances in the presence of frequency offset. It was shown that in the presence of frequency offset the modulation level much influences the system performances. Because of this influence it is desirable to use estimation and frequency offset correction, and also to carefully choose values for the number of subcarriers and modulation level.

REFERENCES

[1] European Standard (Telecommunications Series), Radio Broadcasting Systems; Digital Audio Broadcasting (DAB) to Mobile, Portable and Fixed Receivers, ETSI EN 300 401 V1.3.3 (2001-05), 2001.

[2] European Standard (Telecommunications Series), Digital Video Broadcasting (DVB); Framing Structure, Channel Coding and Modulation for Digital Terrestrial Television, ETSI EN 300 744 V1.4.1 (2001-01), 2001.

[3] Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, IEEE 802.11a-1999, 1999.

[4] Technical Specification, Broadband Radio Access Networks (BRAN); HIPERLAN Type 2; Physical (PHY) Layer, ETSI TS 101 475 V1.2.2 (2001-02), 2001.

[5] Technical Specification, Transmission and Multiplexing (TM); Access Transmission Systems on Metallic Access Cables; Asymmetric Digital Subscriber Line (ADSL), ETSI TS 101 388 V1.3.1 (2002-05), 2002.

[6] A. Dhir, "Wireless Home Networks: DECT, Bluetooth, HomeRF, and Wireless LANs," Xilinx, White Paper 135 (v1.0), Mar. 21, 2001.

[7] T. Pollet, M. Van Bladel, and Moenceclaey, "BER sensitivity of OFDM systems to carrier frequency offset and Wiener phase noise," *IEEE Trans. Commun.*, vol.43, pp. 191-193, July 1995.

[8] R. Van Nee, R. Prasad, "OFDM Wireless Multimedia Communications," Artech House, 2000.

[9] J.van de Beek, M. Sandel, P.O. Borjesson, "ML estimation of time and frequency offset in OFDM systems," *IEEE Trans. Signal Process.*, vol. 45, No. 7, July 1997.

[10] P. H. Moose, "A Technique for Orthogonal Frequency Division Multiplexing Frequency Offset Correction", *IEEE Trans. on Commun*, vol. 42, No. 10, October 1994.

[11] Y. H. Kim, Y. K. Hahm, H. J. Jung, I. Song, "An Efficient Frequency Offset Estimator for Timing and Frequency Synchronization in OFDM System", *IEEE Pacific Rim Conf. on Commun., Comp. and Signal Proc.*, Aug. 1999.

Congestion Control Approaches in Wireless Environment

Toni Janevski[1] and Neslihan Ademi[2]

Abstract – **Modern communication networks using both wired and wireless communications including the Internet, are being designed for fast transmission of large amounts of data, for which congestion control is very important. Wireless networks are characterized by mobility, random changes in connectivity, fluctuations in channel and interference due to neighboring nodes. etc. TCP cannot provide efficient congestion control for wireless networks due to noncongestion losses. To alleviate the performance degradation of conventional TCP in wireless networks, many schemes have been proposed so far. This paper gives out a survey of proposed congestion control mechanisms in wireless environment and points the pros and cons of the wireless congestion control mechanisms, and evaluates their characteristics.**

Keywords – **Congestion Control, TCP, Wireless Networks**

I. Introduction

Communication networks are developing at an outstanding speed, with the evidence of mobile and wireless access networks. For accessing plentiful resources in the Internet through wireless mobile hosts, diverse wireless network standards and technologies have been developed and progressed significantly. The most successful examples include IEEE 802.11, Wi-Max for wireless networks and 3G-LTE/HSPA for cellular communications. Packet switching technologies have merged the traditional voice networks and data networks together into a converged integrated network.

All IP-based applications are the primary motivations to make these networks successful. Most IP-based networks rely on the Transmission Control Protocol (TCP) in the hosts to detect congestion in the network and reduce the transmission rates accordingly. In TCP/IP transmissions, the TCP congestion control operates well in the wired network, but it is difficult to determine an accurate congestion window in a heterogeneous wireless network that consists of the wired Internet and various types of wireless networks. The primary reason is that TCP connections are impacted by not only networks congestion but also wireless link errors. TCP/IP needs to depart from its original wired network oriented design and evolve to meet the challenges introduced by the wireless portion of the network.

Clearly, going forward, our network will become more heterogeneous in which protocols that react to different congestion signals interact.

About congestion algorithms existing literature generally assumes that all sources are homogeneous in that, even though they may control their rates using different algorithms, they all adopt to the same type of congestion signals. But when sources with heterogeneous protocols that react to different congestion signals share the same network as in proposed next generation networks (4G and beyond), the current congestion control framework is no longer applicable.

II. Challenges in Wireless Congestion Control

One of the more challenging environments for the Internet Protocol, and TCP in particular, is that of mobile wireless.

Compared with wired networks, one-hop wireless networks have some inherent adverse characteristics that will significantly deteriorate TCP performance if no action is taken. In essence, these characteristics include bursty channels errors, mobility and communication asymmetry.

III. TCP Modifications For Wireless Congestion Control

Among the various solutions proposed to improve TCP performance, there are four major categories: split-connection solutions, proxy-based solutions, link-layer solutions, and end-to-end solutions. The split-connection solutions attempt to improve TCP performance by splitting a TCP connection into two at the base station so that the TCP connection between the base station and the mobile host can be specially tuned for the wireless links. Realizing the base station is a critical point, approaches based on proxy put an implicit or explicit intelligent agent at the base station, detecting packet losses over wireless links and taking corresponding actions (such as duplicate ACK suppression and/or local retransmission) to ensure the TCP sender responds correctly. For the third category, a reliable link layer is built by adopting some link error recovery mechanisms, seeking to hide link errors from the TCP sender. Unlike the previous three classes, the end-to-end approaches enhance TCP by using SACK to quickly recover from multiple packet losses or by predicting incoming handoffs to avoid unnecessary congestion control invocation.

[1]Toni Janevski is with the Faculty of Electrical Engineering and Information Technologies, University of St. Kiril and Metodij, Skopje, Macedonia, E-mail: tonij@feit.ukim.edu.mk

[2]Neslihan Ademi is with the Faculty of Technical Sciences, International Balkan University, Samoilova 10, 1000, Skopje, Macedonia, E-mail: neslihan@ibu.edu.mk

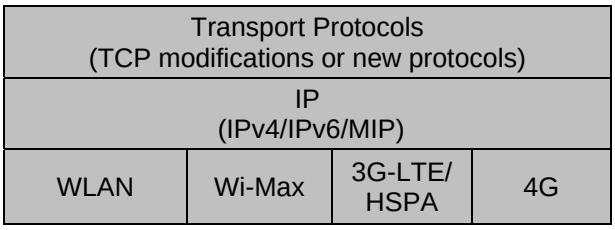


Fig. 1. Future Wireless Internet Protocol Stack

Indirect-TCP (I-TCP) [13] splits the connection on the wireless edge to protect TCP session from wireless media inconsistencies and losses. This results in two different flow and congestion controls working in wired and wireless sections separately and may result in serious inequalities on the two sides. In such attempt, I-TCP eventually violates end-to-end semantics of TCP as well. The M-TCP [14] is similar to I-TCP and splits the connection at super host. It differs only in a way, that super host does not generate acknowledgment (*Ack)* of last received segment until it receives the *Ack* from mobile node. In this way it too breaks end-to-end semantics of the TCP and disparities of two sides flow and congestion control mechanism still exist. Another problem that may arise is the source can take inaccurate decision about receipt of all data on the basis of last byte acknowledgement. If the acknowledgment of last byte is not received, the sender resends the all data which mobile node already received.

The TCP Westwood (TCP-W) and TCP Westwood + (TCP-W+) are true extension of TCP Reno (Reno). TCP-W uses the bandwidth estimation as a parameter to control the congestion window and slow-start (SS) threshold (*ssthresh*). The protocol separates the congestion control from the error control. This mechanism helps it to improve performance over lossy wireless links. Despite, some advantages, TCP-W is unable to present its preeminence over Reno when higher packet losses occur or the blackout time is longer. TCP-Peach (TCP-P) was developed for the satellite network to overcome long propagation delays and elevated link error rate. It uses two new algorithms; namely *sudden start* and *rapid recovery* to overcome the limitations of TCP in satellite network. It also uses dummy packet to estimate the available network resources. The successful delivery of the dummy segments indicates better resource availability and transmission rate can amplify. The TCP-Peach+ (TCP-P+) further improve the network utilization by sending Nil packet instead of dummy packet. It also replaces sudden start and rapid recovery algorithms by jump start and quick start. Jump start is extension of sudden start and it replaces dummy segment with Nil segment. TCP-P+ presents improved performance as compare to TCP-P.

TCPVeno (Veno) [20] introduces sender side solution, by combining the feature of the Reno and TCP Vegas (Vegas). The Reno and Vegas use reactive and proactive congestion control strategies respectively. Veno integrates the congestion detection mechanism of Vegas with Reno to distinguish between the congestive and non-congestive loss. Veno efficiently deals with single packet loss but it suffers form performance degradation when multiple packet losses occur in the network. The reason behind this degradation is the continuous reduction of window size on successive loss. Another sender side solution named; TCP-Jersey (Jersey) is proposed to improve the TCP performance in heterogonous wireless networks [21]. Jersey adopts the similar idea of TCP-W to estimate the bandwidth at sender side by monitoring arrival rate of *Ack*. Jersey uses the *Available Bandwidth Estimator (ABE*) at sender side to estimate the bandwidth and re-computes the *cwnd* on every *rtt*. It uses *ECN* for identifying congestion in the network. For further tuning of congestion control, Jersey also uses two parameters, namely; *intervals of jitter (IJ)* and *jitter ratio (JR)* to determine the ratio of congestion on the network. Freeze TCP (F-TCP) [27] proposed a recipient side solution which improves the performance of the transport protocol TCP in networks where handoffs and disconnections are more common and significant. If MN detects the low quality signal it sends *zero-window-probes (zwp)* with *Ack* to the sender. On the receipt of *zwp* sender freezes the timers and waits for positive *zwp*.

In [25], an inter-layer collaboration protocol for TCP (ILC-TCP*)* presented by introducing a sender side solution for mobile and wireless network. ILC-TCP introduces a new layer parallel to the network protocol stack named as *State Manger (SM)*. The SM communicates with core layers and notifies the status to network. It can handle the temporary disconnection. The ILC TCP is similar to the Freeze TCP except that it is sender side solution while Freeze TCP is a receiver side solution. So both protocol shares the same merits and demerits.

TCP Delayed Congestion Response (TCP-DCR) also improves congestion handling mechanism for wireless networks. It identifies the congestion by sensing retransmission timeout or receipt of DUPACKs. It also suggests modification in calculation of retransmission timer. On the receipt of duplicate acknowledgement, it introduces the bounded delay period τ and sets it equal to one *rtt*. This helps the sender to recover the link-layer loses otherwise the retransmission algorithm is activated.

Adaptive Delayed Acknowledgement (TCP-ADA) [7] is also a sender side solution for mobile ad-hoc networks. It uses Delayed Acknowledgement (DelAck) to block acknowledgement for a specific time period.

Chandran et al., proposed a router assisted solution named as; TCP-Feedback (TCP-F*)* [10][9] for Ad hoc wireless networks. TCP-F detects route failures in case an intermediate node move or fails the next node detects route fail. When route failures detected the next node transmits *route failure notification (RFN)* packet to destination. On the receipt of *RFN* packet each node invalidates the previous route and stops incoming packets. If the alternate path does not exist each node relays the *RFN* packet to source node.

David X. Wei et al., [16] presents a TCP congestion control variant for high speed and high latencies suffering networks. It highlights the four difficulties which TCP faces in high speed networks. It devised a mathematical and analytical model of FAST-TCP. The results show that FAST-TCP shows some improvements over other TCP variants with respect to the throughput and fairness.

TABLE I
SUMMARY OF TECHNIQUES FOR CONGESTION CONTROL IN WIRELESS ENVIRONMENT

Schemes	TCP semantics	Support for mobility	Modification requirement	Targeted application
I-TCP [13]	Split	High	Base station	Cellular
M-TCP [14]	Split	High	Router and end stations	Cellular
SNOOP [15]	End-to-end		Base station	
TCP-Peach [16]	End-to-end	High	Router and end stations	Satellite
ATCP [17]	End-to-end	High	TCP stack	Ad hoc
TCP-ADA [7]			Sender side	Ad hoc
Freeze-TCP [18]	End-to-end	High	Base station	Cellular
TCP-New Reno	End-to-end	Low	Sender side	Heterogeneous
TCP-SACK	End-to-end	Low	Sender side	Heterogeneous
TCP-Vegas [6]	End-to-end	Low	Sender side	Heterogeneous
TCP-Veno [20]	End-to-end	Low	Sender side	Heterogeneous
TCP-Westwood [19]	End-to-end	High	Sender side	Heterogeneous
TCP-Jersey [21]	End-to-end	High	Router and sender side	Heterogeneous
TCP-New Jersey [12]	End-to-end	High	Router and sender side	Heterogeneous
MA-TCP [24]		High		Heterogeneous
ECN			Sender, receiver	
ELN			Sender, receiver	
JTCP [22]	End-to-end		Sender, receiver	Heterogeneous
ILC-TCP [25]	End-to-end	High	Sender side	Heterogeneous

J. Liu et al. Presents ATCP [14] for mobile ad hoc networks. ATCP uses the ECN and ICMP to put the sender in one of four states such as normal congested, loss, and disconnected. Beside some improvement over ad hoc networks it unable to prove strength in situation where blackout and disconnection are frequent.

In [9], challenges in heterogeneous network environment are addressed, and the TCP schemes for different wireless applications (i.e. cellular, ad-hoc and satellite networks) and type of implementation (i.e. split mode or end-to-end approach) are classified. In [10], the performance characteristics of four representative TCP schemes are studied, namely TCP New Reno, SACK, Veno, and Westwood, under the network conditions of asymmetric end-to-end link capacities, correlated wireless errors, and link congestion in both forward and reverse directions. Then a new TCP scheme, called TCP New Jersey is proposed, which is capable of distinguishing wireless packet losses from congestion packet losses, and reacting accordingly, as an improvement of TCP Jersey.

[12]The performance of TCP degrades over wireless links due to high rate of data losses, which are falsely perceived as network congestion state. TCP performance metrics also diminish due to low data rate, since large delays may occur in last link i.e. wireless link. Similarly in heterogeneous wireless network, packet loss may also occur due to mobility-events that can cause burst-losses, service-disconnection. This motivates to reevaluate TCP control operations and embed some mobility related services to optimize its performance for new generation of wireless networks.

In [11] a new wireless congestion control protocol (WCCP) is proposed, based on the channel busyness ratio. In this protocol, each forwarding node determines the inter-node and intra-node fair channel resource allocation and allocates the resource to the passing flows by monitoring and possibly overwriting the feedback field of the data packets according to its measured channel busyness ratio. The feedback is then carried back to the source by the destination, which copies it from the data packet to its corresponding acknowledgment. Finally, the source adjusts the sending rate accordingly. Clearly, the sending rate of each flow is determined by the channel utilization status at the bottleneck node. There are two components in WCCP. One is at the transport layer. It replaces the window adjusting algorithm of TCP with a rate control algorithm to regulate the sending rate. The other is between the networking layer and the MAC layer. It monitors and possibly modifies the feedback field in TCP data packets when it passes the outgoing packets from the networking layer to the MAC layer and the incoming packets in the reverse direction.

Mobility Aware Transmission Control Protocol (MA-TCP) [24]. The MA-TCP consist off the handoff state identification and adaptation mechanism (HIAM) which is responsible for the receipt of the triggers from media independent handoff (MIHF) and generate appropriate message.

In [29] it is foreseen that next generation mobile terminals would be suitable to have transport layer that is possible to be downloaded and installed (Open Transport Protocol - OTP). Such mobiles shall have the possibility to download TCP version which is targeted to a specific wireless technology installed at the base stations. Such transport layer will be open to different implementations of the transport protocols, which may differ from today's main transport protocols.

Another idea about application of congestion control methods is to integrate different schemes into operating system (i.e. Linux) [30].

IV. CONCLUSION

In this paper, we provide a survey of existing congestion control mechanisms for wireless networks.

It is clear that each of the proposed solutions has characteristics which best suit a given environment. Next generation wireless networks, which will support the ABC (Always Best Connected) concept, will integrate heterogeneous wireless communication environments with different characteristics. TCP algorithms do not hold any more for the emerging wireless networks. As it is given in this paper there are many different TCP modifications for congestion control as well as new proposed transport protocols. Since the wireless resources are very scarce the accommodation of transport protocols is necessary with aim to provide higher utilization of radio interface as well as lower latency. Hence, our future work is targeted to development of open transport protocol layer in wireless terminals.

REFERENCES

[1] S. Floyd, Ed., "Metrics for the evaluation of congestion control mechanisms," March 2008, [Online]. Available: http://tools.ietf.org/pdf/rfc5166.pdf

[2] L. Mamatas, V. Tsaoussidis, and C. Zhang, "Approaches to Congestion Control" in Packet Networks, 2004.

[3] C. Jin, D. Wei, and S. Low, "FAST TCP: Motivation, Architecture, Algorithms and Performance," in *IEEE INFOCOM*, Mar 2004.

[4] S. Floyd, "HighSpeed TCP for Large Congestion Windows," in *IETF RFC 3649*, Dec 2003.

[5] L. Xu, K. Harfoush, and I. Rhee, "Binary Increase Congestion Control (BIC) for Fast Long-Distance Networks," in *IEEE INFOCOM*, Mar 2004.

[6] L. Brakmo and L. Peterson, "TCP Vegas: End to End Congestion Avoidance on a Global Internet," in *IEEE J. Selected Areas in Communications*, Oct 1995.

[7] A. K. Singh, K. Kishore, "TCP-ADA: TCP with adaptive delayed acknowledgement for mobile ad hoc networks," IEEE Wireless Communications and Networking Conference, USA, 2004.

[8] S. Shakkottai and R. Srikant, "Network Optimization and Control", Foundations and Trends in Networking, Vol. 2, No. 3 pp. 271-379, 2007.

[9] Y. Tian, K. Xu, and N. Ansari, "TCP in Wireless Environments: Problems and Solutions", IEEE Radio Communications, March, 2005.

[10] Y. Tian, K. Xu, and N. Ansari, "Improving TCP performance in integrated wireless communications netwoks", Computer Networks, Vol.47, pp.219-237, 2005.

[11] M. Kumar, S. Prakash, "Wireless Congestion Control Protocol For Multihop Ad Hoc Networks", (IJCSIS) International Journal of Computer Science and Information Security, *Vol. 7, No. 1, 2010.*

[12] M. S. Akbar, S. Z. Ahmed, M. A. Qadir, "Performance Optimization of Transmission Control Protocol in Heterogeneous Wireless Network during Mobility", IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.8, August 2008.

[13] Bakre, B.R. Badrinath, I-TCP: Indirect TCP for Mobile Host, Proceedings of the 15th international conference on distributed computing system (ICDCS '95), PP.136 – 143, 30 May-2 Jun 1995.

[14] K. Brown and S. Sing, M-TCP: TCP for Mobile Cellular Networks, ACM SIGCOMM Computer Communication Review Volume 27, Issue 5, PP. 19 – 43, October 1997.

[15] H. Balakrishnan, S. Seshan, E. Amir, and R.H. Katz, "Improving TCP/IP performance over wireless networks," in Proc. ACM MobiCom, 1995.

[16] I. F. Akyildiz, G. Morabito, and S. Palazzo, TCP-Peach: A New Congestion Control Scheme for Satellite IP Networks, IEEE/ACM Trans. Net., vol. 9, no. 3, pp. 307–321, June 2001.

[17] J. Liu and S. Singh, ATCP: TCP for Mobile Ad Hoc Networks, IEEE JSAC, vol. 19, no. 7, pp. 1300–1315, July 2001.

[18] Tom Goff, James Moronski, D. S. Phatak and Vipul Gupta, Freeze-TCP: A True End-to-End TCP Enhancement Mechanism for Mobile Environments, Proc. IEEE INFOCOM 2000, vol. 3, Tel Aviv, Israel, pp. 1537–1545, Mar. 2000.

[19] Casetti Claudio Casetti, Mario Gerla, Saverio Mascolo, M.Y. Sanadidi and Ren Wang, TCP Westwood: End-to-End Congestion Control for Wired/Wireless Networks, Wireless Networks, vol. 8, no. 5, pp. 467–479, 2002.

[20] C. P. Fu and S. C. Liew, TCP Veno: TCP Enhancement for Transmission over Wireless Access Networks, IEEE JSAC, vol. 21, no. 2, pp. 216–28, Feb. 2003.

[21] K. Xu, Y. Tian, and N. Ansari, TCP-Jersey for Wireless IP Communications, IEEE JSAC, vol. 22, no., pp. 747–756, 4 May 2004.

[22] E. H.-K. Wu and M.-Z. Chen, JTCP: Jitter-Based TCP for Heterogeneous Wireless Networks, IEEE JSAC, vol. 22, no. 4, pp. 757–766, May 2004.

[23] N. Moller, A. Arvidsson, J. Petersson, C. Fischione, R. Skog, P. Karlsson, K.H. Johansson, Supporting End-to-End Applications Over HSDPA by Cross-Layer Signaling , IEEE Wireless Communications and Networking Conference 2007, pp. 3855-3860.

[24] M. Saeed Akbar, Syed Zubair Ahmed and M. Abdul Qadir, "Mobility Aware Transmission Control Protocol (MA-TCP)", the Fourth International Conference on Mobile Computing and Ubiquitous Networking (ICMU 2008) June 11-13, 2008 Miraikan, Tokyo, Japan. PP 146-151.

[25] M. Chinta, A. Helal, and C. Lee, "ILC-TCP: An Interlayer Collaboration Protocol for TCP Performance Improvement in Mobile and Wireless Environments," Proc. IEEE WCNC 2003, vol. 2, New Orleans, LA, Mar. 2003, pp. 1004–10.

[26] Raisinghani, V.T.; Singh, A.K.; Iyer, S, "Improving TCP performance over mobile wireless environments using cross layer feedback," IEEE International Conference on Personal Wireless Communications, pp. 81–85, 15-17 Dec. 2002.

[27] T. Go , J. Moronski, D. S. Phatak, and V. Gupta, .Freeze-TCP: A true end-to-end enhancement mechanism for mobile environments,. In Proceedings of IEEE INFOCOM 2000, Israel.

[28] Jian Pu and Mounir Hamdi, "QFCP: a Router-Assisted Congestion Control Mechanism for Heterogeneous Networks", SITIS 2006.

[29] T. Janevski, "5G mobile phone concept", In *Proceedings of the 6th IEEE Conference on Consumer Communications and Networking Conference*, Las Vegas, NV, USA, January 11 - 13, 2009.

[30] P. Sarolahti, A. Kuznetsov, "Congestion Control in Linux TCP", In Proceedings of USENIX Annual Technical Conference, Monterey, California, USA, June 10-15, 2002.

Video Transmission Over Slowly Fading Channels Using Diversity and Symbol of Power Adaptation

Slavche Pejoski and Venceslav Kafedziski

***Abstract*—A novel method for video transmission over fading channels, that minimizes end-to-end video distortion, is presented. The method is based on the use of diversity and optimal allocation of the number of symbols used for transmission of each packet. The impact of channel errors is taken into account using outage probability. Optimization is performed using Lagrangian method. Similar procedure is performed to optimize the power allocated to each packet. Symbol adaptation shows superior results to power adaptation.**

***Keywords*—video, cross-layer adaptation, outage capacity, symbol allocation, power allocation.**

I. Introduction

Video transmission is one of the most popular services in the telecommunications world. Much research has been performed in this field, especially for wired channels, but wireless channel still remains in the focus of the research community. The disadvantages of wireless compared to wired transmission are twofold: the available resources are very limited, and the channel conditions change very often. On the other hand, video coding in wireless environment still has its basic features: it is very sensitive to delay and the bit rate needed to encode a single video frame differs significantly from a frame to a frame because it depends on the particular video sequence. This makes the wireless video transmission quite a challenging task, so that optimal use of all available resources and system adaptation to current conditions are necessary. The optimization process is especially needed in real time transmission where the delay should be very low, so retransmission is impossible to be carried out most of the time. In those situations, usually, a combination of forward error correction coding (FEC) and interleaving is employed [1], [2]. In [1] perfect interleaving is employed, which is impossible to do in delay constrained environment. Authors in [2] propose the use of strict set of Reed Solomon (RS) codes that are optimal only for a single SNR value.

The adaptation process has been studied in [1] and [3]. In those works Lagrangian based optimization is used, which is employed as the optimization algorithm in our work, as well. In [3] the channel is assumed to be known and adaptation is done for parts of the video frame, which is not the case here. Authors in [1] use binary modulation only, and the full effect of symbol adaptation is not considered. Review of the achievements in the aforementioned field is presented in [4], [5], [6].

In our previous work [5], we showed that the use of the available channel diversity leads to improvement in the system performance. In that work, we pointed to one of the possible ways to achieve optimal use of the available diversity, based on approximately universal codes. The approach of using approximately universal coding is explained in detail in [7]. In [5] we characterized the slowly varying channel, with no information about the instantaneous channel gain at the transmitter. There, we used the probability density function (pdf) of the channel mutual information to calculate the outage probability of different source coding modes and choose the ones that maximize the end-to-end quality. In the process of choosing video coding modes the transmitter knows that bits obtained from encoding each video slice are mapped to a specific transmission packet. In [5] we assumed that the number of symbols that can be used for transmission of each packet is fixed. But, using the same number of symbols to transmit each packet is suboptimal. The suboptimality is especially pronounced in the region of low symbol rates. Here we use the same approach of calculating the pdf of channel mutual information, mapping parts of the video frame to different video packets and choosing the optimal video coding modes for parts of the video contained in each packet, but extend the work of [5], by including optimal allocation of the available symbols per packet. In order to accomplish this task, we use a nonlinear optimization algorithm based on the Lagrangian optimization.

An optimization algorithm for bit allocation in quantization can be found in [8]. We modify this algorithm to using symbols instead of bits and using end-to-end measure of video quality. This modification requires creation of curves of dependence of the measure of video quality in terms of the number of symbols used for transmission, for each video slice. Then using the calculated curves, Lagrangian optimization is performed. It is known that Lagrangian optimization finds optimal or near optimal solution, which is explicitly shown in [8]. As a measure of the quality of the transmitted video we use the expected end-to-end distortion. The approach is explained in [9] and [10].

We also apply the adaptation to optimal allocation of transmission power to different packets, where we keep the number of symbols per packet fixed, and we compare the two approaches.

The paper consists of four parts. In the second part a description of the transmission system and the formulation of the algorithm itself are given. In the third part, simulation

S. Pejoski is with the Faculty of Electrical Engineering and Information Technologies, University Cyril and Methodius, Skopje, Republic of Macedonia, E-mail: slavchep@feit.ukim.edu.mk

V. Kafedziski is with the Faculty of Electrical Engineering and Information Technologies, University Cyril and Methodius, Skopje, Republic of Macedonia, E-mail: kafedzi@feit.ukim.edu.mk

setup and results are shown and in the last part concluding remarks are presented.

II. System Description

The system we use here is the same as the one in [5] and is shown in Fig. 1. At the transmitting side it consists of video encoder, channel encoder, transmitter and controller.

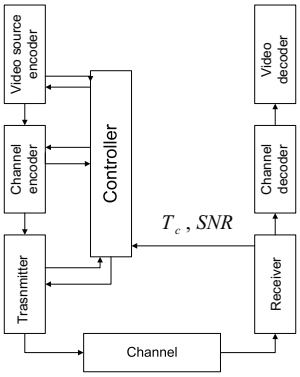


Fig. 1. Video transmission system

In our system the transmitter can have access to certain amount of information about the previously arrived video packets at the receiver. The information delay is dependent on the delay in the transmission system. The transmitter also knows the statistics of the wireless channel i.e. it is familiar with the coherence interval and the probability density function of the channel gain.

The adaptation of system parameters is always based on choosing the optimal modes for different system parameters. In order to evaluate the influence of the different encoding options, an evaluation measure for the received video quality is needed. Here we use the end-to-end expected distortion. The expected distortion is very suitable, because it is calculated at the transmitter and it considers the error probability of video packets and the discrepancy between the available reference frames at the transmitter and the receiver. For calculating the expected distortion of pixel i in video frame n, we use the following expression:

$$D = E[(f_n^i - \tilde{f}_n^i)^2] \tag{1}$$

where f_n^i is the value of the i-th pixel in the n-th video frame, and $\tilde{f}_n^i$ is the value of the i-th pixel of the n-th video frame at the receiver. To calculate the expected distortion in Eq. (1) recursive algorithms can be used. Two such algorithms are described in [9] and [10].

In the calculation of the expected distortion two types of distortion are considered. The first one comes from the quantization process and the second one comes from the packet loss in the channel. These two parts are combined and are considered together because they are both influenced by the same parameters. For example, the choice of quantization level dictates the quantization error and the number of bits which in turn affects the error probability that influences the distortion from packet losses. The distortion due to transmission depends on the packet error probability that makes this parameter important. The packet error rate in wireless systems is given by [7]:

$$P_e = P_{out}P(error|O) + (1 - P_{out})P(error|O^c) \tag{2}$$

where O is the outage event, O^c is the no outage event, $P(error|O)$ is the error probability when the channel is in outage, and $P(error|O^c)$ is the error probability when the channel is not in outage. Similarly to what we did in [5], we assume that our channel coder is a capacity achieving one and that it uses powerful codes that drive $P(error|O^c)$ to zero. Based on these assumptions, the packet error rate becomes $P_e = P_{out}$. For wireless channel in which the coherence interval is L times shorter than the time necessary for the transmission of a single frame and the whole available diversity is used, the outage probability is calculated according to [11]:

$$P_{out}(R) = P\left(\frac{1}{L}\sum_{i=1}^{L}\log_2(1 + |h_i|^2 SNR) < R\right). \tag{3}$$

Eq. (3) allows us to map the channel conditions to the pdf of the channel mutual information. The pdf of the mutual information can be used to calculate the error probability for different source coding modes.

The novelty of this work comes from adaptation of the number of symbols used by different video packets, that are used for transmission of a single video frame, such that the distortion of the video sequence at the receiver is minimized. The algorithm used to allocate the available resources to different video packets is based on nonlinear optimization process, carried out by Lagrangian optimization. Before explaining the allocation process it is necessary to point out that symbol adaptation over different coherence intervals in a time varying channel, when the diversity is not used, brings very low benefit and great complexity. The utilization of the whole available diversity makes the share of the symbols from different channel intervals equal, resulting in the same expression for the mutual information given by Eq. (3) and hence makes the optimization process of symbol adaptation much simpler.

The optimal use of the available diversity for video transmission is illustrated in Fig. 2, where $\gamma_i = |h_i|^2$. In this figure, time flow starts from the upper left corner and ends at the lower right corner, it is always horizontal, and when it reaches the end of the interval it continues from the left side of the next coherence interval. In the example shown in this figure we assumed that the wireless channel coherence interval is one third of the frame duration. The borders of packets when equal number of symbols are allocated to all video packets are shown in black color. Then Δ_s symbols are reallocated form packet $P2$ to packet $P1$, and the time positions of those symbols are placed in such a manner, that the pdf of channel mutual information per symbol doesn't change. The new borders of packets $P1$ and $P2$ are shown in red.

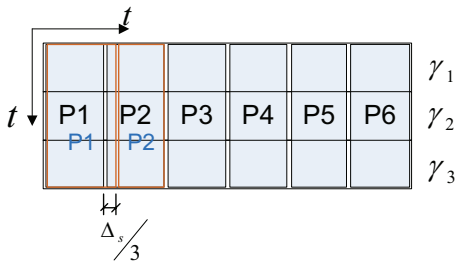


Fig. 2. Optimal use of the available diversity and symbol reallocation

The Lagrangian optimization is a well known method of nonlinear optimization (described, for instance, in [12]). The application of the Lagrangian based algorithm for optimal bit allocation to different subbands in subband coding was first described in [8]. Here, we use this algorithm for optimal allocation of the available symbols to different packets. The algorithm finds an accessible point that lies on the convex hull of the curve that describes the expected distortion in terms of the number of symbols used for transmission, and is closest but lower or equal to the number of available symbols. If the point found during the Lagrangian optimization process has lower number of symbols than the available number of symbols, additional optimization takes place. This additional optimization allocates the rest of the symbols in a greedy manner, i.e. it allocates them to the packet, for which the additional symbols bring largest performance improvement. Obviously, in order for the optimization to take place, curves for expected distortion in terms of the number of used symbols have to be created for each video packet. In the process of curve creation, for every option of used symbols, optimization of the modes used by the video coder must be performed. This additional optimization can be very tedious if a large number of source coding options are available, since it is proportional to the number of source coding options.

We apply the concept of adaptive resource allocation to power allocation as well. In this case, the number of symbols per packet is kept constant, but the available power is allocated to particular transmitted packets in such a way that minimal expected distortion is obtained. For this procedure, we again use Lagrangian optimization. In this case, curves of dependence of the expected distortion in terms of the power used to transmit each packet are created.

III. Simulation Results

In our simulations we use the basic H.263 coder [13]. During the encoding process every video frame is divided into macroblocks, that are independently encoded. Each macroblock consists of 16x16 luma part and two 8x8 chroma parts. In order to obtain better resilience to transmission error the video frame is divided into slices which are independently packetized. Hence, even in situation when a loss of slice occurs, the decoding process for the current video frame can continue. In this work we assume that every slice consists of one row of macroblocks. During the encoding procedure, the encoder can choose between INTRA and INTER modes for every packet, and the quantization level for every slice. The quantization levels are chosen from the set $Q = \{10, 17, 27\}$, which is a subset of the set of all available quantization levels. To get the results we used 298 video frames of the Foreman SQCIF video sequence with maximal receiver video quality measured in Peak Signal to Noise Ratio (PSNR) of 32.94 dB, obtained using quantization level $q = 10$ and assuming no errors during the transmission. For our simulations we used simple error concealment. This concealment method conceals the lost pixels by copying them from the same spatial locations in the previous decoded frame, when packet loss occurs. In all our simulations we anticipate situation with no delay for the information about loss of already sent packets i.e. when coding the current video frame, the video coder knows all the packets that were lost for all video frames in the past. This assumption does not change the conclusions, because the situation with a finite delay can be controlled by the algorithm for calculation of the expected distortion, which results in performance shift only.

In our simulations we use a block fading channel model. The duration of the coherence interval is set to $T_c = 33/6 = 5.5ms$ i.e. there are 6 coherence intervals during the transmission interval of a single video frame. We use the maximum available diversity order 6 if not stated otherwise. The procedure for obtaining the probability density function of the channel mutual information for different diversity order can be found in [5].

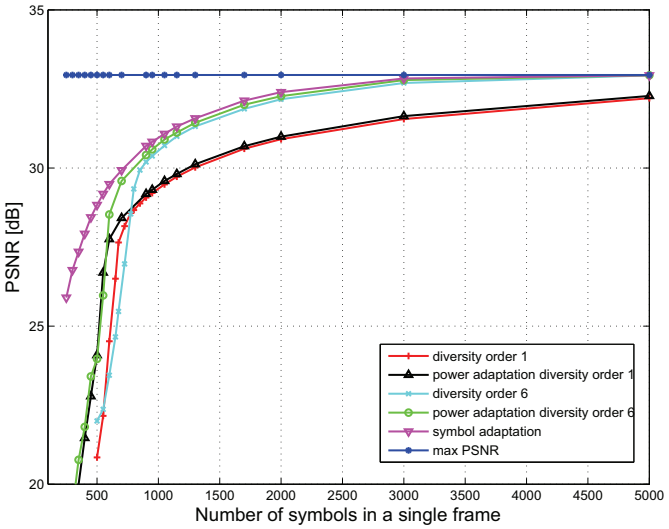


Fig. 3. PSNR vs number of symbols in a single frame

All the simulations are carried out for 20 different channel realizations. The PSNR in all figures is calculated as the mean PSNR over all video frames and channel realizations. In the symbol adaptation algorithm, the symbols are allocated to the packets in portions equal to 10% of the mean available symbols per packet. In the power adaptation algorithm, the power is allocated in portions equal to 10% of the mean power per packet. The resolution of the symbol and power allocation is variable and can be used for trading complexity and perfor-

mance. Namely, finer resolution implies better performance, but results in larger number of points on the curves and hence in increased complexity.

In Fig. 3 the quality of the received video sequence measured in PSNR in terms of the number of available symbols per video frame is shown. For all simulations shown in this figure the mean channel SNR is set to 16 dB. The figure contains performance curves for both symbol adaptation algorithm and power adaptation algorithm. The simulations for the power adaptation algorithm are performed for diversity order of one and six. The figure also contains curves for quality of the received video for diversity order of one and six, when both the power and available symbols per packet are fixed. It is obvious that symbol adaptation brings the largest benefit. This benefit is most pronounced in the region of low symbol rates. In the remaining region symbol adaptation still has the best achievable performance among the adaptation algorithms shown here, but the difference in performance among different algorithms is quite small. Fig. 3 also shows that symbol adaptation is superior to power adaptation for all available symbol rates. Still, the power adaptation algorithm results in performance gain compared to non symbol/power adaptive algorithms. The superiority of symbol adaptation comes from the high SNR that is used to obtain the results in this figure, namely at high SNR the additional energy doesn't modify the pdf of the channel mutual information as significantly as the additional symbols. Other benefit from using symbol adaptation is that, no matter the allocation, all the available resources are used, which is not the case for power adaptation. Namely, when no power is used for transmission of some video packet i.e. the packet is not sent, the symbols dedicated to that packet are not used.

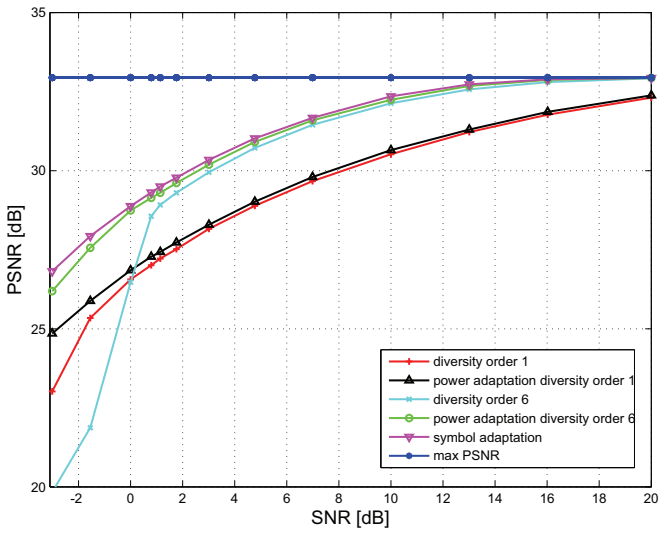


Fig. 4. PSNR vs mean channel SNR

Another comparison of the performance of the proposed algorithms measured in PSNR in terms of channel SNR is made in Fig. 4. In this figure the number of available symbols is kept at 3500 symbols per frame. It can be seen that in this setting the adaptation algorithms have similar performance and they both exceed the performance of algorithms that do not use symbol/power adaptation. Again symbol adaptation brings performance gains compared to power adaptation, but here the difference is not as large as in the constant SNR case.

IV. Conclusion

We present a method for adaptive video transmission in wireless channels. The method uses the pdf of channel mutual information and outage probability to calculate curves of dependence of the video quality on the number of symbols used for the transmission of each packet. For the video part contained in each video packet the optimal mode for source coding is chosen for every point on these curves. Then, symbol allocation procedure based on Lagrangian optimization is carried out. An extension of this concept to power adaptation is also proposed. Comparison between the novel adaptive algorithms and previously developed algorithms with no symbol/power adaptation is performed using simulation. Simulation results show that the adaptive algorithms outperform the constant resource algorithms over the entire region of available resources. The performance gain is highest in the region of limited available resources, where adaptive symbol allocation shows best performance of all.

References

[1] F. Zhai, Y. Eisenberg, T. N. Pappas, R. Berry, A. Katsaggelos, "Joint source-channel coding and power allocation for energy efficient wireless video communications," *Proc. 41st Allerton Conf. Communication, Control, and Computing*, October 2003.

[2] Q. Qu, Y. Pei, J. Modestino, X. Tian, "Error-resilient wireless video transmission using motion-based unequal error protection and intraframe packet interleaving", *ICIP International Conference on Image Processing*, Vol. 2, pp. 837 - 840, 2006.

[3] A. Argyriou, "Distortion-Optimized Video Encoding and Streaming in Multi-Rate Wireless LANs", *IEEE ICASSP*, March 2008, p.p. 2169-2172.

[4] A. Katsaggelos,Y. Eisenberg, F. Zhai, R. Berry, T. Pappas, "Advances in Efficient Resource Allocations for Packet-Based Real-Time Video Transmission", *Proceedings of the IEEE*, Vol. 93, No. 1, January 2005.

[5] S. Pejoski, V. Kafedziski, "Video transmission on slowly fading channels using diversity", Telfor 2009, November 2009, pp. 307-310.

[6] M. Schaar, S. Shankar, "Cross-layer wireless multimedia transmission: Challenges, principles and new paradigms", *IEEE Wireless Communications*, August 2005, p.p. 50-58.

[7] S. Tavildar, P. Viswanath, "Approximately Universal Codes over Slowly Fading Channels", *IEEE Transactions on Information theory*, Vol. 52, No. 7, pp. 3233-3258, July 2006.

[8] Y. Shoham, A. Gersho, "Efficient Bit Allocation for an Arbitrary Set of Quantizers", *IEEE Transactions on Acoustics, speech and signal procession*, Vol. 36, No. 9, September 1988, p.p. 1445-1452.

[9] R. Zhang, S. Regunathan, K. Rose, "Video Coding with Optimal Inter/Intra-Mode Switching for Packet Loss Resilience", *IEEE Journal on selected area in Communications*, Vol. 18, No. 6, June 2000, pp. 966-976.

[10] Y. Zhang, W. Gao, Y. Lu, Q. Huang, D. Zhao, "Joint Source-Channel Rate-Distortion Optimization for H.264 Video Coding Over Error-Prone Networks", *IEEE Transactions on multimedia*, Vol. 9, No. 3, April 2007, pp. 445-454.

[11] D. Tse, P. Viswanath, "Fundamentals of wireless communications", Cambridge university press, 2005.

[12] S. Rao, "Engineering Optimization: Theory and Practice, Forth Edition", John Wiley & Sons, 2009.

[13] Available at: http://euler.slu.edu/ fritts/mediabench/mb2/index/html

Realizable Video Transmission Over Fading Channels with Full Channel State Information

Slavche Pejoski and Venceslav Kafedziski

***Abstract*—A novel method for wireless video transmission is presented, where channel coherence interval is shorter than the video frame duration and causal channel state information (CSI) is available at the transmitter. Triplets of bit error probability, outage probability and obtainable bits for given outage probability are calculated. Based on these triplets, the video sequence is optimally encoded at the beginning of the video frame. During transmission, parameters at physical layer are chosen to provide bit error probability lower than predetermined one.**

***Keywords*—video transmission, fading channel, causal channel state information, packet error probability**

I. Introduction

The real time video transmission has the potential of becoming leading service in the world of telecommunications and it receives a lot of attention.

Real time video transmission is characterized by two most important features: low tolerable delay and variable bit rate. The wireless channel has its own features, among which most important are the restricted resources (bandwidth and power) and the time varying nature of the channel gain. With the continuous advance in telecommunications, there is a possibility to estimate the channel state (channel gain) and feed back this information to the transmitter. This channel state information can be used to choose the optimal modes for different parameters of source and channel encoders. A number of papers [1], [2] [3] show that joint source and channel coding can improve the quality of the received video. This cross-layer paradigm is easily applicable when the channel is known and constant for the whole duration of the video frame transmission and is thoroughly explained in [1] and [2]. In situations when the channel coherence interval (period with equal channel gain) is shorter than the duration of the video frame, it is impossible to know all the channel states that correspond to given video frame at the time instant of encoding that video frame.

In the literature, three approaches are known to be useful in this setting. The first approach employs scalable encoders, for example FGS (Fine Granularity Scalability) video encoders. They encode the video frame using several layers. The adaptation process in this system consists of adapting the physical layer to guarantee some minimal bit error rate (BER) (usually 10^{-6}) and subsequently sending the base layer and part of the enhancement layer, based on their contribution to the overall quality of the decoded video. The down side of this approach is the well known coding inefficiency when low delay is needed. This approach can be found in [4].

The second approach considers the channel as unknown to transmitter and uses coding over subchannels. This approach is quite complicated because it requires interleaving that is adapted to the channel and is usually done with binary modulation. The second disadvantage of this method comes from not using channel information at the transmitter. This approach can be found in [3], [5].

In the third approach the adaptation is done in parts i.e. the video is divided into a number of parts equal to the number of channel coherence intervals corresponding to transmission of a single video frame. So in a given coherence interval specific part of the video and the channel are jointly encoded. The basics of this approach can be found in [1].

In our previous work [6], we investigated the impact of different parameters at the physical layer on the received video quality and we found that variable power has lowest impact. In [6] we used a system with full channel knowledge for all channel states during the video frame transmission. Here we extend the work in [6] by using the conclusions drawn there and making the system causal i.e. the transmitter knows the CSI about the current channel state. In the method several values of outage probability are chosen. Then, based on the statistics of the channel, different BERs are obtained. Each BER corresponds to specific outage probability and for each BER we assume that the packet error rate for video packet with average length, due to outage, is equal to the packet error rate when the system is not in outage. Having the pair of BER and outage probability, the maximal number of bits that can be used in the system at outage probability lower then the calculated outage probability is found. To this number of bits additional bits corresponding to the first channel coherence interval (in this coherence interval the outage probability is zero, but the calculated BER needs to be guaranteed) are added. Based on the triplet (BER, outage probability, number of bits) the source and channel coding are performed. The expected distortion computed at the transmitter is used as the utility function. During transmission, the modes (made of the physical layer parameters) are set to guarantee BER lower than the chosen one. If there are not enough resources to send all the bits, the bits that cannot be sent are discarded.

The paper consists of four sections. In the second section we give description of the system and the proposed algorithm. In the third section we evaluate the proposed algorithm by simulation and in the last section we give conclusion.

S. Pejoski is with the Faculty of Electrical Engineering and Information Technologies, University Cyril and Methodius, Skopje, Republic of Macedonia e-mail: slavchep@feit.ukim.edu.mk.

V. Kafedziski is with the Faculty of Electrical Engineering and Information Technologies, University Cyril and Methodius, Skopje, Republic of Macedonia e-mail: kafedzi@feit.ukim.edu.mk.

II. System and Algorithm Description

The system consists of two entities, one sending the video and the other receiving it. In our system we allow the transmitter to change the parameters of the physical and application layer. Based on the conclusions in [6], we allow the transmitter to choose modes for modulation and coding rate at the physical layer. We use "m" to denote the chosen modulation type, from the set M of all available modulation types and "c" to denote the coding rate, chosen from the set C of all available coding rates. Every combination of "m" and "c" results in mode "v" from the set V of all available combinations. At the application layer the coder can choose the modes of source coding for quantization and coding mode (INTRA/INTER). The chosen mode at the application layer is marked with μ and is chosen from the set of all modes β. The sender starts from the video sequence of raw video frames and encodes it in an adaptive manner in order to achieve best quality at the receiver. During the adaptation process, in order to account for the losses in the wireless channel and the discrepancy between reference frames at the receiver and transmitter, the expected distortion at the decoder is calculated at the encoder using:

$$E[D_n^j] = E[(f_n^j - \tilde{f}_n^j)^2] \tag{1}$$

In (1) f_n^j is the value of the j-th pixel in the n-th video frame, and $\tilde{f}_n^j$ is the value of the j-th pixel of the n-th video frame at the receiver. The calculation of the expected distortion can be done in a recursive manner using the algorithms described in [7] or [8]. The expected distortion depends on the packet error probability, that can be calculated using:

$$P_p = P_{out} + (1 - P_{out})(1 - (1 - P_b)^L). \tag{2}$$

In (2) P_{out} is the outage probability in the system i.e. the probability that the number of bits, used by the video encoder, is larger than the one obtained in the system that guarantees BER lower than P_b. L is the number of bits in the packet.

When coding rate is $r_c = 1$, the BER for binary modulation and MQAM can be respectively calculated from:

$$P_b = Q(\sqrt{2\gamma}) \tag{3}$$

and

$$P_b = 4\frac{\sqrt{M}-1}{\sqrt{M}\cdot\log_2 M}Q(\sqrt{\frac{3\gamma}{M-1}}) \tag{4}$$

If the coding rate is lower than 1, the equations will change in accordance with the channel coding used in the system. In cases where convolutional coding is used, equations for BER calculations can be found in [9].

The optimization process starts at the beginning of the first coherence interval in the video frame transmission interval. The process finds triplets of BER, outage probability and number of bits that can be used to encode the video frame at the specified outage probability. In order to explain the relationship between the elements of the triplet, we will take a look at their calculation for system where the coding rate is $r_c = 1$ and three modulation schemes (BPSK, 4QAM, 16 QAM) are used. In order to guarantee given bit error rate, every transmission mode (combination of coding rate and modulation) can be used above a specific SNR level. If $BER = P_b$ then γ values for different modes can be calculated using the inverse of (3) and (4). These levels are $\gamma_1 = \frac{1}{2}(Q^{-1}(P_b))^2$ for BPSK, $\gamma_2 = (Q^{-1}(P_b))^2$ for 4QAM and $\gamma_3 = 5(Q^{-1}(\frac{4}{3}P_b))^2$ for 16QAM. Best spectral efficiency can be achieved if BPSK is used in the SNR range $[\gamma_1, \gamma_2)$, 4QAM in $[\gamma_2, \gamma_3)$ and 16QAM is used in $[\gamma_3, \infty)$. If the channel probability density function is $\rho(\gamma)$ than the probability of not sending anything is $P_o = \int_0^{\gamma_1}\rho(\gamma)d\gamma$. Similar equations can be used for calculating the probabilities P_{BPSK}, P_{4QAM} and P_{16QAM} that BPSK, 4QAM or 16QAM are used. We form pairs of obtainable bits per symbol and the probability for those bits per symbol to be obtained. We denote those pairs as $(0, P_o), (1, P_{BPSK}), (2, P_{4QAM}), (4, P_{16QAM})$. These pairs can be viewed as the pdf's of the obtainable bits per symbol. Then for a given outage probability P_{out} the obtainable bits per symbol at that outage probability can be calculated as the minimum value of the bits per symbol R_o for which the cumulative probability of all rates (in bits per symbol) lower than R_o is less than P_{out}. This approach can be extended to several consecutive channel coherence intervals by finding the pdf of the obtainable bits per symbol for the coherence intervals included (convolution can be used in order to obtain this pdf). When coding rates lower than $r_c = 1$ are used, a change in the calculation of γ levels for different modes is needed, which can be done by numerical means in cases where no inverse formulas are available.

The number of obtainable bits per symbol for specific outage probability is dependent on the outage probability and maximal allowed BER. In the first part of the algorithm we propose a way to relate the outage probability and the BER. We assume that the BER is such that the packet error probability for video packet with average length due to outage is equal to the packet error probability when there is no outage. Only packets sent during the coherence intervals when the channel state information is not available to transmitter are included in this calculation. For example, if SQCIF video sequence is used and the frame is transmitted during six coherence intervals, then five video packets will be included in the aforementioned procedure, since for the transmission of the first packet, we assume that the transmitter has channel state information. Next, the number of obtainable bits for the coherence intervals for which the channel gain is unknown (R_{out}) is calculated for the specified BER and outage probability. The number of bits in the current coherence interval is then added to R_{out} (the modes chosen at physical layer, for the first coherence interval, guarantee the same maximal BER). Having the triplets $\{P_{b_i}, P_{out_i}, R_{out_i}\}$, for I chosen values of P_{out} the following optimization process is carried out at the transmitter:

$$\min_{\mu, P_{out_i}, P_{b_i}} D = E[(f_n^j - \tilde{f}_n^j)^2] \tag{5}$$

$$B(\mu) \leq R_{out_i}$$

Here $B(\mu)$ is the number of bits used by the video encoder to encode the video frame. At the end of the opti-

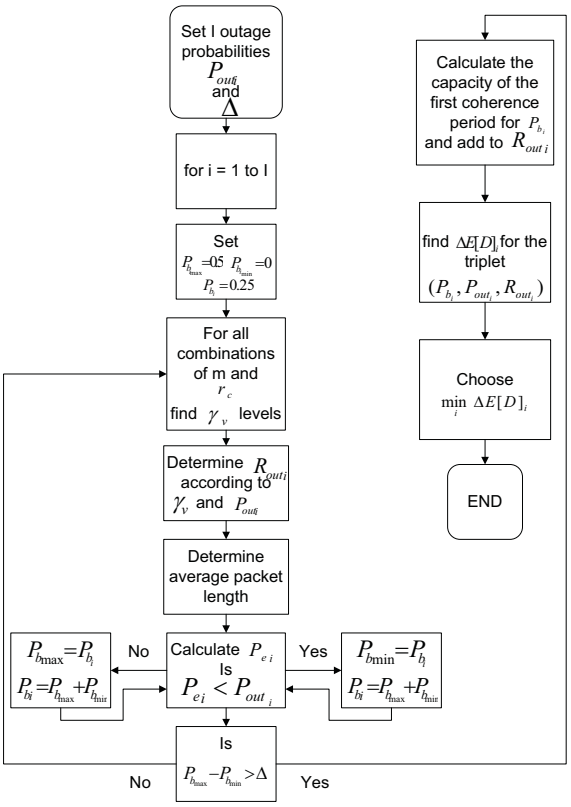


Fig. 1. Flow diagram

mization process coding modes for source coding and triplet $\{P_{b_i}, P_{out_i}, R_{out_i}\}$ that offer minimal expected distortion are chosen. The optimization process in (5) can be done using any optimization approach, for example Lagrangian optimization [3].

The proposed algorithm consists of the following steps:

1) *Set I values for P_{out_i} and set value for Δ (stopping criterion).*
2) *For $i = 1 : I$*
   a) *Set values for $P_{b_{i_{max}}} = 0.5$, $P_{b_{i_{min}}} = 0$, and starting value $P_{b_i} = 0.25$ for the BER.*
   b) *Find thresholds γ_v for all v, that guarantee BER lower than P_{b_i}.*
   c) *Determine the number of obtainable bits R_{out_i}, using γ_v and P_{out_i}.*
   d) *Determine the average packet length L as the ratio between R_{out_i} and the number of packets for which there is no transmitter channel state information. If this value is lower then 100, set the average packet length to 100.*
   e) *Calculate the average packet error probability $P_{e_i} = 1-(1-P_{b_i})^L$ when the system is not in outage. If P_{e_i} is higher then P_{out_i}, set $P_{b_{i_{max}}} = P_{b_i}$, $P_{b_i} = (P_{b_{i_{max}}} + P_{b_{i_{min}}})/2$, else, set $P_{b_{i_{min}}} = P_{b_i}$, $P_{b_i} = (P_{b_{i_{max}}} + P_{b_{i_{min}}})/2$.*
   f) *If $P_{b_{i_{max}}} - P_{b_{i_{min}}} > \Delta$ go to step (b).*
   g) *Calculate the maximal number of bits in the first coherence interval that can be obtained at BER not higher than P_{b_i}. Add this number of bits to R_{out_i}.*
   h) *Given the triplet $\{P_{b_i}, P_{out_i}, R_{out_i}\}$, choose the optimal source coding modes that minimize $E[D]$ (expected distortion) and do not use more bits then R_{out_i}. Save the chosen modes for each i, and set $E[D]_i = E[D]$.*
3) *Choose $\min_i E[D]_i$ and use the saved parameter for that i, to encode the sequence*

The value 100 in step 2) is the average packet length of a video packet when coding a slice at the quantization level offering lowest resolution. The flow diagram of the algorithm is shown in Fig. 1.

III. Simulation results

In all simulations we use the basic version of H.263 [10]. In the coding process the video frame is divided into macroblocks and every macroblock is coded independently. A single macroblock consists of one 16x16 luma part and two 8x8 chroma parts. The packetization is done in error resilient manner, by dividing every video frame into slices. Here a slice is equal to a row of macroblocks. During the coding of the video frame, the encoder can choose between INTRA and INTER coding for every macroblock, and can choose the quantization level for every video slice. The quantization levels are chosen from a set of allowable quantization levels $Q = \{10, 17, 27\}$. Simulations are done using the Foreman SQCIF video sequence composed of 298 video frames. The video sequence has maximal video quality measured in Peak Signal to Noise Ratio (PSNR) of 32.94 dB, obtained by using the quantization level $q = 10$, and assuming that there are no channel errors. At the decoder, simple error concealment is used. In case of error, this error concealment method copies the pixel at the same spatial location in the previous decoded video frame.

In the simulations Rate Compatible Punctured Convolutional Codes (RCPCC) with coding rates $r_c \in \{\frac{4}{5}, \frac{2}{3}, \frac{4}{7}, \frac{1}{2}\}$ is used. The convolutional code has two generator polynomials 171 and 133 and the constrained length of the encoder is $k = 8$. For all simulations hard decoding is used.

In our simulations the modulation types used are from the set $M = \{BPSK, 4QAM, 16QAM, 64QAM\}$.

We model the wireless channel as a block fading Rayleigh channel. The coherence interval of the channel is set to $\frac{1}{6}$ of the duration of one video frame.

In order to evaluate the performance of our algorithm we compare it to algorithm that uses a hypothetical transmitter which knows the channel states for the whole duration of the video frame and has more capabilities at the physical layer, having the option to adjust the power level in different coherence intervals. In this algorithm, which we call JSCCPM [6], the physical layer supplies the application layer with 5 different budgets that correspond to the following bit error rates $\{0.05, 0.01, 0.005, 0.001, 0.0005\}$, and then the video coder finds the source parameters for each value of BER resulting in the lowest expected distortion, $E[D]_i$. The parameters obtained for the lowest $E[D]_i$ are used for transmission.

JSCCPM differs from the newly proposed algorithm in two aspects: its outage probability is equal to zero and it has the ability to use power adaptation. In order to get better insight in the performance of the proposed algorithm, we use yet another algorithm for cross layer optimization of video transmission. This algorithm jointly optimizes the video and physical parameters for every coherence interval. The algorithm maps a slice to a video packet, so constant parts of the video are mapped into strictly defined channel intervals. Therefore, we call this algorithm "algorithm with constant parts".

In the optimization process for the newly proposed algorithm, we used five different values for the outage probability $P_{out} \in \{0.25, 0.1, 0.05, 0.025, 0.005\}$.

In Fig. 2, a comparison of the performance (measured in terms of PSNR) of the proposed algorithm and the other two algorithms, at a mean channel SNR of 10 dB, is shown.

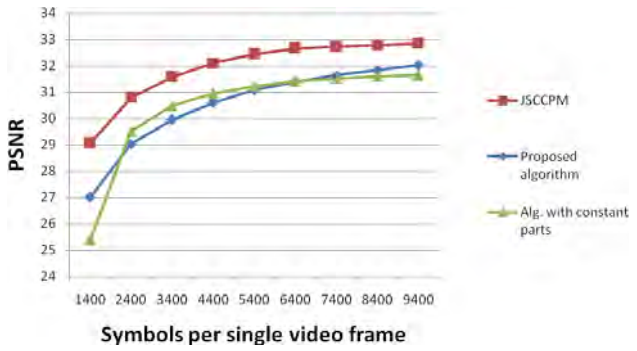


Fig. 2. PSNR vs number of symbols in a single frame

From Fig. 2 it can be seen that the proposed algorithm shows worse performance than JSCCPM. Still, the difference is within 2 dB and reduces to 1 dB for high symbol rates. The proposed algorithm shows better performance compared to the algorithm with constant parts, in the regions of low and high available resources. This is due to the ability of the proposed algorithm to send the video packets for different parts of the video frame in different coherence intervals, so that in the area where extra bits are needed to improve the quality of those video parts improvement in the performance of the received video is achieved. In the remaining region, the use of joint source and channel coding for different coherence intervals, brings improvement for the algorithm with constant parts, compared to the proposed algorithm.

In Fig. 3 we compare the performance of the three algorithms in terms of the mean channel SNR, assuming symbol rate of 2500 symbols per video frame. This figure is for the region of available symbols per video frame where the proposed algorithm shows somewhat worse performance than the algorithm with constant parts. Still, these two algorithms perform closely over the entire region of mean channel SNR. When compared to the JSCCPM algorithm we can see that at high SNR the performace loss is very small. At low SNR the performance loss is significant.

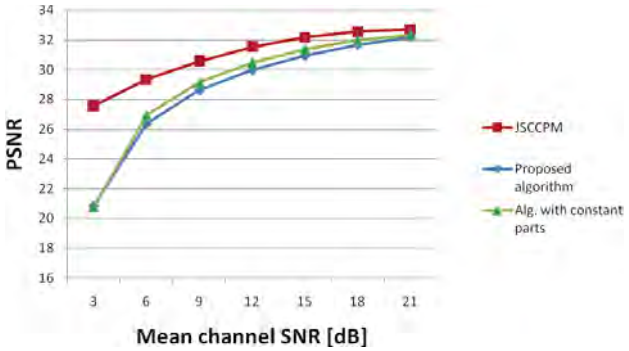


Fig. 3. PSNR vs mean channel SNR

IV. Conclusion

We propose an algorithm for real time video transmission over a wireless channel. In the transmission system causal CSI is available at the transmitter and channel coherence interval is shorter than the transmission interval of a video frame. Encoding procedure is carried out at the beginning of the video frame transmission interval. In the coding process triplets of BER, outage probability and number of bits obtainable at that outage probability are presented to the video encoder. The encoder determines the optimal video coding mode, for each given triplet. The triplet resulting in minimal expected distortion is chosen for coding and transmission. At reasonably high symbol rate and SNR, the proposed algorithm is capable of performing within 1 dB in PSNR of the received video, compared to the algorithm that has ideal knowledge of CSI at all times.

References

[1] A. Argyriou: Distortion-Optimized Video Encoding and Streaming in Multi-Rate Wireless LANs, *IEEE ICASSP*, March 2008, p.p. 2169-2172.

[2] M. Schaar, S. Shankar: Cross-layer wireless multimedia transmission: Challenges, principles and new paradigms, *IEEE Wireless Communications*, August 2005, p.p. 50-58.

[3] F. Zhai, Y. Eisenberg, T. N. Pappas, R. Berry, A. Katsaggelos, "Joint source-channel coding and power allocation for energy efficient wireless video communications," *Proc. 41st Allerton Conf. Communication, Control, and Computing*, Octomber 2003.

[4] G. Su, Z. Han, K. Liu, "Multiuser Cross-Layer Resource Allocation for Video Transmission over Wireless Networks", IEEE Networks, Vol. 20, pp. 21-27, April 2006.

[5] A. Katsaggelos,Y. Eisenberg, F. Zhai, R. Berry, T. Pappas, "Advances in Efficient Resource Allocations for Packet-Based Real-Time Video Transmission", *Proceedings od the IEEE*, Vol. 93, No. 1, January 2005.

[6] S. Pejoski, V. Kafedziski, "Error resilient wireless video transmission", Procceding of ETAI 2009 National COnference (in Macedonian), September 2009, Ohrid.

[7] R. Zhang, S. Regunathan, K. Rose, "Video Coding with Optimal Inter/Intra-Mode Switching for Packet Loss Resilience", *IEEE Journal on selected area in Communications*, Vol. 18, No. 6, June 2000, pp. 966-976.

[8] Y. Zhang, W. Gao, Y. Lu, Q. Huang, D. Zhao, "Joint Source-Channel Rate-Distortion Optimization for H.264 Video Coding Over Error-Prone Networks", *IEEE Transactions on multimedia*, Vol. 9, No. 3, April 2007, pp. 445-454.

[9] J. Hagenauer: Rate-compatible punctured convolutional codes (RCPC codes) and their applications, IEEE Transactions Communications, Vol. 36, April 1988, p.p. 389 - 400.

[10] Available at: http://euler.slu.edu/ fritts/mediabench/mb2/index/html

SESSION RCMTA I

Radio Communications, Microwave Technique and Antennas I

Dispersive TLM Z-transform based 3D Model of Left-Handed Metamaterials

Nebojša Dončov, Bratislav Milovanović, Tatjana Asenov, Zoran Stanković

Abstract – **In this paper, numerical three-dimensional (3D) model of electromagnetic left-handed metamaterials (LH MTM), is derived by using the Transmission Line Matrix (TLM) method based on Z-transforms techniques. Model, implemented in time-domain TLM solver, is capable to account for frequency dispersive behaviour of LH MTM permittivity and permeability described by the Drude function. Model accuracy and efficiency are verified on some characteristic examples in comparison with analytical solutions.**

Keywords –**metamaterials**, **TLM method, Z-transforms**, **numerical dispersive model**

I. INTRODUCTION

Metamaterials (MTM) are broadly defined as an artificial media with extreme values of effective permittivity and permeability. These artificial electromagnetic (EM) structures exhibit highly unusual properties not readily found in nature, such as the antiparallelism between phase/group velocities and the reversal of classical phenomena (Doppler effect, Vavilov-Cherenkov radiation, Snell's law, Goos–Hanchen effect) [1,2]. In recent years, metamaterials, particularly the left-handed metamaterials (LH MTM) with negative refractive index (i.e. ε, μ < 0), have been the subject of numerous researches that promise to bring about important technological and scientific advancements in telecommunications, radars and defense, nanolithography with light, microelectronics and medical imaging.

In order to validate specific EM properties and fundamental physics of LH MTM and to design advanced components based on metamaterials, a number of numerical techniques have been used. Some of them have been enhanced in the form of numerical model of metamaterials offering a direct specification of permittivity and permeability of LH MTM and much faster analysis than analog implementation of the MTM transmission line networks on circuit simulator. If the case of differential numerical techniques in the time-domain, such as Finite-Difference Time-Domain (FD-TD) method [3] and Transmission-Line Matrix (TLM) method [4] as most popular, such numerical model, if developed, can allow for time-harmonic and transient simulation of MTM and analysis of their dispersive behaviour for different excitation functions.

As far as FD-TD method is concerned, several techniques have been already developed to incorporate frequency dispersion of MTM, some of them explained and referenced in [3]. TLM method is a numerical network model of Maxwell's field equations and therefore perfectly suited to realization of metamaterials as host transmission lines with embedded lumped series capacitors and shunt inductors [2]. Following this loaded transmission line approach of metamaterials, TLM method has been enhanced in [5] with a model based on insertion of reactive periodic elements into the conventional TLM mesh of so-called link lines. However, in contrast to FDTD method, that numerical TLM model allowed the direct specification of MTM properties only at one design frequency, for which the embedded reactive stubs were calculated.

In order to develop TLM model that will capture dispersive behaviour of LH MTM in a wide frequency range, TLM method based on Z-transforms techniques is applied in this paper. This approach is fully verified for accurate time-domain description of general frequency-dependent properties in isotropic, bi-isotropic, anisotropic and nonlinear materials that can be found in nature [6,7]. Such enhanced TLM method is extended in the paper by the model based on Drude function for describing frequency dependant behaviour of LH MTM permittivity and permeability and bilinear transformation technique to transfer this dependence in the time domain. Model is incorporated into three-dimensional (3D) TLM mesh and implemented in 3D TLM*scn*-Z software, designed at the Microwave Lab at the Faculty of Electronic Engineering in Nis. This software has been already successfully applied to resonant frequency calculation of metallic cavity loaded with dielectric slab exhibiting frequency-dependent complex permittivity [8]. Accuracy and efficiency of dispersive TLM model of LH MTM are verified on some characteristic examples in comparison with analytical solutions.

II. DISPERSIVE 3D TLM MODEL OF LH MTM

In [6,7], the scattering procedures for various types of frequency-dependent so-called right-handed materials (i.e. natural materials) have been developed. Exponential Z-transform technique was mainly used to incorporate right-handed material EM properties into 1D and 3D TLM algorithm. Main task was to determine, by using an appropriate conductivity or susceptibility model, elements of some of fraction expansions forms given below:

$$(1+z^{-1})g_e(z) = g_{e0} + z^{-1}(g_{e1} + \overline{g_e(z)}) \quad (1)$$

$$(1+z^{-1})r_m(z) = r_{m0} + z^{-1}(r_{m1} + \overline{r_m(z)}) \quad (2)$$

$$(1-z^{-1})\chi_{e,m}(z) = \chi_{e,m0} - z^{-1}(\chi_{e,m1} + \overline{\chi_{e,m}(z)}) \quad (3)$$

where: g_e is normalized electric conductivity, r_m is normalized magnetic resistivity, χ_e is electric susceptibility and χ_m is magnetic susceptibility.

Authors are with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: [doncov, bata, tatjana.asenov, zoran]@elfak.ni.ac.rs

Realistic LH MTM can be characterized by using either the Lorentz or Drude dispersion function which differs in term of frequency region bandwidth in which real parts of EM properties are negative. In this paper the Drude function for both the permittivity $\varepsilon(\omega)$ and permeability $\mu(\omega)$ with identical dispersion forms [3] is used to find partial fraction expansions forms elements:

$$\varepsilon(\omega) = \varepsilon_0\left(\varepsilon_\infty - \frac{\omega_{pe}^2}{\omega^2 - j\omega\gamma_e}\right) \quad (4)$$

$$\mu(\omega) = \mu_0\left(\mu_\infty - \frac{\omega_{pm}^2}{\omega^2 - j\omega\gamma_m}\right) \quad (5)$$

where the Drude function parameters $\omega_{pe,m}$ and $\gamma_{e,m}$ represent the electric or magnetic plasma frequency and the corresponding collision frequency, respectively.

In this paper, *y* component of electric field and *z* component of magnetic field will be used as an illustration how to incorporate the Drude function of permittivity and permeability into 3D TLM Z-Transform method algorithm following notation used in [6]. Similar expression can be obtained for other electric and magnetic field components. Also, it is assumed that normalized electric conductivity and magnetic resistivity, g_e and r_m, are frequency independent. From Eqs. (4) and (5) follows that electric and magnetic susceptibility of LH MTM have a form in the **s**-domain as:

$$\chi_e(s) = \chi_{e\infty} + \frac{\omega_{pe}^2}{\gamma_e}\left(\frac{1}{s} - \frac{1}{s+\gamma_e}\right) \quad (6)$$

$$\chi_m(s) = \chi_{m\infty} + \frac{\omega_{pm}^2}{\gamma_m}\left(\frac{1}{s} - \frac{1}{s+\gamma_m}\right) \quad (7)$$

Applying rather bilinear Z-transform technique $s \rightarrow 2(1\text{-}z^{-1})/[dt(1+z^{-1})]$ to Eqs. (6) and (7), instead of exponential Z-transform, as produces much better results, electric and magnetic susceptibility can be represented in the **z**-domain as:

$$\chi_e(z) = \chi_{e\infty} - K_e(1+z^{-1})\left(\frac{1}{1-z^{-1}} - \frac{1}{B_e(1-z^{-1}A_e/B_e)}\right) \quad (8)$$

$$\chi_m(z) = \chi_{m\infty} - K_m(1+z^{-1})\left(\frac{1}{1-z^{-1}} - \frac{1}{B_m(1-z^{-1}A_m/B_m)}\right) \quad (9)$$

where the coefficients: $K_{e,m} = -\omega_{pe,m}^2 dt/(2\gamma_{e,m})$, $B_{e,m} = 1+\gamma_{e,m}dt/2$ and $A_{e,m} = 1-\gamma_{e,m}dt/2$.

Taking partial fraction expansions to represent frequency dependence of electric and magnetic susceptibility as function of the value at the previous time-step leads to:

$$(1-z^{-1})\chi_e(z) = \chi_{e\infty} - K_e + K_e/B_e - z^{-1}\left(\chi_{e\infty} + K_e - a_{1e}K_e/B_e + z^{-1}\frac{b_{1e}/4}{1-z^{-1}a_{1e}}\right) \quad (10)$$

$$(1-z^{-1})\chi_m(z) = \chi_{m\infty} - K_m + K_m/B_m - z^{-1}\left(\chi_{m\infty} + K_m - a_{1m}K_m/B_m + z^{-1}\frac{b_{1m}/4}{1-z^{-1}a_{1m}}\right) \quad (11)$$

and with their comparison with Eq. (3), elements of partial fraction expansions forms can be found as:

$$\chi_{e,m0} = \chi_{e,m\infty} - K_{e,m} + K_{e,m}/B_{e,m} \quad (12)$$

$$\chi_{e,m1} = \chi_{e,m\infty} + K_{e,m} - a_{1e,m}K_{e,m}/B_{e,m} \quad (13)$$

$$\overline{\chi_{e,m}(z)} = z^{-1}\frac{b_{1e,m}/4}{1-z^{-1}a_{1e,m}} \quad (14)$$

where the coefficients: $a_{1e,m} = A_{e,m}/B_{e,m}$ and $b_{1e,m}/4 = K_{e,m}4\gamma_{e,m}dt/[2B_{e,m}{}^3]$.

Update scheme for *y* component of electric and *z* component of magnetic field includes additional accumulators S_{ed} and S_{md}, respectively, that can be evaluated by using state variables defines as: $X_{1e} = z^{-1}V_y/(1-z^{-1}a_{1e})$ and $X_{1m} = z^{-1}i_z/(1-z^{-1}a_{1m})$:

$$S_{ed} = 4\overline{\chi_e(z)}V_y = b_{1e}X_{1e} \quad (15)$$

$$S_{md} = 4\overline{\chi_m(z)}i_z = b_{1m}X_{1m} \quad (16)$$

Finally, the complete key step in scattering calculation related to *y* component of electric and *z* component of magnetic field, with a frequency independent normalized electric conductivity and magnetic resistivity, g_e and r_m, can be represented as:

$$V_y = T_e(2V_y^r + z^{-1}S_{ey}), \quad i_z = T_m(-2i_z^r + z^{-1}S_{mz}) \quad (17)$$

$$S_{ey} = 2V_y^r + k_eV_y + b_{1e}X_{1e}, \quad S_{mz} = -2i_z^r + k_mi_z + b_{1m}X_{1m} \quad (18)$$

$$X_{1e} = z^{-1}a_{1e}X_{1e} + z^{-1}V_y, \quad X_{1m} = z^{-1}a_{1m}X_{1m} + z^{-1}i_z \quad (19)$$

with the coefficients: $T_e = (4+g_e+4\chi_{e0})^{-1}$, $T_m = (4+r_m+4\chi_{m0})^{-1}$, $k_e = -(4+g_e-4\chi_{e1})$ and $k_m = -(4+r_m-4\chi_{m1})$.

III. Numerical Analysis

Accuracy of developed dispersive 3D TLM model of metamaterials is verified on the example of uniform 10-GHz TEM wave propagation through metamaterial slab inserted between two air slabs. Problem geometry is shown in Fig.1.

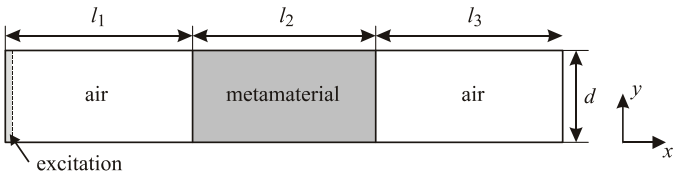


Fig.1 Metamaterial slab inserted between two air slabs

The length of each slab is l_1= l_2= l_3=70 mm and width is d=10 mm. Parameters of the Drude function are chosen so that at 10 GHz, the refractive index of the LH MTM is n=-2 (ε_r = μ_r = -2 at 10 GHz) and its characteristic impedance is 377 Ω. Lossless case is considered ($\omega_{pe,m} >> \gamma_{e,m}$). Excitation is in the form of monochromatic 10-GHz wave with electric field polarized in z direction. For modelling of this basically two-dimensional structure, uniform TLM mesh with 210×10×1 nodes is used. Electric and magnetic walls, as boundary conditions, are applied in z and y planes, respectively, in order to support TEM wave propagation. Electric field distribution at times t_1 and t_2, separated by 10dt interval, is shown in Fig.2 where it can be seen that the wave impedance of LH MTM slab is matched to air, its phase velocity is negative and, as its wavelength, is half than in air.

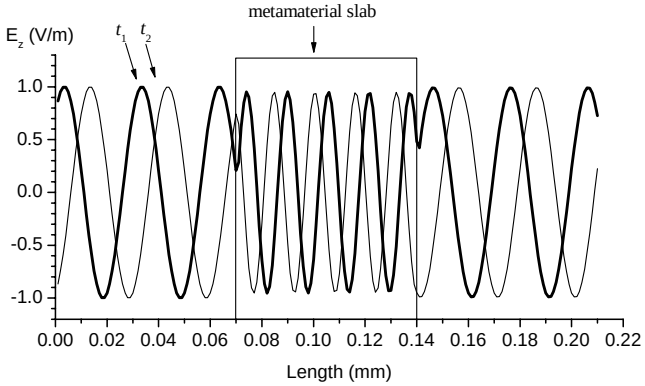


Fig.2. Electric field distribution, t_2– t_1=10dt

Capability of dispersive 3D TLM model to account for frequency dispersive behaviour of LH MTM is illustrated on the example of reflection coefficient calculation of air-metamaterial interface in a wide frequency range. Parameters of the Drude function are chosen differently for permittivity and permeability of LH MTM so that for an example ε_r = -2 and μ_r = -1 at 10 GHz, giving the reflection coefficient magnitude of 0.172. Frequency dependance of real part of relative permittivity and permeability of LH MTM is shown in Fig.3.

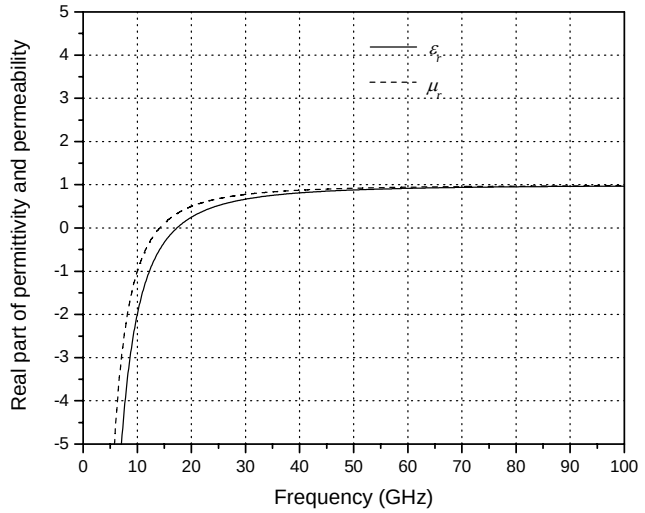


Fig.3: Real part of relative permittivity and permeability of LH MTM versus frequency

Initial pulse of Gaussian form is used to inject energy in the considered frequency range. The numerical procedure for wide bandwidth reflection coefficient calculation of air-metamaterial interface was performed in two TLM time domain simulation steps. At first step, the entire problem space was filled with air and electric field strength versus time, at a position one cell in front of the place where air-water interface should be placed, gave the incident field. A second simulation was performed when air-metamaterial interface existed, applying the dispersive TLM model of LH MTM properties. The result of the second simulation provided the total field at a position one cell in front of the interface. The reflected field was obtained by subtracting the incident field result (no metamaterial) from the total field result with the metamaterial interface present. The incident and reflected electric field versus time are shown in Fig.4.

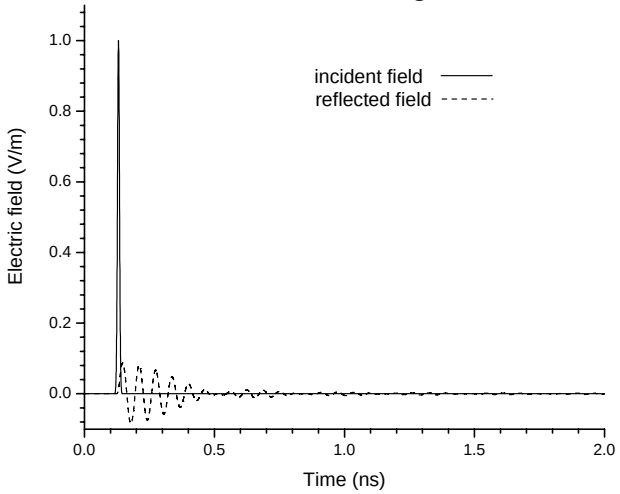


Fig.4 The incident and reflected electric field, at a position one cell in front of the air-metamaterial interface, versus time

The incident and reflected electric field data in the time domain were transformed to the frequency domain via discrete Fourier transform. The reflection coefficient magnitude at each frequency, calculated by dividing the transforms of the reflected field and the incident field, is in very good agreement with the analytic solution (Fig.5).

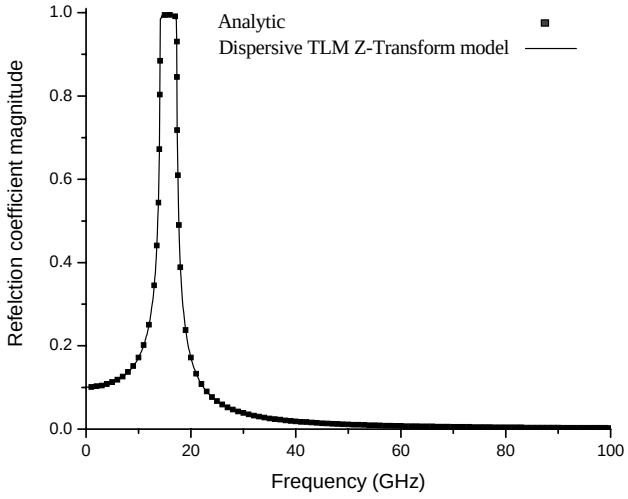


Fig.5 Reflection coefficient magnitude of air-metamaterial interface versus frequency

IV. Conclusion

3D dispersive TLM model of metamaterials based on Z-Transform of the Drude function, that describes EM poperites of LH MTM, has been derived and implemented. Accuracy of the model to account for frequency dispersive LH MTM behaviour is illustrated on the several examples which confirmed the theoretically predicted behaviour of metamaterials and reproduced the results of an excellent agreement with analytical solutions.

References

[1] G.V.Eleftheriades, K.G. Balmain, *Negative refraction metamaterials*, John Wiley & Sons, 2005.

[2] C. Caloz, T. Itoh, *Electromagnetic Metamaterials: Transmission Line Theory and Microwave Applications*, John Wiley & Sons, 2006.

[3] Y. Hao, R. Mitra, *FDTD Modeling of Metamaterials – Theory and Applications*, Artech House, 2009.

[4] C. Christopoulos, *The Transmission-Line Modelling (TLM) Method in Electromagnetics*, Publication in the Morgan & Claypool Publishers series, 2006.

[5] P.P.M. So, H. Du, W.J.R. Hoefer, *Modeling of Metamaterials with Negative Refractive Index using 2-D Shunt and 3-D SCN TLM Networks*, IEEE Transactions of Microwave Theory and Techniques, Vol.53, No.4, pp.1496-1505, 2005.

[6] J. Paul, C. Christopoulos, D.W.P. Thomas, *Generalized Material Models in TLM – Part I: Materials with Frequency-dependent Properties*, IEEE Transactions on Antennas and Propagation, Vol.47, No.10, pp.1528-1534, 1999.

[7] J. Paul, C. Christopoulos, D.W.P. Thomas, *Generalized Material Models in TLM – Part II: Materials with Anisotropic Properties*, IEEE Transactions on Antennas and Propagation, Vol.47, No.10, pp.1535-1542, 1999.

[8] B. Milovanović, N. Dončov, J. Joković, T. Dimitrijević, *TLM-Z Method Modelling of Microwave Cavity Loaded with Frequency-Dependent Dielectric Slab*, Proceedings of the 43rd International Scientific Conference on Information, Communication and Energy Systems and Technologies, ICEST 2008, Nis, Serbia, Vol.2, pp.355-358, 2008.

Modeling Shielding Effectiveness Measurements for Nano-scale Samples

Szilvia Nagy[1] and András Fehér[2]

Abstract **– If thin metallic layers grown on compound semiconductor substrates are heat treated Ohmic contacts can be formed. In case of the layers being thin enough, fractal-like topology of the metallic clusters can occur instead of continuous layers. This fractal behavior can be seen in many other cases for various material systems. The layers arising are mostly analyzed by a kind of scanning microscopy or other 2D and 3D image generating methods. Mechanic and electronic properties, like the conductivity, can be determined, too.**

In the present work a sample holder for measuring another electromagnetic property, namely the shielding effectiveness is being studied. The shielding effectiveness of these structures carry information about the quality of the Ohmic contact built during the growth process. Furthermore, the connection of the topology and the shielding properties of various material systems can be mapped. The equipment for determining the shielding effectiveness or the reflectivity of such small scale, thin, planar samples is modeled by finite element method.

According to the model, the *S* matrix elements of the developed sample holder depends only slightly on the shape and the position of the specimen under test.

Keywords **– Shielding effectiveness, EMC, compound semiconductor**

I. INTRODUCTION

Ohmic contact can be generated on compound semiconductor substrates by growing thin metallic layers upon the surface and heat treating the system. The Ohmic contacts occur depending on the composition of the materials and the growing and thermal treating circumstances. If the layers of are thin enough, the metal can form clusters with fractal-like topology [1,2,3], instead of continuous layers. Fullerene layers molecular beam epitaxially grown on vanadium selenide substrate form a system of islands consisting of more plateaus with different fractal dimensions and inner structure [4]. Measuring shielding effectiveness of these materials with nano-scale patterns on their surface can provide a lot of information about the layers without actually measuring them on scanning electron or atomic force microscopes.

For measuring shielding effectiveness or permittivity more methods are available. Standard IEEE-STD-299 requires a shielded enclosure with a hole in one of the side walls and a really large sample, thus its prescriptions can not be fulfilled with compound semiconductor slices. Within the other possibilities, like partially filling a microwave line with the sample, or touching a specially designed head to the bulk specimen, we have chosen the only method, which can be adapted for a usually 300 micron thick, planparallel, but otherwise irregularly shaped, rigid and fragile slice of compound semiconductor sample. The sample holder of this method is a coaxial aerial, cut in two halves perpendicular to its axis, similar to the device in the former standard [5]. The variation of this sample holder appears in many fields of electromagnetic compatibility literature, from metal coated fibers [6,7], polymers [8] to nanomaterials [9, 10]. In the later applications, only one of the coaxial aerial halves is used, the other part is substituted by a prefect metallic reflector, which can easily be gilt, and gives a possibility to heat the sample.

II. SAMPLE HOLDER DESIGN

The sample holders are designed so, that they would match the standard N-type connectors of the network analyzer. The planned measuring setup is according to standard [5], except for the frequency range, which is in our case 100 MHz to 10 GHz. Probably, later wider frequency range could be also studied.

The sample holder is an enlarged coaxial aerial separated in two halves, and the material under test is held between these parts as it is shown in the upper part of Fig. 1. The lower part of Fig. 1 illustrates the method with only one coaxial lin part and a reflector [10].

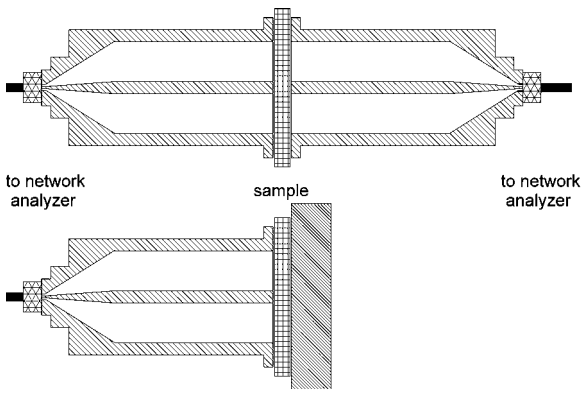


Fig. 1. Sample holder for the ASTM D4935 measurement, and a variation of it with a reflector for nano materials.

[1]Szilvia Nagy, Faculty of Engineering Sciences, Széchenyi István University, Egyetem tér 1, H-9026 Győr, Hungary. E-mail: nagysz@sze.hu

[2]András Fehér, Faculty of Engineering Sciences, Széchenyi István University, Egyetem tér 1, H-9026 Győr, Hungary. E-mail: afeher@sze.hu

For the sample holder given in the upper subplot of Fig. 1 one side of the coaxial transmission line is driven by a network analyzer while the transmitted signal is measured on the other side. The size of the equipment, according to ASTM D4935, is 5.24" for the outermost diameter, the inner diameter of the outer cylinder is 3", while the diameter of the inner conductor copper rod is 1.3". Of course, these parameters can be scaled, in [2] a 133 mm flange diameter specimen holder with the coaxial aerial diameters 76 mm:33 mm, and a smaller, 14 mm:6 mm coaxial line was tested with 33 mm and 44 mm flanges. The size of the sample holders for nano materials was significantly smaller. Our specimen holder is filled with Teflon, thus the scaling is a little bit different (according to

$$Z_0 = \frac{60}{\sqrt{\varepsilon_r}} \ln \frac{b}{a} \tag{1}$$

with a and b being the inner and outer radii of the coaxial line, which were in our case 1.5 mm and 5 mm.

According to the standard [5], in order to have the same capacitive coupling between the two halves of the device with or without sample, a reference should be prepared having the same material properties as the sample, but covering only the conductive surfaces of the sample holder surface. The shape of the reference and the sample according to the standard is given in Fig. 2.

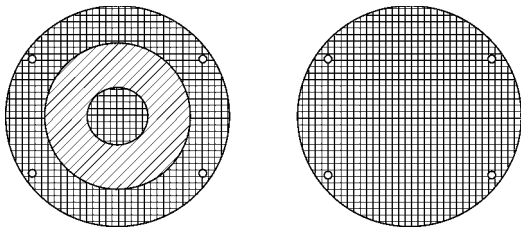


Fig. 2. Geometry of the reference and the sample according to [5].

In case of the reference, the dark circle and ring in Fig. 2 is made of the same material as the sample. The reference as well as the sample can be deposited on the surface of a dielectric disc, which is symbolized by the lightly covered area between the parts of the reference. Both objects should be measured at the same frequencies, and the shielding effectiveness *SE* is the ratio of the received reference and the load power,

$$SE = 10 \log \frac{P_{\text{ref}}}{P_{\text{sample}}}. \tag{2}$$

The thickness of the studied material must be much less than the free wavelength of the electromagnetic wave propagating in the coaxial transmission line in TEM mode.

With a network analyzer or a simple directional decoupler the reflected signal can also be measured. Using both the transmitted and the reflected signals, the *S* matrix parameters S_{21} and S_{11} can be measured. [11]

The sample holder operates correctly up to the frequency, where besides the TEM mode, higher order modes appear [5], i.e., until

$$f_c = \frac{1}{\pi\varepsilon} \frac{2c}{a+b} \tag{3}$$

with c being the speed of the light. For the above described sample holder this cut-off frequency is a little less than 10 GHz.

Studying Fig. 2, one can clearly say, that it is almost impossible to prepare reference and sample from GaAs or other compound semiconductor of such shape, and it is rather hard to mask the reference during the growth and thermal treatment processes, according to the figure, and still providing the same material properties as those of the sample.

The results theoretically calculated [3] and measured according to standard ASTM D4935 are usually in sufficient correlation. This gave the idea, that instead of the reference, which can not be prepared for compound semiconductor substrates with thin metallic layers with nano-scale topology, we can model the results and measure only the sample and the empty substrate.

III. SYMULATIONS

For simulation we have used the RF toolbox of the finite element program package COMSOL Multiphysics. Although the sample holder was almost cylindrical, because of the rectangular flanges, no symmetry of the device could be used. However, because of the almost cylindrical samples, i.e., the slices of a grown GaAs crystal, the model contained cylindrical sample. This sample was shifted and cut for studying the effects of various sized samples with various geometry. The applied frequency domain was 100 MHz to 10 GHz, the geometries can be identified from the resulting plots.

First, the sample holder without sample and with air filling was tested, and after verifying, that the results agree with the theory, we have placed a cylindrical sample coaxially with the specimen holder in between the flanges. The flanges are designed to be large enough, so that the field at its edges should be negligible. The resulting electric and magnetic field is summarized in Fig. 3.

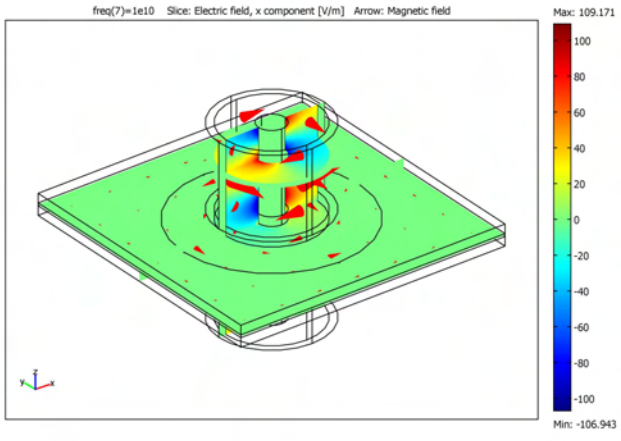


Fig. 3. Electric and magnetic field plotted in color slices and arrows for the centrally placed sample. The inner conductor is of diameter 3 mm, the outermost conductor's diameter is 10 mm.

The slightly disturbed TEM mode of a coaxial line is recognizable. Clearly, the electromagnetic field concentrates at the centre of the sample. Similar results arose in all of the studied frequencies, the energy density between the flanges was very small. This result suggests, that the shape of the sample is indifferent as long as it covers the high energy density central region of the sample holder. To test this suggestion, we have modified the sample, by shifting it 3 mm off the center, and by cutting a slice of it, and filling with air at the outer cylinder of the sample holder. The two resulting electromagnetic field plots are given in Fig. 4.

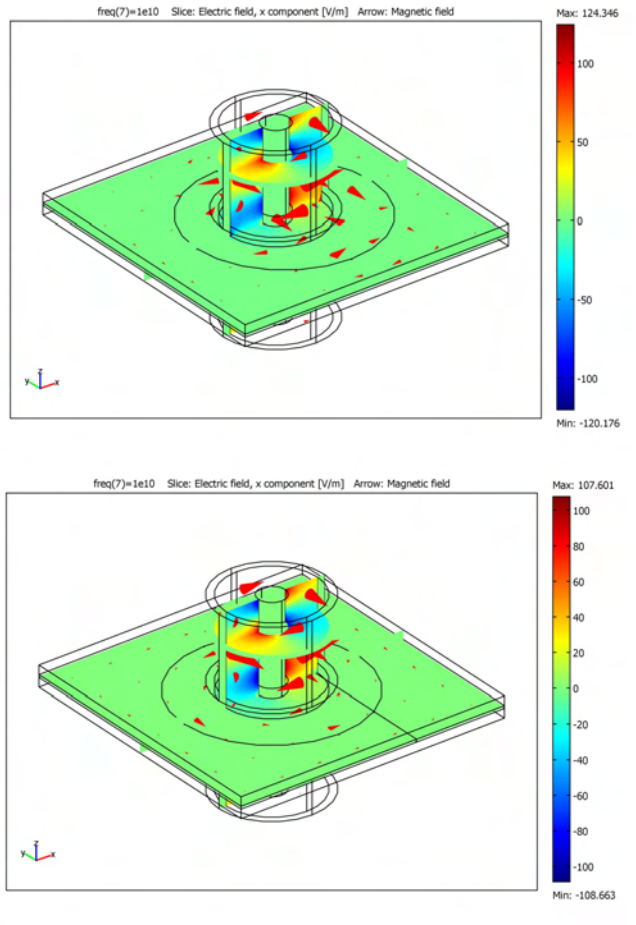


Fig. 4. Electric and magnetic field plotted in color slices and arrows for the off-centered placed and sliced samples, respectively.

Differences of the electromagnetic field are visible not only in the outer regions of the flanges, but in the coaxial line, too.

The matrix elements S_{11} and S_{21} for all the samples were calculated for the 100 MHz to 10 GHz region of frequency, and plotted in Fig. 5. The results of the various samples are not distinguishable on this scale, except at 10 GHz, which is out of the operating domain of the sample holder.

In order to study the differences of the *S* matrix elements of the modified samples from those of the original one are given in Fig. 6. It can be seen, that for lower frequencies the shifting of the samples cause less error, than the irregular, not cylindrical shape of it, while at higher frequencies, especially around the cut-off frequency, it results in resonances. The irregular shape of the sample seem to shift the *S* matrix elements with a slightly larger value, but it is not affected by reaching the cut-off frequency. Since the values of S_{21} are larger at the lower frequency domain and smaller at high frequencies, its relative error is less for the cylinder slice than for the shifted cylinder. Opposite statement holds for S_{11}.

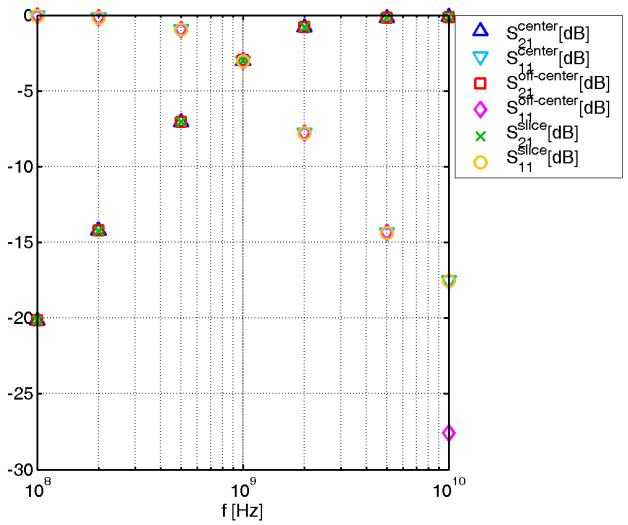


Fig. 5. *S* matrix elements for the, centered, off-centered placed and sliced samples, respectively.

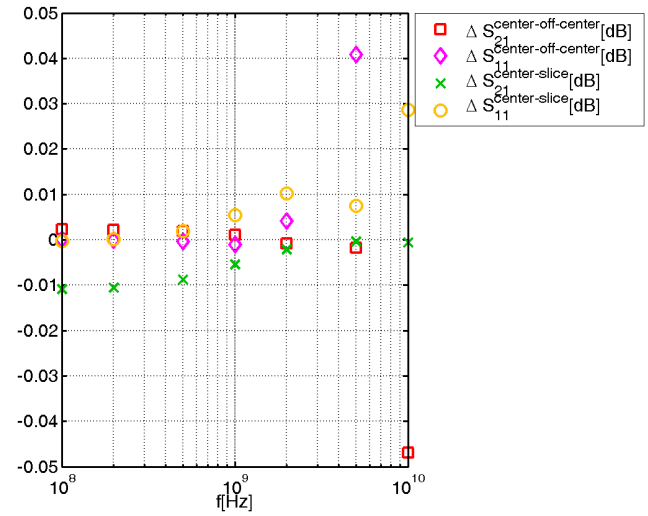


Fig. 6. Differences of the *S* matrix elements from the original, centered, cylindrical sample for the off-centered placed and sliced samples, respectively.

IV. Conclusion

A coaxial line based sample holder for thin, small, rigid, irregularly shaped planar samples is studied both by numerical simulations and by measurements. The numerical study showed, that due to the large flanges of the specimen holder

the non central regions of the sample affect only slightly the main results, which was also proven by network analyzer measurements previously. Comparing numerical and measured data can result in shielding effectiveness and dielectric constant determination methods without reference sample.

ACKNOWLEDGEMENT

Financial support of the Bolyai János Research Fellowship of the Hungarian Academy of Sciences is gratefully acknowledged.

REFERENCES

[1] I. Mojzes, Cs. Dominkovics, G. Harsányi, Sz. Nagy, J. Pipek, L. Dobos, "Heat treatment parameters effecting the fractal dimensions of AuGe metallization on GaAs", Appl. Phys. Lett., vol. 91, art.No. 073107, 2007.

[2] M. Schuszter, L. Dobos, K.A. Demcu, Sz. Nagy, I. Mojzes, "Analysis of Morphology Changes of Heat Treated Metallization of Compound Semiconductors by the Fast Wavelet-transform based on B-Spline", J. Optoelectron. Adv. M., vol. 9, pp. 2241-2244, 2007.

[3] L. Dávid, L. Dobos, B. Kovács, I. Mojzes, B. Pécz, "Fractal character of in situ heat treated metal-compound semiconductor contacts", J. Mater. Sci: Mater. Electron, vol. 17, pp. 321 2006.

[4] Á. Nemcsics, Sz. Nagy, I. Mojzes, R. Schwedhelm, S. Woedtke, R. Adelung, L. Kipp, "Investigation of the surface morphology on the epitaxially grown fullerene structures", Vacuum, vol. 84, pp. 152-154, 2010.

[5] Former American Standard ASTM D4935.

[6] K. Bula, J. Koprowska and J. Janukiewicz, "Application of Cathode Sputtering for Obtaining Ultra-thin Metallic Coatings on Textile Products", Fibers and Textiles in Eastern Europe Vol. 14, pp. 75, 2006.

[7] T. W. Więckowski and J. M. Janukiewicz, "Methods for Evaluating the Shielding Effectiveness of Textiles", Fibers and Textiles in Eastern Europe Vol. 14, pp. 18, 2006.

[8] R. B. Shulz, V. C. Plantz and D. R. Brush, IEEE Trans. EMC, Vol. 30, pp. 187, 1988.

[9] J. E. Fischer, H. Dai, A. Thess, R. Lee, N. M. Hanjani, D. L. Dehaas and R. E. Smalley, Phys. Rev. B, Vol. 55, pp. 4921, 1997.

[10] H. Xu, S. M. Anlage, L. Hu, and G. Gruner, Appl. Phys. Lett., Vol. 90, 183119, (2007).

[11] A. Fehér, Sz. Nagy, and I. Mojzes, "Developing sample holders for measuring shielding effectiveness of thin layers on compound semiconductor substrates", Proceedings of PIERS 2009, Moscow, 18-21 August, 2009, pp. 1458.

Verification of TLM Compact Wire Model in Cylindrical Mesh applied to Determining of Transmission Coefficient in Cavity

Tijana Dimitrijević, Jugoslav Joković, Nebojša Dončov

***Abstract* – This paper presents a verification of a compact wire model implemented in the 3-D TLM cylindrical mesh based on comparison of numerical and measured results of cavity transmission characteristics. The TLM compact wire model in a cylindrical mesh is based on wire structures parameters calculation in conditions of variable cross-section of the TLM nodes through which wire conductor passes due to nature of cylindrical grid along the wire path. Obtained numerical results have been compared with the corresponding results reached by TLM method based on a rectangular grid, and the correctness of the method has been experimentally verified.**

***Keywords* – TLM method, cavity, wire model, cylindrical mesh, transmission characteristic**

I. INTRODUCTION

A cylindrical metallic cavity represents a configuration very suitable for good modelling of devices used in some practical microwave applications, such as applicators used in the processes of dielectric material heating and drying or power dividers used for distribution of information in communications systems [1,2]. Even though there are many ways to couple energy into the cavity [1], input and output ports of microwave cavity devices are generally realized by coaxial probe that ensures coupling with corresponding electromagnetic (EM) field component. Therefore, the reflection (S_{11}) and transmission (S_{21}) characteristics are common parameters in cavity exploration.

For purpose of an analysis of metallic cavities the TLM (Transmission-Line Matrix) time-domain method [3], as a general, electromagnetically based numerical method, is very suitable [4,5]. In recent years, TLM enhancement in form of a compact model for wire structures has been developed [6], yielding a significant improvement in the required computer resources compared to the traditional TLM method. Model allows for accurate modelling of wires with a considerably smaller diameter than the TLM node size. It uses a special wire network embedded between nodes to model signal propagation along the wires, while allowing for interaction with the electromagnetic field. Its implementation into the rest of TLM mesh, based on rectangular grid, is straighforward as mean cross-section dimensions of nodes through which wire conductor runs can be easily kept constant along the wire path.

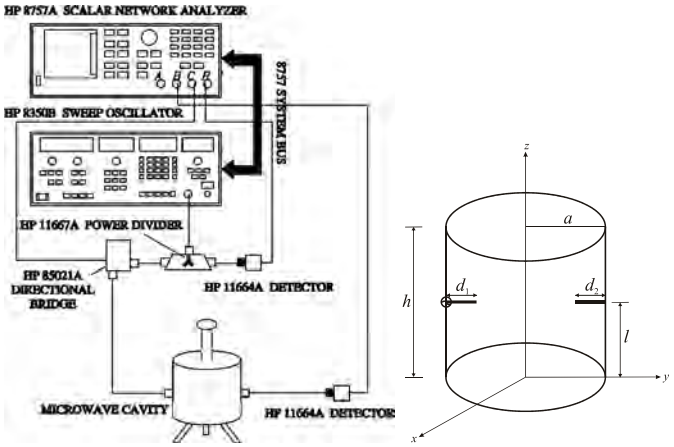
Authors are with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: [tijana.dimitrijevic,jugoslav.jokovic,nebojsa.doncov]@elfak.ni.ac.rs

When modelling of structures of rectangular geometry is concerned, the network of TLM cubic shaped nodes in a Cartesian grid comes up as a good solution. However, for problems with cylindrical or spherical symmetry it is more appropriate to use nodes that model directly non-cubic shaped blocks, enabling problem boundaries to be described more accurately. If a rectangular mesh is used to model a cylindrical structure, a curved boundary would have to be described in a step-wise fashion which might result in deviation of resonant frequencies values as well as in excitation of unwanted modes. Since a numerical error caused by step-wise approximation depends on the mesh resolution, it could be reduced by applying the TLM mesh of greater resolution, which, on the other hand, results in increasing a simulation time. Moreover, increasing of a mesh resolution is limited since an implementation of the compact wire model into the rest of the TLM mesh demands that the ratio between dimensions of a wire conductor and dimensions of nodes through which the wire conductor passes is optimal. Thus, a greater resolution of applied rectangular TLM mesh enables a cylindrical cavity to be precisely modelled only if a probe of relatively small radius is used [7, 8].

The fact that two opposite demands have to be fulfilled is the main disadvantage of using a rectangular grid for modelling of a probe-fed cylindrical cavity. In order to overcome such limitations, even though a procedure of a compact wire model implementation is much easier in a rectangular grid compared to a cylindrical grid, we found more profitable to implement the wire model into the cylindrical grid. This approach enables the precise modelling of cavity boundaries independently of mesh resolution applied, but an error may arise due to variable mean cross-section dimensions of cylindrical TLM nodes, through which a wire conductor (placed in the radial direction) passes, resulting in different wire network properties from one node to another. This has been solved by implementation of an additional connecting procedure for wire segments belonging to TLM nodes with different cross-sections into the TLM algorithm.

Efficiency of the wire model adapted to cylindrical mesh was verified on the case of an empty cylindrical metallic cavity of radius *a* and height *h*. The transmission coefficient, based on two probes inserted into the cavity, has been considered numerically and experimentally (Fig.1). For a numerical modelling purpose, according to experimental procedure, one probe was used as the real feed, while other was used to monitor established distribution of EM field inside the cavity.

The TLM results, corresponding to the transmission coefficient obtained for various probes length in the frequency range f = [1.5 ÷ 3.5] GHz, have been evaluated against the results reached by the TLM method based on the rectangular coordinate system and experimental results obtained by using the set-up shown in the Fig.1a.



a) b)

Fig. 1. a) Experimental set up for resonant frequency measurement b) Cylindrical cavity with feed and receiving probes

II. TLM MODELLING

In the TLM time-domain method, EM field strength in three dimensions, for a specified mode of oscillation in a cylindrical metallic cavity, is modelled by filling the field space with a network of link lines and exciting a particular field component through incident voltage pulses on appropriate lines [4]. EM properties of different cavity loads are modelled through network of interconnected nodes, known as a symmetrical condensed node (SCN)]. Each node describes a portion of the medium shaped like a cubic (Cartesian rectangular mesh) or a slice (Non-Cartesian cylindrical mesh) depending on the coordinate system applied. Additional stubs can be incorporated into TLM model to account for inhomogeneous materials and/or electric and magnetic losses.

When cylindrical structures are concerned, a non-Cartesian cylindrical mesh in the coordinate system (φ, r, z) can be used for the modelling purpose. The coordinate system used and the port designations are shown in Fig. 2. Simulation proceeds exactly as for a SCN with stubs in a Cartesian grid. The only modification involves the calculation of stub parameters where account must be taken of the details of the new geometry.

The TLM wire node is based on a SCN with one small modification in the form of additional link and stub lines interposed over the existing network to account for increase of capacitance and inductance of the medium caused by wire presence [6]. The single column of TLM nodes, through which a wire conductor passes, can be used to approximately form the fictitious cylinder which represents capacitance and inductance of a wire per unit length. Its effective diameter, different for capacitance and inductance, can be expressed as a product of factors empirically obtained by using known characteristics of the TLM network and the mean dimensions of the node cross-section in the direction of wire running [6].

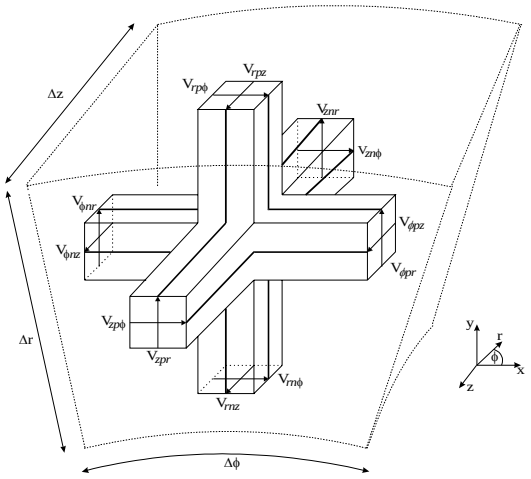


Fig. 2. A cylindrical SCN

An equivalent radius of the fictive cylindre in a cylindrical grid for calculating the capacitance and inductance, r_{Cr} and r_{Lr}, respectively, for wire segment running along r direction are $r_{Cr} = k_{Cr}\Delta r_c$ and $r_{Lr} = k_{Lr}\Delta r_c$, where Δr_c represents mean dimension of the node cross-section in r direction ($\Delta r_c = \left(\frac{(r_i + r_{i+1})}{2} \Delta\varphi + \Delta z \right) / 2$, where r_i and r_{i+1} are lower and upper limits of the TLM wire node in radial direction (Fig.3), while k_{Cr} and k_{Lr} are factors empirically obtained by using known characteristics of TLM network.

Distributed capacitance and inductance per unit length, needed for modelling of wire segments, may be expressed as:

$$C_{wr} = \frac{2\pi\varepsilon}{\ln(r_{Cr} / r_w)},\ L_{wr} = \frac{\mu}{2\pi}\ln(r_{Lr} / r_w) \qquad (1)$$

where r_w is a real probe radius.

An equivalent radius of the fictitious cylinder can be easily kept constant along nodes column in a rectangular grid. However, in a cylindrical grid for a wire conductor in the radial direction, as it is shown in Fig. 3, mean cross-section dimensions of TLM nodes, through which a wire passes, vary making difficult to preserve distributed capacitance and inductance of a wire per unit length.

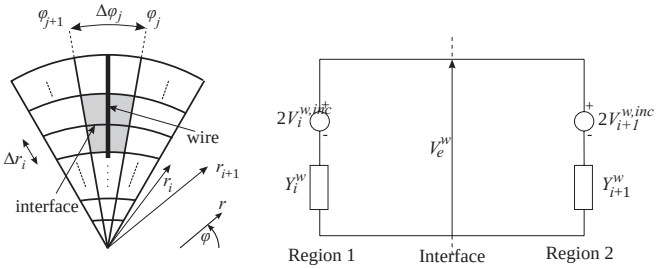


Fig. 3. TLM nodes in $r\varphi$ plane through which wire runs and an interface between two nodes

As result, admittance of the wire network link line, interposed over the existing network to account for wire presence, varies from one TLM node to another (Fig.3). To

solve this, an additional connecting procedure for wire segments with different link-lines admitances has been implemented into the existing TLM-based software.

Reflected voltages on both directions of the interface between nodes with different cross-section, which at the same time represent incident voltages respect to the node center for the next time step, can be expressed as follows:

$$V_i^{w,ref} = \frac{Y_i^w - Y_{i+1}^w}{Y_i^w + Y_{i+1}^w}\left(V_i^{w,inc} - V_{i+1}^{w,inc}\right) + V_{i+1}^{w,inc} \quad (2)$$

$$V_{i+1}^{w,ref} = \frac{Y_i^w - Y_{i+1}^w}{Y_i^w + Y_{i+1}^w}\left(V_i^{w,inc} - V_{i+1}^{w,inc}\right) + V_i^{w,inc} \quad (3)$$

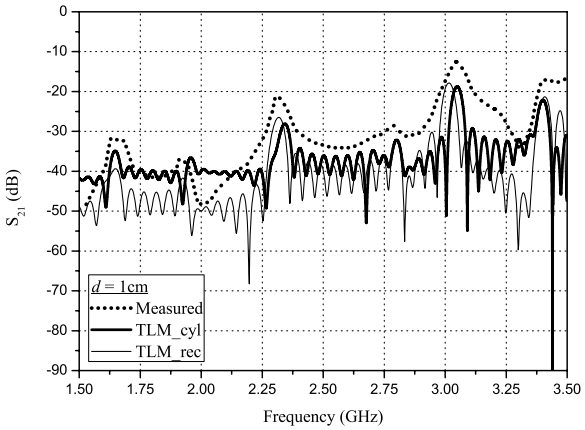
where $V_i^{w,inc}$ and $V_{i+1}^{w,inc}$ are the incident voltages.

III. Numerical Results

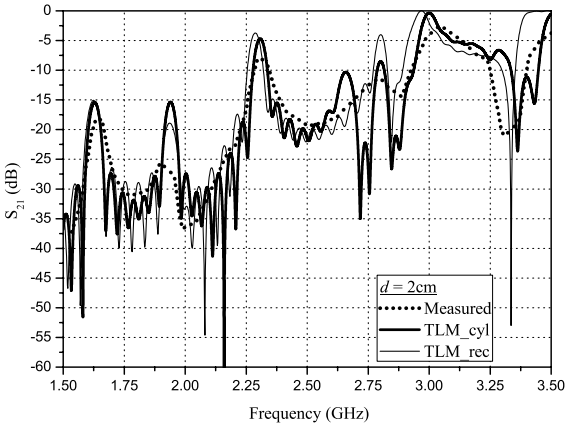
The TLM method based on a cylindrical grid enhanced with the wire node model has been used to analyze resonant modes distribution within the considered cavity containing two wire probes, representing feed and receiving probe. Dimensions of the modelled cavity, a = 7 cm and h = 14.24 cm, was chosen to follow the experimental ones (Fig. 1b). The feed probe was placed at the height l = 7.4 cm from the cavity bottom, slightly different from $h/2$, in the radial direction, whereas the receiving probe was placed at the same height and direction, opposite to the feed probe. In this way, it is possible to excite and simultaneously detect modes which have the radial component of the electrical field in the cavity. In order to model a real coaxial cable characteristics, the feed probe was connected, through the TLM wire port, to the real voltage source: V_{source} = 1 V with the resistance R_{port1} = 50 Ω, whereas the resistance of the receiving port was R_{port2} = 50 Ω. For cavity space modelling, a cylindrical TLM mesh of resolution ($\varphi \times r \times z$) = (36×28×32) was used.

The radius of probes has been kept constant r = 0.5mm, whereas theirs lengths have been varied in the range $d_1 = d_2 = d$ = [1÷5]cm. In Fig. 4. the transmission coefficient characteristics obtained by TLM method based on a cylindrical grid are compared with corresponding results reached by a rectangular grid based TLM method and with experimental results as well. As can be seen, there is a very good agreement between considered results, in terms of resonant frequencies values and EM field level. Thus, possibilities of the TLM method based on a cylindrical grid as a tool for modelling and analyses of transmission procedure in an empty cavity are confirmed. It should be noted that the difference between numerical and experimental results regarding smoothness of graphs emerges as a result of numerical modelling procedure where has been assumed that the walls have been made out of perfectly conducted metal.

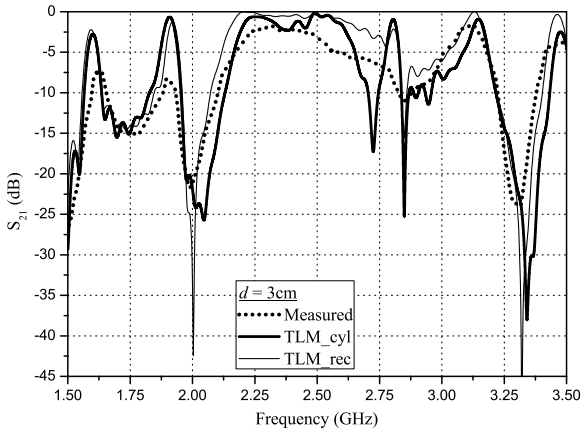
However, in the case of a small probe length, as for the length of d = 1cm, it is apparent that, compared to the TLM method based on a rectangular grid, a cylindrical grid based TLM method gives better agreement with the experimental results in terms of the EM field level. There is a deviation of about 10dB when a rectangular grid based TLM method is used. One of the reasons for this feature is that in the case of a small probe length a numerical error occurs due to applied TLM rectangular network resolution which results in a small number of nodes through which wire conductor passes. This error could be avoid if the TLM rectangular network of greater resolution is being used, but, it demands that a radius of a probe is reduced in order to preserve the optimum ratio between a probe radius and TLM node dimensions requested when the compact TLM wire model is applied.



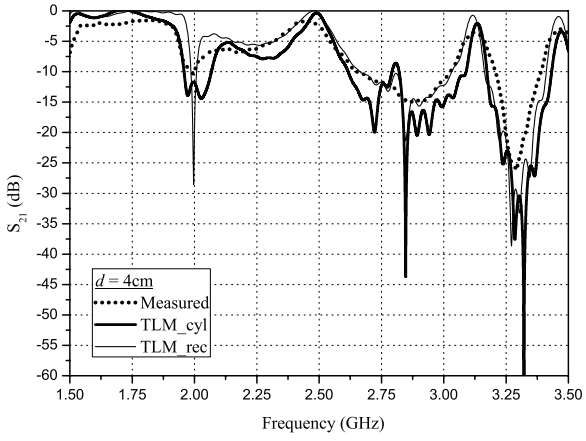
a)



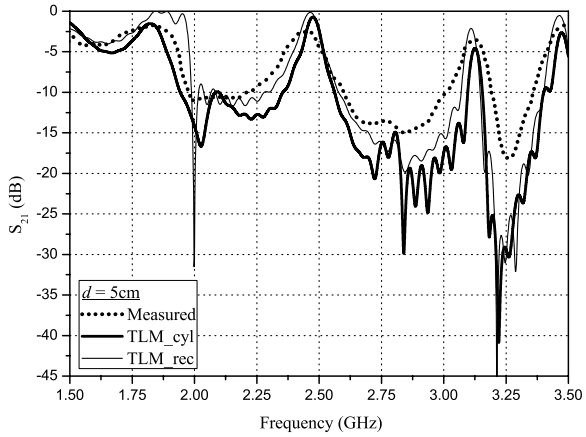
b)



c)



d)



e)

Fig. 4. The transmission coefficient for various feed and receiving probes lengths a) 1cm, b) 2cm, c) 3cm, d) 4cm, e) 5cm

Furthermore, a cylindrical grid enables modelling of probes with much greater radius compared to the TLM method based on a rectangular grid. Fig. 5 shows dependence of a maximum radius that could be modelled on the probe length. In the case of cylindrical cavity modelled in the rectangular grid of resolution ($x \times y \times z$) = (43×43×32), the probe radius can be maximally 0.5 mm [8] so the modelling process could be carried out. On the other hand, when the cylindrical grid ($\varphi \times r \times z$) = (36×28×32) is being used the maximum value of the probe radius that could be modelled is much greater and depends on the probe length.

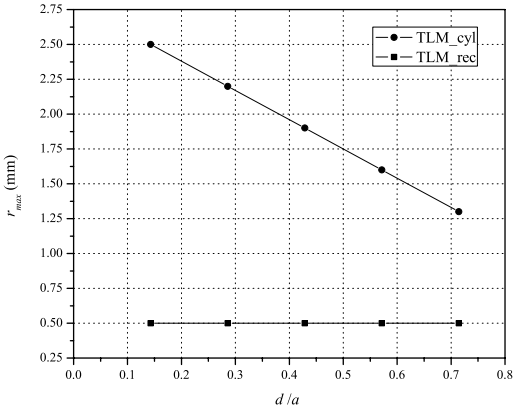


Fig. 5. Comparison of maximum probe radius vs. probe length in rectangular and cylindrical grid

IV. CONCLUSIONS

In this paper, the compact wire model implemented to TLM method based on a cylindrical coordinate system, have been experimentally verified. However, since mean cross-section dimensions of TLM nodes, through which wire conductor passes, is not constant along the nodes columns in a cylindrical grid in a radial direction, an implementation procedure is more complex to carry out in a cylindrical than in a rectangular mesh. This means that wire network parameters have to be calculated taking into account inconsistence of the capacitance and inductance of the wire per unit length. For that reason, an additional connecting procedure for wire segments with different EM properties has been implemented into the existing cylindrical grid based TLM software.

The wire model adapted to cylindrical TLM mesh has been applied to the problem of an analysis of transmission characteristics in an empty cylindrical cavity, when an EM field has been excited and detected via wire probes. The results have been verified by comparison with both experimental and results based on a TLM rectangular grid. The main advantage of using a cylindrical instead of rectangular grid for the purpose of modelling a probe-fed cylindrical cavity is that enables accurate modelling of boundaries, since a numerical error caused by the step-wise approximation of boundaries is avoided. Also, a cylindrical grid enables modelling of probes with much greater radius compared to the TLM method based on a rectangular grid.

REFERENCES

[1] T.V.C.T. Chan, H. C. Reader "Understanding microwave heating cavities", Artech House, Boston, London, 2000.

[2] S. Amnartpluk, C. Phongcharoenpanich, S. Kosulvit, M. Krairiksh, A Power Divider using Linear Electric Probes Coupling Inside Conducting Cylindrical Cavity, Proc. of the 2003 International Symposium on Circuits and Systems, Vol. 3, pp.419-422, 2003.

[3] C. Christopoulos, The Transmission-Line Modelling Method, IEE/OUP Press, 1995.

[4] B. Milovanovic, N. Doncov, TLM Modelling of the Circular Cylindrical Cavity Loaded by Lossy Dielectric Sample of Various Geometric Shapes, Journal of Microwave Power and Electromagnetic Energy,USA, Vol.37, No.4, pp.237-247, 2002.

[5] B. Milovanovic, N. Doncov, A. Atanaskovic, Tunnel Type Microwave Applicator Analysis using the TLM method", Proc. of the 4th International Workshop on Computational Electromagnetics in the Time Domain, CEM-TD 2001, Nottingham, UK, pp.77-84, 2001.

[6] V.Trenkic, A.J.Wlodarczyk, R.Scaramuzza, "Modelling of Coupling between Transient Electromagnetic Field and Complex Wire Structures", International Journal of Numerical Modelling: Electronic Networks, Devices and Fields, vol.12, pp.257-273, 1999.

[7] J. Jokovic, B. Milovanovic, N. Doncov, TLM Analysis of a Cylindrical Metallic Cavity Excited with a Real-Feed Probe, International Journal of RF and Microwave Computer-Aided Engineering, Publication of the Wiley, USA, Vol.16, No.4, pp.346-354, 2006.

[8] J. Joković, B. Milovanović, N. Dončov, Numerical Model of Transmission procedure in Cylindrical Metallic Cavity Compared with Measured Results, International Journal of RF and Microwave Computer-Aided Engineering, Publication of the Wiley, Vol. 18, No. 4, pp.295-302, 2008.

Wideband Stacked Microstrip Antenna

Slavi R. Baev[1]

Abstract – **New configuration of stacked microstrip antenna is proposed in order to enhance its bandwidth. Additional parasitic elements - microstrip dipoles - are introduced in antenna construction in order to excite more resonances in frequency characteristic. Numerical and experimental comparison between proposed and typical stacked antenna is conducted and derived results are presented in the paper.**

Keywords – **microstrip antenna, bandwidth, parasitic dipoles.**

I. INTRODUCTION

Microstrip antennas are preferable type of radiator for many communications purposes. Some typical applications are: antennas for compact mobile terminals (GSM, GPS), wireless radio links, large antenna arrays with radiation pattern control for satellite communications. There is tendency in communication systems to decrease physical size of mobile devices and increase their functionality. Therefore antenna system should be small and should have wide frequency band. Wide bandwidth allows integration of different services and higher capacity in a mobile system.

Widespread method for bandwidth enhancement is excitation of multiple resonances in the impedance characteristic of the antenna by introduction of parasitic elements [1]. Parasitic elements can be placed in horizontal direction (planar configuration) [2], [3] or in vertical direction (stacked configuration) [4], [5] compared to the main radiator.

Combination of horizontally and vertically aligned radiating elements to achieve compact microstrip antenna with improved bandwidth is applied in [6], [7]. These publications present modified construction of aperture-coupled stacked microstrip antenna, which exhibits wider bandwidth compared to typical antenna of this type. In the proposed antenna two pairs of microstrip dipoles with different size are placed above the main radiator, instead of one wide rectangular patch. In this way two additional resonances are introduced in the impedance characteristic, instead of one.

In the present publication the idea used in [6] and [7] is further developed in order to explore the potential of the proposed method for bandwidth improvement. New antenna configuration is presented which exhibits three resonances in its characteristic similar to the radiator from [6] and [7]. In proposed antenna there is main patch on bottom substrate, parasitic patch on top substrate and two parasitic dipoles positioned adjacent to the main patch. Proposed stacked microstrip antenna is compared with typical one through simulations and experiment.

[1]Slavi R. Baev is with the Department of Radiocommunications and Video Technologies, Faculty of Telecommunications, Technical University – Sofia, 8 "Sv. Kliment Ohridski " Blvd., Sofia 1000, Bulgaria, E-mail: sbaev@tu-sofia.bg

II. ANTENNA GEOMETRY

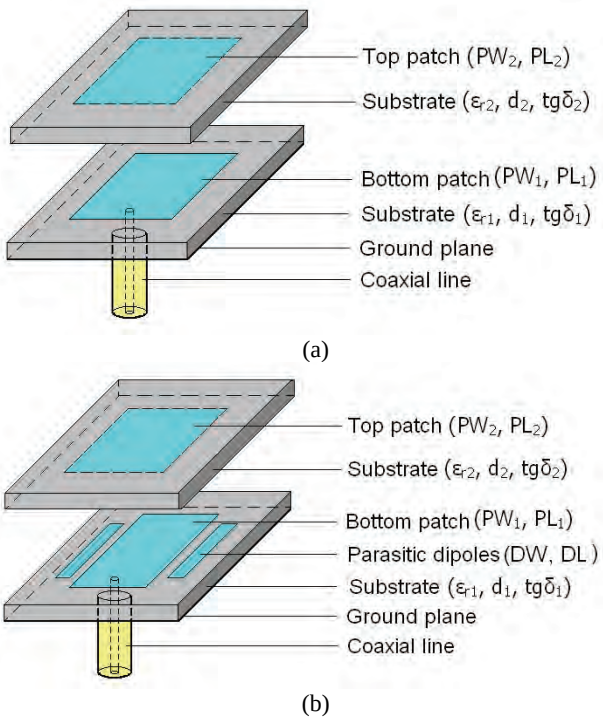


Fig. 1. 3D view of investigated antennas: (a) typical antenna (Antenna 1); (b) proposed antenna (Antenna 2)

Fig.1 presents typical stacked microstrip antenna (Antenna 1 – Fig.1a) and proposed stacked radiator (Antenna 2 – Fig.1b).

Coaxial line is used as feeding method in both antennas. It is convenient because there is no strong back radiation compared to aperture coupling method (microstrip line exciting bottom patch through slot cut in ground plane). Another problem with aperture feed is also omitted with coaxial line. In certain cases aperture may act as separate resonator and additional resonances may occur in frequency characteristic of the antenna. With coaxial feed resonances are formed only by the radiating elements above the ground plane. In this way the effects of main and parasitic patches can be clearly seen and evaluated.

The two antennas shown in Fig.1 have similar vertical structures. Directly fed main patch is formed on dielectric substrate with thickness d_1=0,81mm while top parasitic patch is formed on the bottom side of substrate with thickness d_2=0,508mm. RO4003 (ε_r=3,38) is used as dielectric material of the substrates. Additional air gap d_a separates top and bottom substrates providing proper element coupling. Dielectric spacers positioned in the four corners of the antenna element ensure this air gap.

Table 1 contains main construction parameters of two antennas. Designations are as follows:

PL_i, PW_i	Length and width of the patches,
ε_{ri}, d_i, $tg\delta_i$	Relative dielectric permittivity, thickness and loss tangent of the substrate materials,
DL,DW	Length and width of the two microstrip dipoles,
S	Spacing between bottom patch and parasitic dipole.

TABLE I
CONSTRUCTION PARAMETERS OF INVESTIGATED ANTENNAS

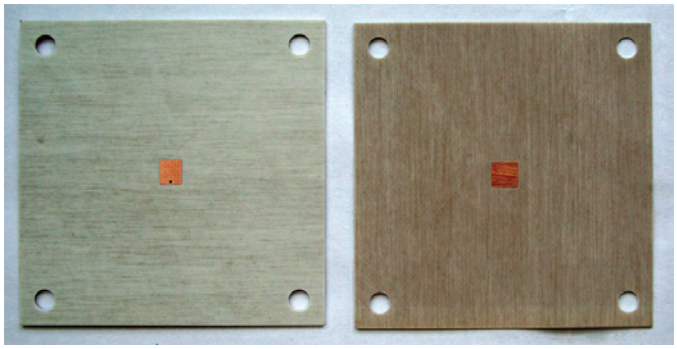
Designation	Value	
	Antenna 1	Antenna 2
PL_1	5,65 mm	5,65 mm
PW_1	5,15 mm	4,55 mm
PL_2	6 mm	5,8 mm
PW_2	6,5 mm	6,1 mm
DL	–	5,15 mm
DW	–	1,275 mm
S	–	0,2 mm
d_1	0,81 mm	0,81 mm
d_2	0,508 mm	0,508 mm
d_a	1,7 mm	2,1 mm

In typical stacked microstrip antenna (Fig.1a) coaxial line feeds bottom patch and top patch is excited through electromagnetic coupling with bottom element. The two elements have slightly different geometrical sizes and therefore resonate at adjacent frequencies. Therefore in VSWR or return loss characteristics only two clear resonances are observed. They overlap in a way that forms wide impedance bandwidth.

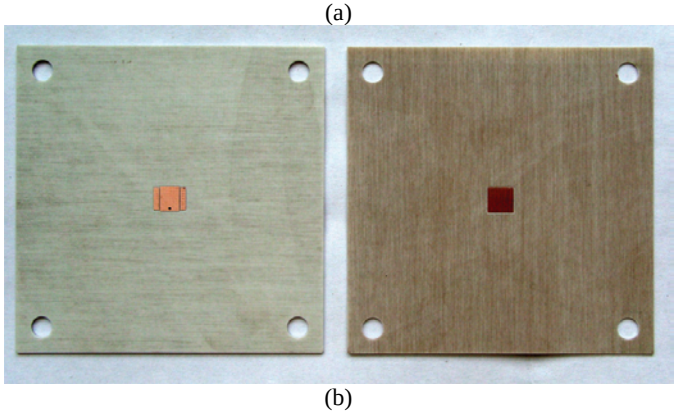
Proposed antenna (Fig.1b) acts in similar way as typical structure. Main bottom patch and parasitic top patch are coupled and form two resonances in VSWR. The presence of two more parasitic elements - microstrip dipoles - leads to excitation of third (high frequency) resonance. This modification allows additional improvement of antenna bandwidth.

III. ANTENNA ANALYSIS

Initial design and tuning of the investigated antennas is done through the use of software based on Finite Element Method – Ansoft HFSS. This is precise electromagnetic simulator widely used for modeling of different microwave structures. Experimental verification of simulation results is conducted afterwards in order to evaluate real antenna behavior. Photographs of top and bottom radiating elements etched on dielectric substrates for Antenna 1 and Antenna 2 are shown in Fig.2



(a)



(b)

Fig. 2. Test samples of investigated antennas:
(a) Antenna 1; (b) Antenna 2

Experimental measurements of two antennas are conducted in laboratory conditions. S-parameters are measured with network analyzer N5230A from Agilent Technologies. Fig.3 shows S-parameters measurement setup – assembled antenna connected to the network analyzer.

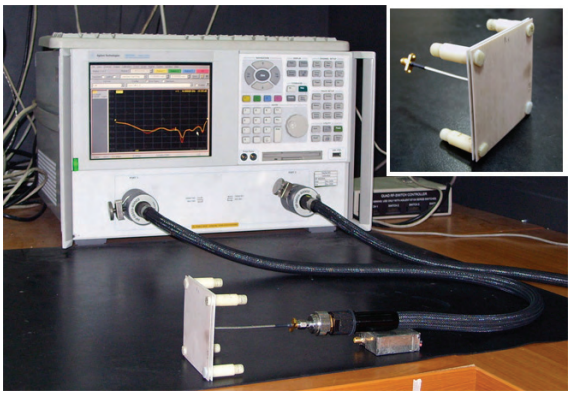


Fig. 3. S-parameters measurement setup

Results for return loss S_{11} measured for Antenna 1 and Antenna 2 are shown in Fig.4a and 4b respectively. Two types of results are presented in one coordinate system – HFSS simulation (thin line) and measurement (thick line). This allows direct evaluation for the accuracy of software simulations. Very good match between the two types of data is observed concerning frequency bandwidths and return loss values.

Fig.4 clearly shows the advantage of proposed antenna (Antenna 2) compared to typical stacked microstrip element (Antenna 1). In ideal situation (HFSS results) Antenna 1 has only two resonances in return loss: f_1=12,9GHz and

f_2=15,6GHz. Antenna 2 exhibits three clear resonances: f_1=13,1GHz, f_2=15,45GHz and f_3=16,4GHz. Therefore bandwidth improvement in Antenna 2 is caused by the excitation of the third high frequency resonance (f_3=16,4GHz). As mentioned above this resonance is due to the presence of additional parasitic dipoles adjacent to the main bottom patch.

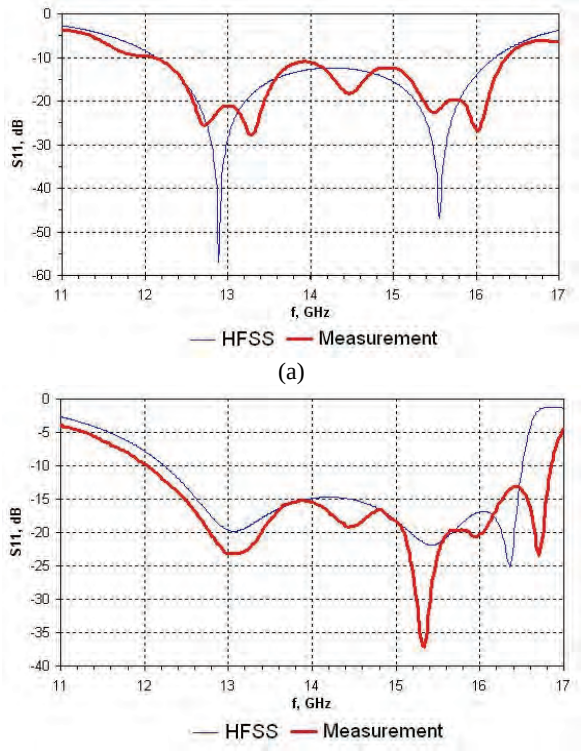


Fig. 4. Simulated and measured return loss:
(a) Antenna 1; (b) Antenna 2

TABLE II
IMPEDANCE BANDWIDTH OF INVESTIGATED ANTENNAS

Variant			Antenna 1	Antenna 2
BW	$f_L \div f_H$	HFSS	12,17÷16,25	12,3÷16,51
		Measurement	12,17÷16,4	12,05÷16,85
	%	HFSS	28,7	29,2
		Measurement	29,6	33,2

Table 2 contains simulated and measured impedance bandwidths of Antenna 1 and Antenna 2. There is slight difference between simulated and measured value of bandwidth for Antenna 2. It is probably caused by very small geometrical dimensions of the test sample and the production tolerances. This is the reason for the shift in frequency of third resonance – measured resonance is 450 MHz higher than simulated. Antenna 2 has another advantage compared to Antenna 1 except wider bandwidth. Impedance matching for the middle region of the working band is much better for Antenna 2. For frequencies around f=14 GHz Antenna 2 has return loss $S_{11} \leq -15,4$dB, while for Antenna 1 it is fulfilled $S_{11} \leq -11$dB. This means that if we assume $S_{11} \leq -10$dB as limiting condition Antenna 2 has more potential for bandwidth enhancement than demonstrated here.

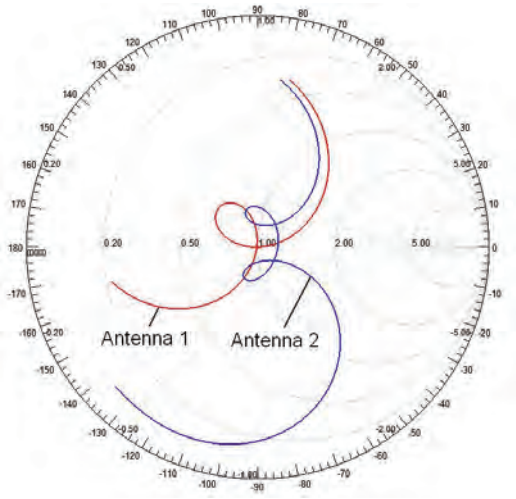


Fig. 5. Input impedance of Antenna 1 and Antenna 2

Fig. 5 shows input impedance of Antenna 1 and Antenna 2 plotted on Smith chart. As seen Antenna 1 has one loop near the center of the chart. It corresponds to the two resonances in the return loss characteristic. This impedance behavior is typical for two resonance system such as stacked microstrip element. On the other hand Antenna 2 has two loops in Smith chart. Such shape of input impedance corresponds to three resonances in frequency characteristic of the radiating element. Antenna 2 has three resonant elements in its construction. Main bottom patch and top parasitic patch determine low frequency loop in input impedance. Microstrip parasitic dipoles on bottom substrate excite third resonance and therefore determine higher frequency loop in Smith chart.

Radiation properties of investigated structures are evaluated through measurement of antenna gain in anechoic chamber. Fig.6 is a photograph of measured antenna mounted on a positioner in anechoic chamber. Antenna under test is fitted in specially designed measurement box (black square object in photograph), which is covered with absorber for elimination of unwanted reflections.

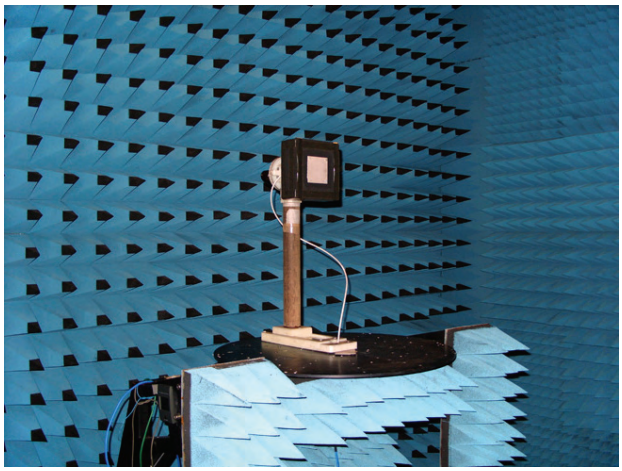


Fig. 6. Measurement in anechoic chamber

Measured gain for Antenna 1 and Antenna 2 is presented in Fig.7. Measured and approximating curves for the gain are plotted on the graph in order to precisely evaluate antenna operation. Results give very close values of gain for the two antennas. This is normal, because constructions of Antenna 1 and Antenna 2 are very similar. It is interesting to note that Antenna 1 has slightly higher maximum gain (G_{max}=8,3dB) but it drops much quicker at higher frequencies. Antenna 2 has lower gain (G_{max}=8,1dB) but it is more stable with frequency, which determines wider gain bandwidth of this radiator. Gain bandwidth is defined as the frequency region at the edges of which gain drops with 1dB from its maximum value. Table 3 presents gain bandwidths for Antenna 1 and Antenna 2.

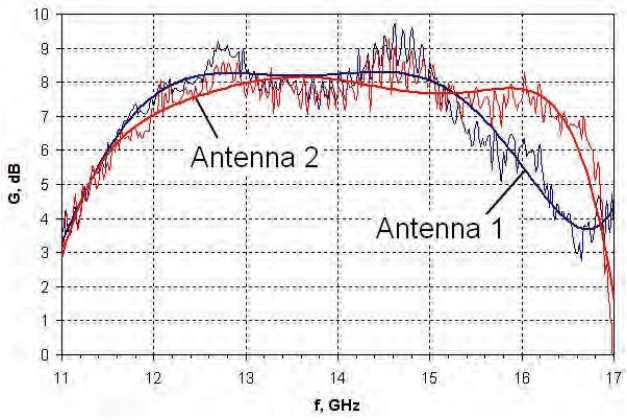


Fig. 7. Gain vs. frequency for Antenna 1 and Antenna 2

TABLE III
GAIN BANDWIDTH OF INVESTIGATED ANTENNAS

Variant		Antenna 1	Antenna 2
BW	$f_L \div f_H$	11,8÷15,5	12,1÷16,4
	%	27,3	30,2

IV. CONCLUSION

This paper presents new variant of stacked microstrip antenna, which exhibits improved impedance bandwidth compared to typical antenna of this type. The idea to combine vertically and horizontally positioned parasitic elements presented in [6] is further developed here. As a result alternative three resonance antenna is designed. It differs from typical stacked antenna by the addition of two parasitic microstrip dipoles positioned on both sides of main bottom patch.

Numerical and experimental investigations of proposed and typical antenna are conducted. As a result impedance bandwidth of proposed antenna is *BW*=33,2% compared to *BW*=29,6% for typical construction. This is not very big difference but it must be noted that Antenna 2 (proposed) has much better input match ($S_{11} \leq -15$dB) than the typical element - Antenna 1 ($S_{11} \leq -11$dB). The new radiating element has also wider gain bandwidth – 30,2% compared to 27,3% for typical antenna.

It should be noted that proposed antenna preserves all advantages of typical stacked patch – it has symmetrical construction which determines symmetrical and stable radiation pattern in working band; small physical area of the antenna (there is just slight enlargement of the size in proposed element); small size makes the radiator suitable for array applications; etc.

Presented three resonance stacked element radiates one sense of linear polarization. Modern communications systems often demand operation with different types of polarization – RHCP and LHCP, two orthogonal linear polarizations. Therefore one potential goal for future work of the author is development of similar antenna, which can radiate different types of polarization.

REFERENCES

[1] R. Garg, P. Bhartia, I. Bahl, and A. Ittipiboon, *Microstrip Antenna Design Handbook*, Norwood, Artech House, 2001

[2] R. Garg, and V. S. Reddy, "A Broad-Band Coupled-Strips Microstrip Antenna", *IEEE Trans. Antennas Propagat.*, vol. 49, pp. 1344–1345, 2001.

[3] G. Kumar and K. C. Gupta, "Broad-Band Microstrip Antennas Using Additional Resonators Gap-Coupled to the Radiating Edges," *IEEE Trans. Antennas Propagat.*, vol. 32, pp. 1375-1379, 1984.

[4] F. Croq and D. M. Pozar, "Millimeter Wave Design of Wide-Band Aperture-Coupled Stacked Microstrip Antennas," *IEEE Trans. Antennas Propagat.*, vol. 39, pp. 1770–1776, 1991.

[5] S. D. Targonski, R. B. Waterhouse, and D. M. Pozar, "Design of Wide-Band Aperture-Stacked Patch Microstrip Antennas," *IEEE Trans. Antennas Propagat.*, vol. 46, pp. 1245–1251, 1998.

[6] S. R. Baev, and N. I. Dodov, "Bandwidth improvement of an aperture-coupled stacked microstrip antenna", *Proc. ICEST 2006*, June 29–July 1, Sofia, Bulgaria, pp. 21–24, 2006.

[7] S. R. Baev, and N. I. Dodov, "Investigation of a stacked microstrip antenna with wide impedance bandwidth", *Telekom 2006*, October 12–13, Varna, Bulgaria, pp. 294–299, 2006.

Minimizing HUM Modulation Products in Coaxial Part of HFC Networks

Oleg Borisov Panagiev[1]

Abstract **- Variants of combined AC/DC remote supply of cable amplifiers used in coaxial parts of HFC networks are offered in the report. The goal is to minimize the number and levels of products from HUM modulation that occurred in passive as well as in active devices of cable distribution network with AC supply of trunk and subtrunk amplifiers. Different methods and circuit solutions of cable amplifiers models, produced by European and world manufacturers are described. Corresponding conclusions are made on the base of practical realization of offered circuit decisions.**

Keywords **– HFC/CATV, cable amplifier, fiber node, HUM, switching-mode power supply.**

I. INTRODUCTION

In the coaxial part of each hybrid fiber coaxial network, as also in the cable distribution network (CDN) of the existing CATV systems, the transfer of the signals to the subscribers is made by trunk and subtrunk. Because of the damping of the signals in the passive devices (coaxial cables, taps, splitters, attenuators and others), the maintenance (the support) of their normal levels in CDN is accomplished by cable amplifiers (Fig.1). The supply of the contemporary cable amplifiers is an alternating current (AC from 30 V to 65 V) and the supply unit (SU) is a switching-mode power supply (SMPS). The supply is submitted distantly from a supplying source, placed in the Head end and sometimes in some apartment building that is close to de trunk or the subtrunk. The division and the summation of supply current and group signal is made by diplexers, which are made of chokes (L_1, L_2, L_3, L_4), capacitors (C_3, C_4) and resistors (R_1,

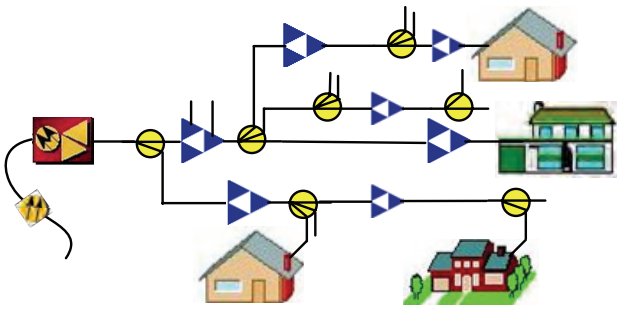


Fig.1. Cable distribution network on HFC/CATV system

R_2, R_3…R_8), (Fig.2). Also, trough the diplexers the supply current is being transferred from the input to the output (the outputs of the amplifier) and backwards. That is accompanied by an occurrence of HUM modulation in the active and in the passive devices [1], [2], [3], [4], [5]. By the cable amplifiers with DC supply (+30 V to +65 V) and ordinary power supply unit, the levels of the HUM modulation products are insignificant (suppressed to more than 90 dB).

In the present paper are offered circuit solutions for minimizing of the HUM modulation products, using:

a) Replacing of L_1, L_2, L_3 and L_4 with power passing chokes F26 of the MMG-North America [6]. By doing that it is accomplished not only the minimize of the effects of HUM modulation from 1MHz to 1GHz but also an increased bandwidth, improved reliability, expanded services, and technology upgrades. The result is chokes with proprietary windings that negate the need for dampening resistors and can withstand 15 A continuous current and a peak of 25 A for two hours without loss of performance.

b) Supplying of the cable amplifiers [7], [8] and the fiber nodes [9], [10] with AC, as well as with a DC. This way the HUM modulation products are minimized and is facilitated the upgrade of the old cable distribution networks, in which is used DC supply. A short interruption of the signal to the subscribers and a minimal temporary deterioration of the image quality are fulfillable without parallel trunks or subtrunks. However, it is necessary the respective changes in the SMPS units to be made, so that they can be supplied also with DC. On a later point of the process, when all of the amplifiers are replaced and the output level of every amplifier is optimized in the whole cable distribution network, it is possible for the supplying source for the cable amplifiers in the Head end to be replaced from DC to AC, without needing of any readjustment of the amplifiers. To be minimized the HUM modulation is recommended continuing of the usage of DC supply for the amplifiers.

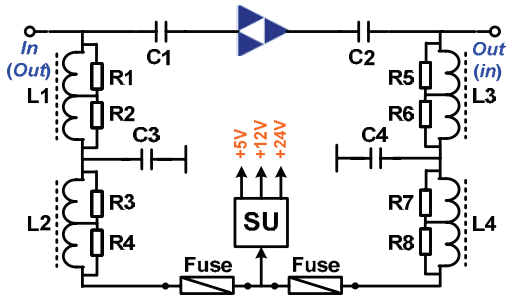


Fig.2. Electrical circuit on power supply in cable amplifier

[1]Oleg B. Panagiev is with the Technical University of Sofia, Bulgaria, E-mail: olcomol@yahoo.com

II. Influence of the HUM modulation on the transmited signals

Essentially HUM modulation represents undesired modulation of the television visual carrier by the fundamental or low-order harmonics of the power supply frequency, or other low-frequency disturbances. HUM can comes from a defective AC/DC power supply of an amplifier and/or fiber node, to amperage draw thru a passive equipment or water in connectors. Also HUM modulation occurs due to imperfect power supply filtering and as a result of parametrical modulation of magnetic components in both amplifiers and passive devices. Modulation due to power supply ripple should logically be more or less independent of RF frequency. To the extent that the parametric modulation of magnetic components is a factor, however, its effects may vary with frequency. The primary mechanism is that magnetic cores may partially saturate at peaks of the AC current. This will affect there impedance at RF, and so create a change in return and transmission loss that will vary at a 50 Hz rate. Since the impedance change may vary with frequency, as may the effect of the impedance change, there is no way of predicting how the modulation percentage will vary across the spectrum. At a given frequency, however, the percentage modulation of a video signal should be the same as that of a digitally modulated signal though the latter may be able to tolerate a higher modulation percentage.

HUM is calculated in % or in dB:

a) converting the measured logarithmic value (dB) in %

$$HUM[\%] = 100/(10^{HUM[\mathrm{dB}]/20}); \quad (1)$$

b) converting the percentage in logarithmic value (dB)

$$HUM[\mathrm{dB}] = 20\lg(100/HUM[\%]). \quad (2)$$

The influence of the HUM modulation on the parameters of the carried signals in cable distribution network is unfavorable for the analog signals as well as for the digital. In the 50 Hz video oscillogram (Fig.3) this can be seen as a wave-form line on the video signal black shoulder or on the 50 Hz synchronization pulse (picture repetition impulse). A

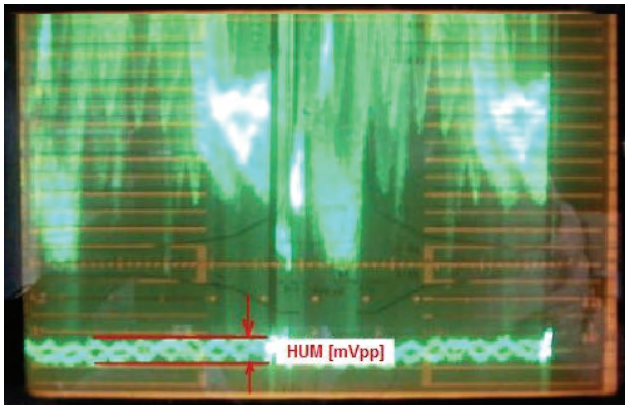


Fig.3. Hum on an analogue signal

HUM modulation on a QAM signal caused by a faulty power supply in the distribution line or in the signal processing will be seen as a low MER value. Typical disturbance in the constellation diagram with HUM transmission is presented on Fig.4. Remember that: At least 32 dB is needed for 64-QAM and 36 dB for 256-QAM at the subscriber.

III. Circuit solutions

The accomplishment of combined remote supply is made in two ways:

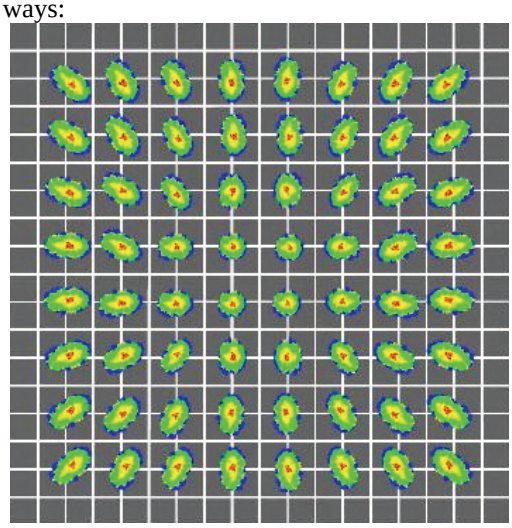


Fig.4. Typical disturbance in the constellation diagram with HUM transmission

- by building-in of new (additional) module, which carries out the initial starting of the SMPS unit;
- by making a change in the SMPS unit circuit to ensure possibility for working with DC (removal and joining of elements and electrical connections).

On Fig. 5a is shown diagrams of increase of start supply voltage $U_{cc,start}$ (with dash line) and enable voltage U_e (continuous line) for integrated circuit (IC) of SMPS unit when original (alternating) supply voltage is used.

Fig. 5b shows increase of mentioned above voltages bun when use DC supply voltage. Both voltages increase linearly and almost simultaneously because time constant is very small. That is why it is not possible to start in operation the SMPS unit. Moreover $U_{cc,start}$ must grow more slowly than U_e and time Δt_2 depends on values of R_1 and C_2, (Fig.6).

The idea, which is suggested and realized below, is to increase the time for building-up the initial voltage in exponential law and to equalize time constant with time constant obtained when alternative current supply is used.

Experiments were made by using of cable amplifiers of two different companies – C-COR/ADC PHASOR and WISI.

A. Building-in of new module

Elements of schematic diagram (Fig.7), that forms start impulse, is mounted on separate printed circuit board (PCB). Connection with SMPS unit is made using short wires without interruption of existed links, i.e. mounting is parallel in shown points on the figure. The required time constant is achieved selecting the values of elements C_1, R_1 and R_2.

Formed impulse is inverted and supplied by transistor T_1 of respective pin of integrated circuit, on which is received start initial supply voltage. The minimal input voltage is +30 V.

B. Change of circuit of SMPS unit

When the start impulse is formed according to diagram shown on Fig. 8 it is necessary to break some connection on PCB of SMPS unit and to create a new ones. Moreover, to obtain the time constant, C2 is replaced with a new capacitor C2* with the same operating voltage as the original one but with more capacity (C2* > C2).

In series of resistors $R_1||R_2$ is connected diode D (1N4002) to provide reliable positive start impulse on pin 6 of integrated circuit (TDA4605-3), altering on exponential law determined by R_1, R_2 and C_2^*. At the same time diode D cause discharge of capacitor C_2^* trough pin 6 of integrated circuit and isolation between main and secondary supply obtained from pulse transformer (PT). The minimal input voltage is +35 V.

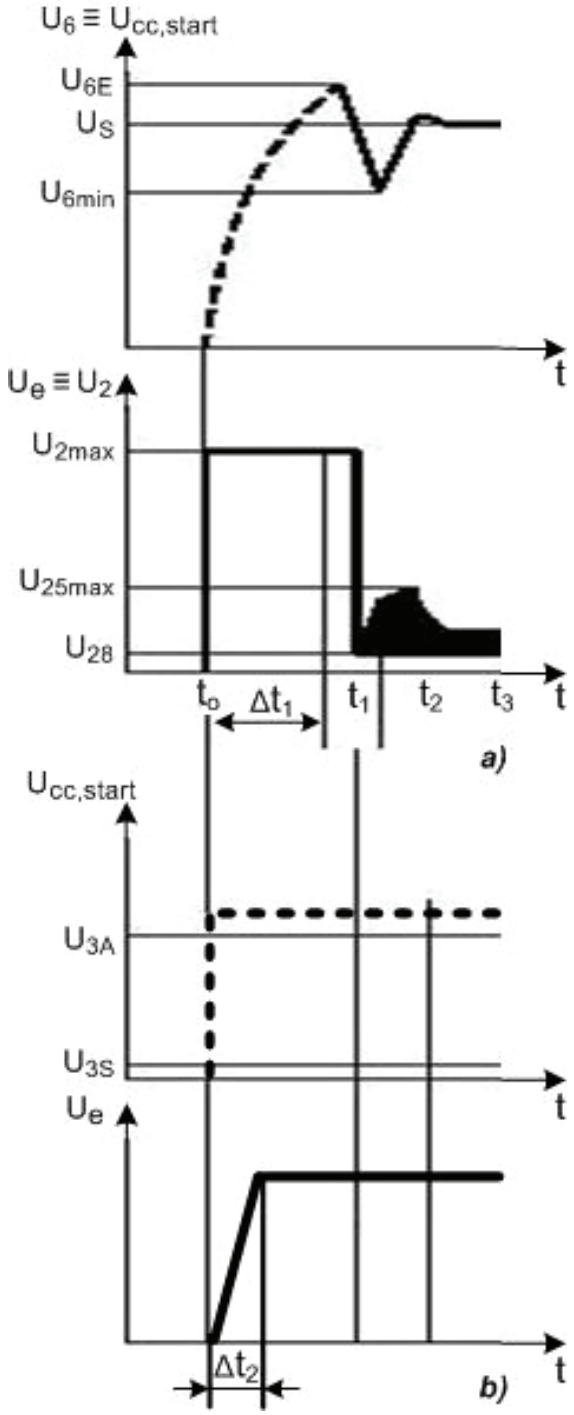


Fig.5. Diagrams of starting and enable voltages

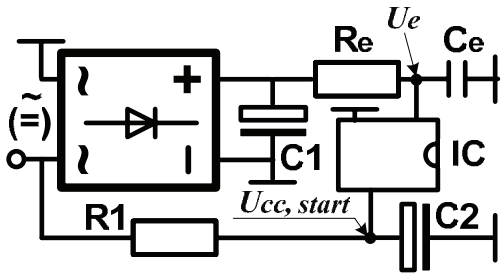


Fig.6. SMPS unit with original (alternating) supply voltage

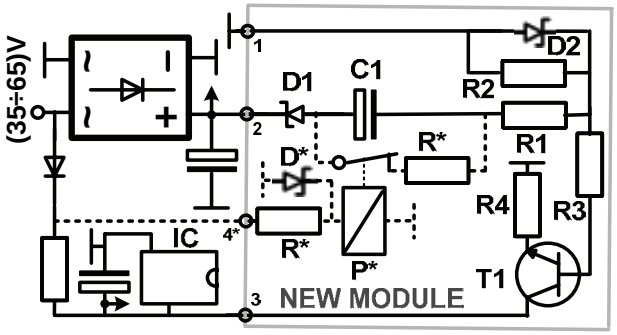


Fig.7. SMPS unit with new module

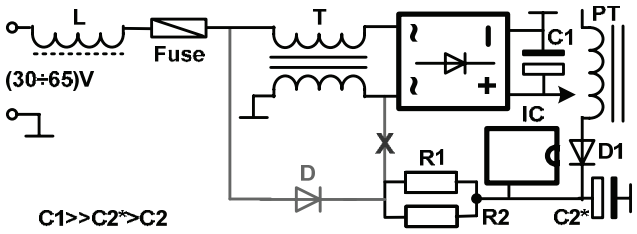


Fig.8. Change of circuit of SMPS unit

IV. CONCLUSION

The circuit solutions, discussed in the report, allow cable amplifiers and fiber nodes to work with AC supply (50/60 Hz) as well as with direct current one. Choice of supply depends on cable operator.

Experiments for circuit shown on Fig.7 did not show problems in switching on, switching off and normal operation of SMPS unit. For circuit shown on Fig.8 it is established that if supply because of any reason stops for a short time it is necessary to wait 2 or 3 minutes to discharge the capacitor C_1

and to start SMPS unit again. This disadvantage could be removed by adding elements (on figure they are shown by dash line) that when input voltage is switched off makes short connection of capacitor pins C_1 and discharge it.

On Fig.9 is shown graphs of HUM modulation by three cases:

1) MMG F26 core without resistors:
2) Competitor's core with resistors;
3) SMPS unit with DC supply.

HUM modulation levels are measured according to European standard CENELEC EN 50083-7. For cases 1) and 2) graphs indicate a comparison of materials of similar geometries, using the same test equipment and set up.

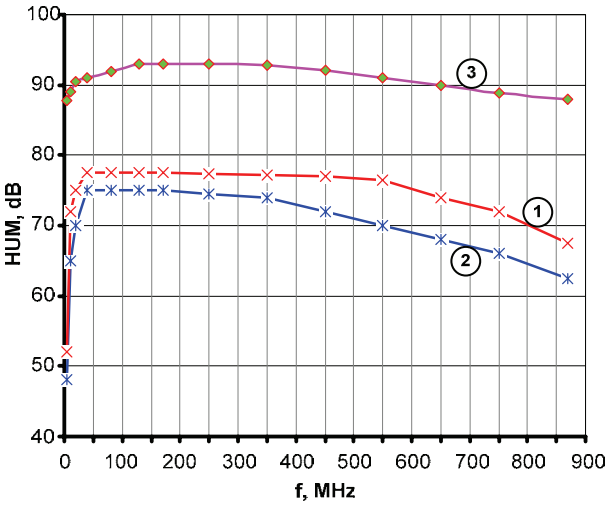


Fig.9. Graphs of HUM modulation

REFERENCES

[1] W. Ciciora, J. Farmer. D. Large and M. Adams, *Modern cable television technology*.2nd edition, Elsevier Inc., 2004.
[2] PCT international. Amplifiers, 2005.
[3] C-COR Inc. Catalog: RF line passives and Amplifiers, 2007.
[4] SDA 15A Amplifier Operations manual. ACI Communications Inc. Washington, 2006.
[5] R. Hranac.,"Digital signal measurements",Cisco systems, 2003.
[6] Hum modulation suppression components for broadband applications: Power Pass. Chokes. MMG-North America, 2003.
[7] www.wisi.de
[8] www.triax.com
[9] BTN-2: Telecommunications Optical Node. Motorola/ Starline®, www.motorola.com.
[10] K. Maki. AC8000 FIBRE OPTIC PLATFORM, Product specification, Teleste, 2008.

Reduction of Nonlinear Products from Composite Distortions of the Signals in Optical Link

Oleg Borisov Panagiev[1] and Galya Veselinova Raicheva[2]

Abstract **- Nowadays application of hybrid fiber coaxial (HFC) networks in cable communication systems is a key technology for two-way transmission of information between users. In spite of that, common transfer of AM-VSB and M-QAM signals is attended with unfavorable effects that analogue channels have on digital ones: arise of composite distortions of second and third order, increase of BER, decrease of noiseimmunity etc. Methods for supporting minimal levels of distortions in HFC networks conformed to network topology and parameters of optic transmitter, receivers and amplifiers are offered.**

Keywords **– HFC, composite distortions, HUB, BER, M-QAM**

I. INTRODUCTION

With the increasing demands of interactive services, existing coaxial TV systems with its inherent problems and limitations can not be used for large and very large distances (> 10 km), [1], [2].

Solution of the problems and elimination of limitations are possible with building up of hybrid fiber coaxial (HFC) networks, which application nowadays in cable communication system is a key technology for two-way transmission of information between users. Better quality of video services could be achieved by applying of subcarrier multiplexing (SCM) in the HFC systems that use M-ary quadrature modulation (M-QAM) for transmission of compressed digital video signal (MPEG) together with ordinary amplitude modulation of analogue signals (AM-VSB). Frequency diagram of one HFC network is shown on Fig.1. Using MPEG compression and QAM modulation is possible to transfer some digital video programs or some narrow-band channels with data, telephone signals etc. in one channel with frequency band 8 MHz instead of one analogue TV program with AM/VSB modulation. Internet access request from the subscriber is transmitted via Upstream services, and the receipt of requested information from the WEB by in the Downstream services.

In really operation of HFC networks with common transmission of AM/VSB and M-QAM signals in determined conditions could be observed unfavorable effect from analogue on digital channels [3], [4], [5]. As a result nonlinear products are obtained (composite distortions – CSO, CTB and CXM). The most important factors for their origin are: laser "clipping", Rayleigh backscattering and reflection noises.

In general case complex distortion are regarded as a noise, because their adverse effect is accompanied by increase of binary error rate BER.

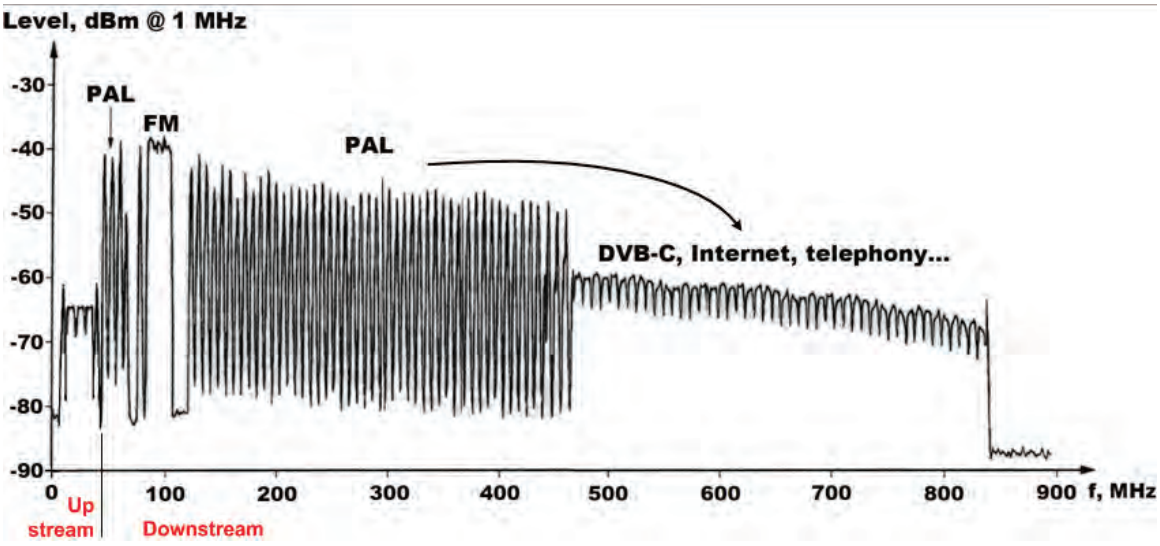


Fig.1. Frequency diagram of HFC network

II. DETERMINATION OF OPTICAL POWER IN HUB/FN

In HFC networks are established a lot of distortion factors, lowering the quality of digital channels, some of them are related to network topology [6] and others – with way of

[1]Oleg B. Panagiev is with the Technical University of Sofia, Bulgaria, E-mail: olcomol@yahoo.com

[2]Galya V. Raicheva is with the Technical University of Sofia, Bulgaria, E-mail: raicheva@tu-sofia.bg

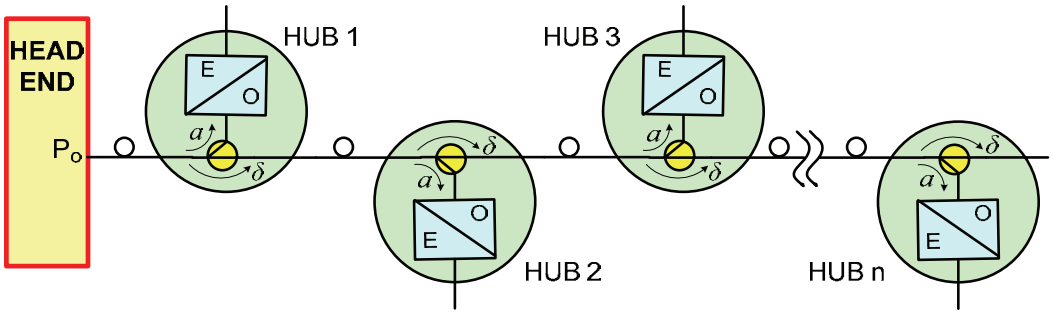


Fig.2. Bus topology

mixing of AM-VSB and M-QAM signals [4] and parameters of optical devices [7], [8] – optical receivers, optical transmitters and optical amplifiers.

Topology of a HFC network usually is tree type (branched), but for large networks it is non-effective and expensive, that is why recently star topology is used, BUS - is already used in local networks (LAN) and ring topology.

In many cases, the three types topology are used together (for large cable TV systems) in order to combine their advantages in particular application.

A. Bus topology (Fig.2)

In Bus topology one more parameter appears in expression for determination of optical power in particular HUB, this is the loss of power a (tap-off attenuation in dB) by derivation of light signal:

$$P_{n,b}[dBm] = P_o - a - (n-1).\delta\,, \qquad (1)$$

where $n \le 60$. On Fig.3 is shown diagram of $P_{n,b} = func\,(P_o, n)$, where P_o= (4÷16) dBm, n = 1÷60, δ =1,8 dB and a =5,6 dB, (WISI LK06).

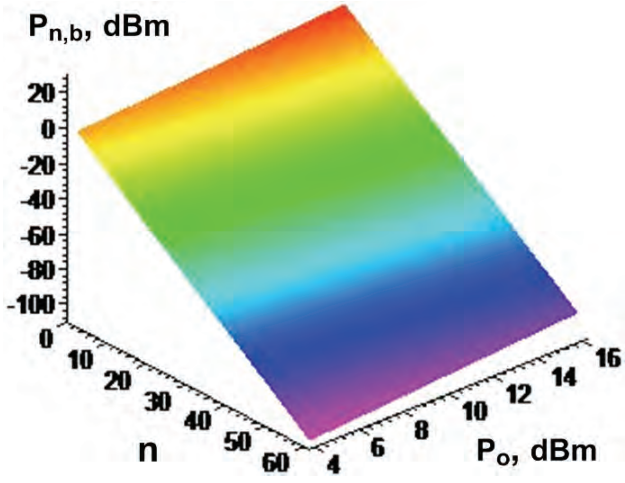


Fig.3. Diagram of $P_{n,b} = func\,(P_o, n)$

B. Star topology (Fig.4)

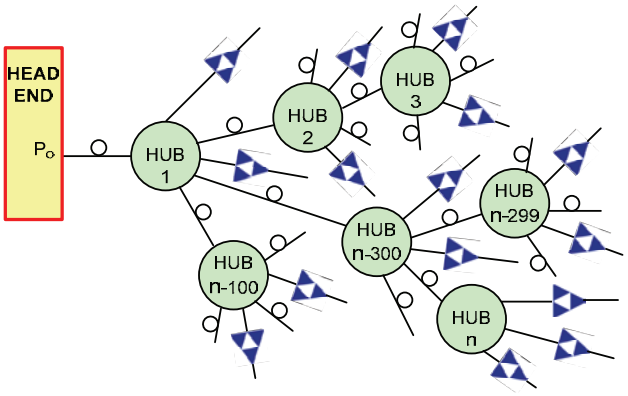


Fig.4. Star topology

For star topology, optical power in any HUB depending on output power of optical transmitter could be determined by:

$$P_{n,s}[dBm] = P_o - \delta - 10.[1 + 3{,}322.\lg(1 - 10^{-\frac{\delta}{10}})].\lg n\,, \qquad (2)$$

where

P_o is output power of optical transmitter in dBm;

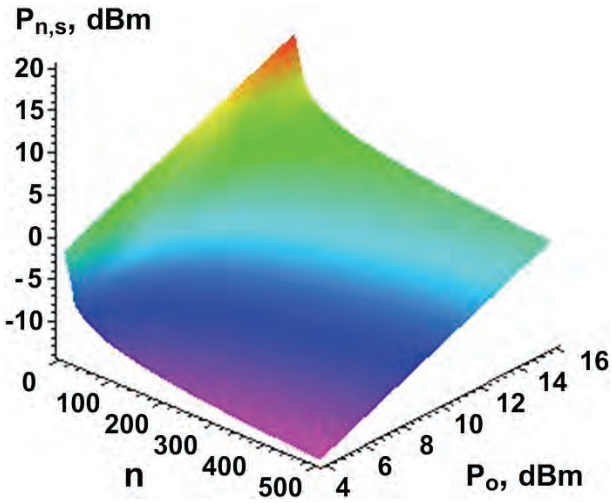


Fig.5. Diagram of $P_{n,s} = func\,(P_o, n)$,

n – HUB number, $n \le 500$;

$P_{n,s}$ – power available at n-th fiber node (FN/HUB) in dBm;

δ - HUB losses (insertion losses, throughpass attenuation) in dB.

On Fig.5 is shown diagram of $P_{n,s} = func\,(P_o, n)$, where Po= (4÷16) dBm, n = 1÷500 and δ = 5,7 dB, (WISI LK13).

III. Reduction of Composite Distortions in Optical Link

Every main trunk in HFC systems consists of fibers and branching – of coaxial cables. In optical part except the thermal noise and non-linear distortions there are created impulse noises from laser "clipping". [3], [4], [5]. The impulse noise inherent to multiplexed signal by frequency division causes a drop of output power of laser diode near zero and its input control current is lower than current setting its work point. [3], [5]. As a result, nonlinear products appear that cause increase BER. It is necessary to analyze carefully this product in order to determine requirement not only to the system, but also to its components: optical transmitters/receivers, bridge and linear amplifiers.

For determination of composite distortions it is necessary to use simplified expressions to obtain values for CTB, CSO and CXM.

Maintaining minimum levels of complex distortions in HFC networks is an essential task that must be solved depending on the network topology and parameters of active devices into it. It should be noted that in optical systems of different manufacturers exploitation parameters are with similar values (Table 1) as regards to laser diodes used in transmitter as well as photodiode used in the receiver. Significant impact on output parameters have the index of optical modulation m. Increase of this index causes deterioration of CTB, CSO and CXM (they are ratio carrier/interference according to CENELEC EN 50083-6) and improvement of ratio carrier /noise (CNR), (Fig.6).

The main reasons leading to the need to recalculate the composite distortions of the system are:

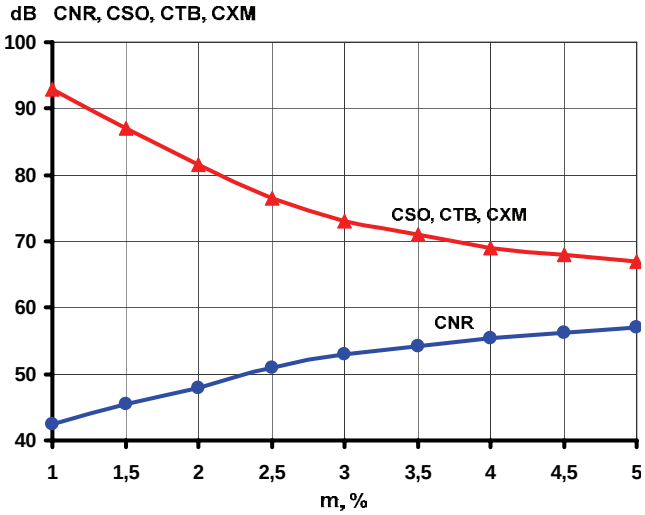


Fig.6. Influence of m on CSO, CTB, CXM and CNR

TABLE 1

	Optical Tx			Optical Rx		
Type	LT61S	OTXS06	CHP GQTX	LR 82	ORB901	CHP GFRX
Optical (input) output level, dBm	6	8	10	(-8...0) 0	(-8...+3) 0	(-3…+3)
RF – (input) output level, dBμV	(79…88)	(85)	(97)	(77…95) 90	(85) 104	101
CNR,dB	>50	>50	51	50	52	62
CSO,dB	>60	>60	55	≥ 60	35*	>65
CTB,dB	>60	>65	60	≥ 60	35*	>85
CXM,dB	60	60	63	60	55*	80
Producer	WISI	Triax	C-COR	WISI	Triax	C-COR

A). Reducing the level of modulating signal P_{in} (respectively of output optical power P_o) for one (any/unspecified) channel.

Expressions for determination of complex distortions are:

$$CSO_{new}\,[dB] = CSO + \Delta CSO, \quad (3)$$

$$CTB_{new}\,[dB] = CTB + \Delta CTB, \quad (4)$$

$$CXM_{new}\,[dB] = CXM + \Delta CXM, \text{ where} \quad (5)$$

$$\Delta CSO\,[dB] = -\Delta Po, \quad (6)$$

$$\Delta CTB\,[dB] = -2.\Delta Po, \quad (7)$$

$$\Delta CXM\,[dB] = -2.\Delta Po, \text{ a} \quad (8)$$

$$\Delta P_o\,[dB] = (P_o - P_{o,nom}), \text{ for a channel.} \quad (9)$$

***B). Change of optical modulation index* (OMI, *m*).**

$$m = \Delta P_o\,[mW] / P_o\,[mW] \text{ или} \quad (10a)$$

$$m = 10^{\frac{\Delta P_o\,[dBm] - P_o\,[dBm]}{10}} \quad (10b)$$

In logarithmic units

$$m\,[dB] = \Delta P_o\,[dBm] - P_o\,[dBm] = 10 \lg m. \quad (10c)$$

Change of optical modulation index in dB is

$$\Delta m\,[dB] = 20lg(m_1 / m_2), \quad (11)$$

m_1 and m_2 in practice is difficult to be measured that is why according to [9]:

$$\Delta m\,[dB] = \Delta P_o\,. \quad \text{Then} \quad (12)$$

$$m_{new}\,[dB] = m + \Delta m. \quad (13)$$

Influence of m on CSO, CTB, CXM and CNR is shown on Fig.6.

C*). *Reduction of the level of output optical power P_o of the transmitter depending on number of transfer channels N.

$$\Delta P_{in}\ [dB] = 10 \lg N, \tag{14}$$

$$P_{in,new}\ [dBm] = P_{in} - \Delta P_{in}, \tag{15}$$

$$P_{o,new}\ [dBm] = G_p + P_{in,new}, \tag{16}$$

where G_p is gain of power amplification.

D*). *Cascade connection of optical systems

Total complex distortions in conformity with these of n-th system and Table 2 are:

$$CSO_{\Sigma}\ [dB] = CSO_n + \Delta' CSO, \tag{17}$$

$$CTB_{\Sigma}\ [dB] = CTB_n + \Delta' CTB, \tag{18}$$

$$CXM_{\Sigma}\ [dB] = CXM_n + \Delta' CXM, \tag{19}$$

TABLE 2

Δ' , dB	Reduction, dB		
	Δ'CSO	Δ'CTB	Δ'CXM
0	3,61	4,52	6,0
1	3,14	4,03	5,53
2	2,71	3,59	5,08
3	2,33	3,19	4,65
4	1,99	2,82	4,25
5	1,69	2,48	3,88
6	1,43	2,18	3,53
7	1,21	1,91	3,21
8	1,02	1,67	2,91
9	0,85	1,46	2,64
10	0,71	1,27	2,39
11	0,60	1,10	2,16
12	0,50	0,96	1,95
13	0,41	0,83	1,75
14	0,34	0,72	1,58
15	0,29	0,62	1,42
16	0,24	0,54	1,28
17	0,20	0,46	1,15
18	0,16	0,40	1,03
19	0,13	0,34	0,92
20	0,11	0,30	0,83

where $\Delta'\ [dB] = Cxx_{n-1} - Cxx_n$, and *xx* takes the symbol: *SO*, *TB* or *XM*.

IV. CONCLUSION

When the parameters of optical devices in both topologies are equal, the optical power could be determined in respective HUB and could be made the following conclusions:

- in star topology it is possible to connect at last five times more HUB than Bus topology;
- in star topology when the number of HUB is increased, for example twice (from 16 to 32), the optical power in the last HUB is twice less ($P_{32} \leq P_{16}/2$);
- in Bus topology the twice increase of HUB leads to fivefold decrease of optical power in the last HUB ($P_{32} \leq P_{16}/5$).

REFERENCES

[1] C. Cox III, "*Analog optical links theory and practice*," Cambridge University press, 2004.

[2] K. Angelov, K. Koitchev, S. Sadinov. "An investigation of noise influences in optical transmitters and receivers in cable TV networks", ICEST, Proc. of Papers, Sofia, pp. 102-105, 29 June – 01 July 2006.

[3] R. Howald. "HFC-Reloaded: The service capabilities being unleashed by hybrid fiber coax make a further examination mandatory", Comm. systems design, pp. 7-10, June 2003.

[4] S. Bette, V. Moeyaert, M. Lhoir and etc. "Temporal Binary Error Distribution of QAM Transmission Impaired by Clipping-Induced Noise in CATV Lightwave Subcarrier Multiplex Systems", Proc. Symposium IEEE/LEOS Benelux, pp.167-170, 2004.

[5] O.B. Panagiev, "Some Capabilities of the Optical Technologies in the Cable Television Systems", IX NSASC "Electronics'2000, Sozopol 20-22 sept. 2000, pp. 163-170.

[6] S. Patela. "CATV and other fiber-optic networks", 2001.

[7] Triax: Main Catalogue 2009.

[8] http://www.wisi.de/

[9] Hirshmann. Vortrag Optische AM-CATV-Ubertragung. 1996.

Analysis of Wide Band Unknown Microwave Signals Detected by One Antenna

Nemanja M. Mitrović[1], Dragan D. Obradović[2] and Predrag S. Manojlović[3]

Abstract – **Acquisition of multi-channel signal where each channel consists a piece of arrived information could be complex operation, if it is considered wide band microwave signals, especially if one of the crucial information is level of arrived signal. First, at all, matching impedance between antenna and detector for a wide frequency band could not be same for all frequencies. Another possible troubleshooting point is transition between analog and digital domain and at end validation of arrived data. This paper should present, at particularly example, how some physical imperfections could be furbish by digital domain with software. Results of this analysis could be used in nuclear, radar, sonar, and medicine application, since all noted applications are dealing with unknown signals and demands hard real-time processing.**

Keywords – **Signal acquisition, Digital Signal Processing, Wide band receiver.**

I. INTRODUCTION

Today, many systems are basing on multi-channel [1], [2] acquisition. Some of them use multichannel acquisition for frequency hopping that allows low signal level detect ability, interference rejection, anti-eavesdrop protection, code division multiplexing and high-resolution ranging, some for one or more process parallelization, some for smart systems that are hoping from channel to channel with lowest noise or interference, and etc.

Let us simplify and consider some system that only acquires signal form the one antenna at the only one channel. That system could provide some data information. Note that for interpreting those information system need some "key" for rightly decoding but we consider unknown signals. There are more information that could be estimate from detected signal such as level and operating frequency. It should be note that frequency information in this system is discrete value and depends on channel wide. For more prescient frequency calculation, this channel should be narrow. At hardware level channel wide is represent with band pass filters that could not be narrow if area of interest is very high frequencies. Now, consider system that is equipped with acquired signal tracking device. In this case, there is one more information that could be detected and that is signal shape in time domain. This information is very powerful, since it can give some answers, like is it some data signal or some triggers or some pulse signal or something else. Furthermore, useful information will be parameters of that signal in both cases – parameters of base signal or parameters of modulation of data signal.

That was simplified version, since that was description of mono-channel system. Multi-channel has same characteristic with one little difference – arrived signal is disperse on several channels. In hardware domain that required a lot of memory and processing power. Now, put in fire wide band of high frequencies with many channels, for more precise frequency resolution, and only one antenna to obtain that band, made analog domain very complex and in real systems very harmful. Wide temperature stability, small size and low coast are the biggest enemy of such systems. Second two input fix degrade characteristic in system but first one not and required software correction.

Is it unique correction in system? Certainly not, since analog to digital converters has non-linear characteristic. On other side, matching impedance between any analog scopes is not same for all frequencies. At least, antenna could be specific problem. Antennas with a narrow beam of radiation should not been big problem, but antennas with wide beam of radiation need some software correction. Both have same problems with reflections. From this perspective, there are many corrections, what imposed main questions: is there sense to make them, could this system be robust and could it works at all? There are questions and let see some answers.

II. ANALYSIS OF BACKGROUND HARDWARE

For purpose of analysis, was build a hardware platform prototype that provides wide band, multi-channel microwave signals processing. Mainly, it consists of three different blocks: antenna, detector and digital signal processing unit.

The task is wide band signals observation and analysis. Horn antenna [3] that covers wide frequencies band, form 8GHz up to 12GHz is shown on Fig. 1.. Main characteristic of horn antenna is wide beam of radiation. Chosen horn antenna has maximal gain 8.5dB. Corresponding to 3dB beam wide [4], as a standard for antennas, beam limits of considered antenna are ±60º (antenna beam wide is 120º). Mostly, in microwave applications antennas dimensions dictates dimensions of whole devices (antennas are the biggest parts of

[1]Nemanja M. Mitrović is with the IMTEL Komunikacije, Bul. Mihajla Pupina 165b, 11070 Novi Beograd, Serbia, E-mail: nemanja@insimtel.com

[2]Dragan D. Obradović is with the IMTEL Komunikacije, Bul. Mihajla Pupina, 11070 Novi Beograd, Serbia, E-mail: obrad@insimtel.com

[3]Predrag S. Manajlović is with the IMTEL Komunikacije, Bul. Mihajla Pupina, 11070 Novi Beograd, Serbia, E-mail: pedja@insimtel.com

devices). That occasion, dimensions [4] of this horn antenna are small 48mm x 38mm x 35mm.

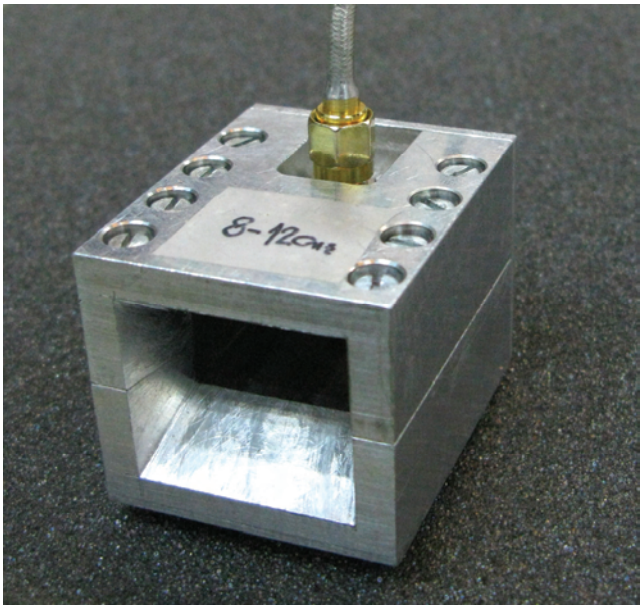


Fig. 1. Small size, wide band, high frequency horn antenna

Theoretical diagram of radiation, obtain by simulation, represented with blue line and measured diagram of radiation represented with red line are shown on Fig. 2..

As it was suppose, measured diagram of radiation has same shape as theoretical, but somewhere with smaller and somewhere with bigger deviation.

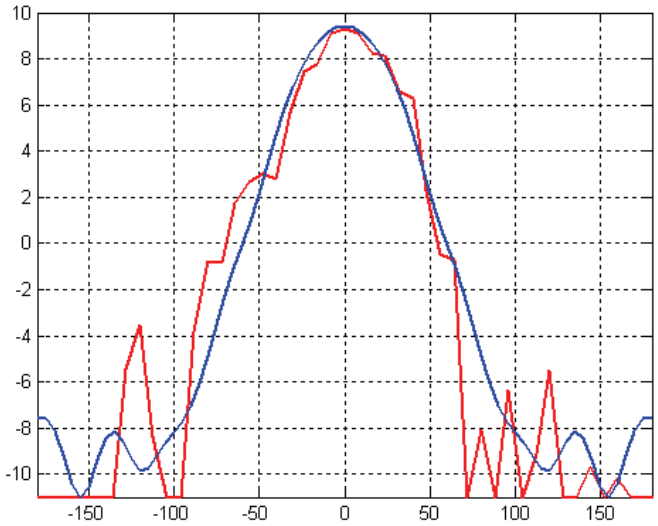


Fig. 2. Theoretical and measured diagrams of radiation

Theoretical and measured diagrams have been shown for 10 GHz frequency of source radiation and they are not same for all operating frequencies of antenna. There are two parents of this anomaly: (1) it is impossible to create wide band antenna with same diagram of radiation for all frequencies, (2) more problems are cause by matching impedance between antenna and detector microwave board. Both problems are physical nature and digital processing unit must deal with them. Fig.3. represent network analyzed horn antenna gain.

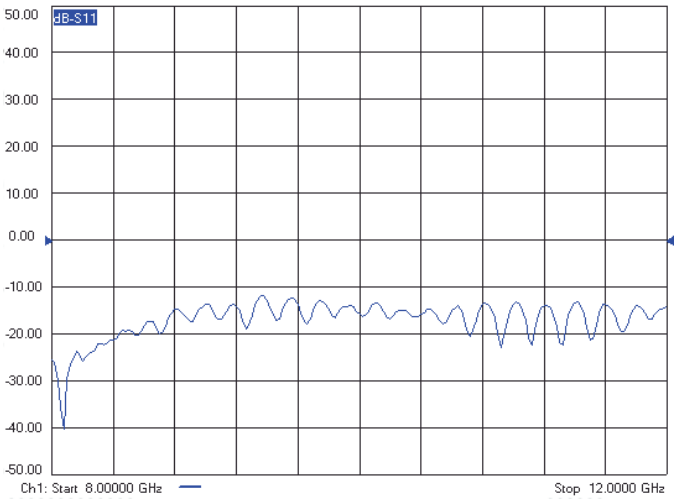


Fig. 3. Network analyzed horn antenna gain

Microwave board with signal detector block in this case is basing on AD8317 chip [5] as it is shown on Fig. 4.. That is wide band chip and operates from 1GHz up to 10GHz, and could detect signals slower then 25ns. Since this detector could not cover complete antenna band, mixer down converted input signal to the band from 1GHz to 4GHz.

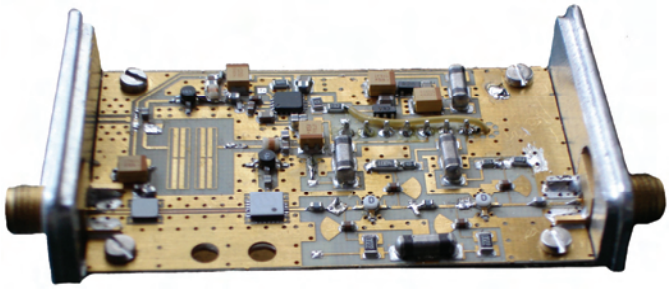


Fig.4. Logarithmic Detector

Note that this describes hardware could obtain only mono-channel processing. Main question is how and could it been modified to afford multi-channel processing? Fortunately, there is solution with using switching synthesizer. Idea is in hopping from channel to channel at some scheme and collecting data. Simplify scheme is hopping to the next channel until the last one is reached and then return to the first one. Little be more complicated scheme would be hopping to channel with lowest noise.

This method collecting data do not capture data exactly at the same time from all channels. This little disadvantage could not be eliminated but could be decrease speeding up switching. On the other side, speeding up has physical limits shown in synthesizer PLL set up time. Today, there are very speed PLL-s and set up time could be around 1μs. Full solution to avoid quasi-parallelization is independent detectors for all channels. This solution is certainly better in electrical sense, but size and price of system are pretty much bigger.

Central part of digital signal processing unit [6] is digital signal processor TMS320F2809 [7] as it shown on Fig.5..

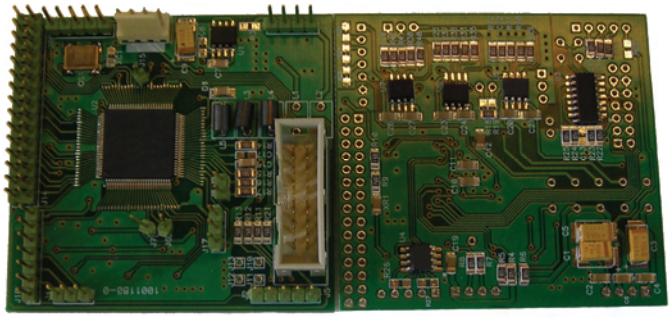


Fig. 5. Digital processing unit

Signal detector hardware allows very high-speed signals detecting that must be capture by interface, corrections of many physical imperfections demands a lot of memory, so DSP is a natural solution for this hardware. Digital signal processing unit has many tasks:

1) Processing digital data information.
2) Synthesis control.
3) Correcting physical imperfection of system.
4) Communication.

Last task is not crucial for determining right signal processing task, but is important for advance system applications.

Synthesis control is proceeding simultaneously with collecting and processing arrived signal data - DSP triggered synthesis to hop on next channel, then read data and at the end process received data.

III. ARRIVED SIGNALS ANALYSIS

Hardware described in pervious section can obtain, with more or less accuracy, next information about arrived signal and radio silence:

1) Observing radio silence (covered band 8-12GHz).
2) Signals frequencies.
3) Each signal power level.
4) Signal shape in time domain.
5) Parameters of signal shape in time domain.

The first level of analysis is continuous process observing radio silence. Breaking radio silence may be caused by one or more signals. Information about breaking radio silence has been triggering unknown signal/signals processing.

Next level is getting number of signal that break radio silence, identifying frequencies and power level of arriving signal or signals. From the moment of trigger, until trigger disappears this level should be pursuing. Hopping from frequency to frequency obtain covering antenna band, but it is not its only purpose. Hopping, also, gives information about arrived signal frequency. That information is not exactly, since it is discrete. That roughly information about frequency is determined by channel filter bandwidth and in this hardware, it is 100MHz.

Third level of signal/signals analysis is determining signal/signals type. This area of analysis has many possibilities and many unknown variables. Arrived signal could be modulated signal, pulse signal, continues wave signal etc. Further, today data signal frequently comes with analog, digital, base band or pulse modulation.

From this perspective, base band signals (modulated or not) could be threat as same thing and in addition base band modulated signals could be demodulated. How to distinguish is arrived signal modulated or not? Unfortunately, if arrived signal is totally unknown there is no way to be sure 100%, but always left solution to start search for some characteristic signals parts. This search needs a lot of time and processor memory and in lot of cases could not be processing in real time. For this case, communication part of digital processing unit, describe in pervious section, may send arrived data for further analysis to specialized non real time hardware and continue with other real time tasks. Sometimes exists some expectations about arrived signal and in those cases, decoding could be obtained by real time.

Digital modulations PSK, ASK and FSK, are often generated and detected using the principle of QAM and required deeper analyze. The I and Q signals are combining into a complex-valued signal I+jQ, where *j* is the imaginary unit. The resulting so-called equivalent low-pass signal or equivalent baseband signal is a complex-valued representation of the real-valued modulated physical signal, the so-called pass-band signal or RF signal. Digital modulation schemes are possible only if the transmitter-receiver pair have prior knowledge of how data is encoded and represented in the communications system. Since, here transmitter is unknown parameter demodulation of arrived signals is impossible, except there are some expectation about it, such as a modulation frequency and quantization parameters.

One bit of data signal is nothing more then presents or absents of pulse signal. Now, consider only ASK where pulse and pause duration and power level are unknown. Usually, short pulses consider a lot of energy followed by long pause, and wide pulses are following by short pause. It is logical, since average signal in both cases are similar. Those two cases are worst cases for analyzing pulse parameters and have one big similarity and one difference. Similarity is in shape since one case is inversion of second, on the other side difference is in power level. Quality of measuring parameters of arrived pulses is in directly connection with resolution of measuring shorter part of pulse (pulse or pause). On a hardware level, quality of measuring depends on AD converter resolution. In this case, resolution of AD converter is 12MSPS that mean all pulses slower then 80ns could be measure. Using pick detectors, signals faster then 80ns could be only conclude, but exactly measuring of pulse can not be obtain. Why this is important? It is important for measuring pulse level. This way is possible to measure pulses level, even, they are faster then AD converter and with backward reconstruction pulse time could be estimate. Estimation must use assumption that arrived signal has consistence at the energy filed. At the end for periodicals pulses, frequency and pulse/pause ratio could be measuring too.

IV. TROUBLESHOOTING

Observing and analyzing some radio band is complex task, since it is requiring understanding of a several electrical

science fields. At first antennas, second RF, third digital signal processing and connections between those fields. Each junction produced one non-linearity in system. Until that non-linearity is monotone, everything is fine, since it represent bisection in system.

Bottlenecks on hardware level are matching impedance between antenna and detector, conversion from analog to digital domain and temperature and frequency stability. Temperature stability refers to logarithmic and picks detector exerciseing on detecting level. It is recommended to capture operating abilities of detectors in a wide temperature band and use them in digital domain for correction input data, before digital signal processing start. This solution is a kind of normalization that provides same reference level for the considered temperature band. Same, temperature has some influence on stability of synthesized frequencies. Solution could use similar principle, only here temperature variation brings corrections in setting parameters of synthesis. In both cases, system must be equipped with temperature sensor. Those calibrations are necessary for more prescient measuring and they do not entering degradation of system functionality.

Influence of temperature to synthesizing frequencies is one trouble, but system response on different frequencies is something completely different. At first, antenna and logarithmic detector are providing different matching impedance in wide frequency band. Consequences are non-consistent dedicating power level in considering band. It shows on two levels: first same level is representing with different value and second increment/decrement is not appropriate. Those deviations in some applications have not large influence but here they have to be press. On the other side, this correction leads to decreasing dynamic range of detectors.

V. CONCLUSION

Topic of this paper is analysis and dedicating radio signals and that is always unfathomable. Fortunately, there are stuffs that could be reachable and analyzable. First, observing radio silence is certainly possible. If radio silence is broken then it is possible to determinate carrier frequencies and power levels in each channel of arrived signals. In addition, some signals shapes in time domain are easily to determinate (basic shape), but modulated signals are more sophisticated and required more patients. Some types of modulation are possible to admeasure and some not. It is interesting to create advanced analysis for possibility of signals demodulation and investigate probability of success demodulation.

Hardware, made for this analysis has many advantages, since it provide all further analysis. Another good thing of this hardware is many communication protocols that allow this system to be a STIM sensor part that communicate with some NCAP like in IEEE1451 protocol [8].

Results of this analysis could be used in nuclear, radar, sonar, and medicine application, since all noted applications are dealing with unknown signals and demands hard real-time processing.

REFERENCES

[1] N. M. Mitrović, D. D. Obradović, S. A. Tasić, "High frequency low cost fast channel switching synthesizer", TELSIKS'09, IEEE Conference Proceedings, pp., Nis, Serbia, 2009.

[2] N. M. Mitrović, D. D. Obradović, S. A. Tasić, Z. . Gajić, "Digital frequency synthesizing (DDS and PLL) control using Silabs 8051F121 microcontroller", TELFOR'09, IEEE Conference Proceedings, pp., Beograd, Serbia, 2008.

[3] N. M. Mitrović, D. D. Obradović, P. S. Manojlović, "One method source radiation azimuth finding using horn antennas", ETRAN'09, Conference Proceedings, pp., Vrnjacka Banja, Serbia, 2009.

[4] C.A. Balanis, *Antenna theory – Analysis and design*, New York, John Wiley & Sons. Inc., 1997.

[5] *1 MHz to 10 GHz, 50 dB Log Detector/Controller*, Analog Devices, USA, 2005.

[6] J. Basílio Simões, João Cardoso, Nuno Cruz, and Carlos M. B. A. Correia, "A PC104 Multiprocessor DSP System for Radiation Spectroscopy Applications", ICSPAT'99, Conference Proceedings, Orlando, USA, 1999.

[7] *TMS320F2009, TMS320F2008, TMS320F2006, TMS320F2002, TMS320F2001, TMS320F2801, TMS320F2802, TMS320F2801x DSPs Data Manual*, Texsas Instruments, USA, 2009.

[8] *IEEE Standard for a Smart Transducer, Interface for Sensors and Actuators-Common Functions, Communication, Protocols, and Transducer Electronic Data Sheet (TEDS) Formats*, IEEE Instrumentation and Measurement Society, New York, USA, 2007.

Block-based Wave Digital Network of an Elliptic Filter in MATLAB/Simulink

Biljana P. Stošić

***Abstract* – Microstrip structures with two cascaded transmission lines in parallel branches can be modeled and analyzed by use of one-dimensional wave digital approach. A complex structure, divided into uniform segments, is efficiently modeled here by wave digital network. Frequency response is obtained by direct analysis of the block-based network formed in Simulink toolbox of MATLAB environment. An Elliptic filter, proving the response accuracy of the new technique, is given.**

***Keywords* – Wave digital approach, wave digital networks, microstrip circuits, Elliptic filter.**

I. Introduction

Modeling of the planar microstrip structures by wave digital elements, based on well known theory of wave digital filters [1-4], can be efficiently used for analysis in both the time and the frequency domains.

Wave digital network (WDN) represents a wave digital model of a complex microstrip structure. A structure has to be divided into several uniform transmission lines (uniform segments) where each segment is modeled by unit wave digital elements [1-2]. A lossless uniform transmission line is modeled by a two-port digital element with a delay occurring in the forward path – called the unit element (UE) [2]. The port resistances of UE are equal and correspond to the characteristic impedance of uniform segment.

There does not essentially exist just one type of microstrip structures, but a whole variety of quite distinct subclasses. This reflects the richness of WDNs, and the most appropriate one have to be chosen for structure at hand. Till now, stepped-impedance filters, nonuniform structures with linearly tapered lines, and stub-line structures are analyzed by use of suggested one-dimensional (1D) wave digital approach [5-12].

The basic idea of the 1D wave digital approach is to treat the complex structure as a typical connection of several uniform segments. The delays of uniform segments vary from one another, and because of this each segment has to be represented as cascade connection of a certain number of UEs. A way of determining a minimal number of sections in WDN is described in paper [7].

The wave digital model of two cascade-connected transmission lines, which is presented in Section II, represents the background to the modeling strategies that follow.

The WDN is formed directly in the Simulink toolbox of MATLAB environment. Signal flow diagrams are the basis for block-oriented simulation programs such as Simulink. In Section III, a block-based wave digital model of the structure is described.Response in WDN can be found by use of formed block-diagram network and some basic MATLAB functions.

Finally, a simulation validation of the proposed modeling and analysis approaches provided by means of an Elliptic filter is presented and discussed in Section IV.

The author is with the University of Niš, Faculty of Electronic Engineering, Department of Telecommunications, Aleksandra Medvedeva 14, Niš, Serbia. E-mail: biljana.stosic@elfak.ni.ac.rs

II. Cascaded Transmission Lines and their Wave Digital Model

In this section, the equivalence between cascade connection of a transmission line and an open stub and its wave digital model is described.

Consider now cascade connection of two transmission lines where the second one is an open stub, Fig. 1.

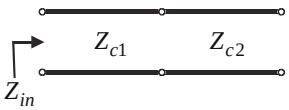


Fig. 1. Cascaded transmission lines

An input impedance of a lossles transmission line terminated by impedance Z_l has complex value

$$Z_{in} = Z_c \cdot \frac{Z_l + S \cdot Z_c}{Z_c + S \cdot Z_l}. \quad (1)$$

According to equation (1) and the fact that $Z_l \to \infty$ for an open line, an input impedance of the casacaded lines shown in Fig. 1 is

$$Z_{in} = \frac{Z_{c1}}{Z_{c1} + Z_{c2}} \cdot \left(S \cdot Z_{c1} + \frac{Z_{c2}}{S} \right). \quad (2)$$

Billinear frequency transformation is

$$S = \left(1 - z^{-1}\right) / \left(1 + z^{-1}\right), \quad (3)$$

where $z = e^{ST}$ is a complex variable and $S = j\Omega$ is a normalized complex frequency.

By replacing the normalized complex frequency in the relation (2) and assigning a new coefficient as

$$\alpha = \left(Z_{c1} - Z_{c2}\right) / \left(Z_{c1} + Z_{c2}\right), \quad (4)$$

the relation for input impedance can be written in form

$$Z_{in} = Z_{c1} \cdot \frac{1 - 2\alpha z^{-1} + z^{-2}}{1 - z^{-2}}. \quad (5)$$

Finally, the reflection coefficient is

$$\frac{B}{A} = \frac{Z_{in} - Z_{c1}}{Z_{in} + Z_{c1}} = \frac{(z^{-1} - \alpha) \cdot z^{-1}}{1 - \alpha \cdot z^{-1}}. \quad (6)$$

Further, a wave digital network depicted in Fig. 2 is observed.

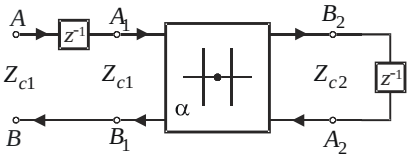


Fig. 2. Wave digital network

An equation system for two-port series adaptor is

$$B_1 = -A_2 - \alpha \cdot (A_1 + A_2),$$
$$B_2 = -A_1 + \alpha \cdot (A_1 + A_2), \quad (7)$$

where the adaptor's coefficient is

$$\alpha = (Z_{c1} - Z_{c2})/(Z_{c1} + Z_{c2}). \quad (8)$$

For the network, the additional relations can be written

$$A_2 = z^{-1} \cdot B_2, \quad A_1 = z^{-1} \cdot A \quad \text{and} \quad B = B_1. \quad (9)$$

The reflection coefficient is found by solving the new equation system obtained by substitution of the additional relations into equation system given by (7)

$$\frac{B}{A} = \frac{(z^{-1} - \alpha) \cdot z^{-1}}{1 - \alpha \cdot z^{-1}}. \quad (10)$$

Comparing the reflection coefficients given by the relations (6) and (10), it can be concluded that cascade connection of two transmission lines (Fig. 1) can be transform into wave digital network (Fig. 2) by using billinear transformation.

III. MATLAB/SIMULINK MODEL

To understand the modeling principle, consider for example the microstrip structure which comprises one parallel branch with cascaded transmission lines. Uniform segments have to be connected as depicted in Fig. 3.

A wave digital network which corresponds to the observed segment connection is depicted in Fig. 4. The uniform segments have different characteristic impedances Z_{ck}, $k = 1,2,3,4$. All uniform segments are modeled by several cascaded UE ($n_k \times T$ blocks), i.e. several sections.

Models of uniform segments UTL1, UTL2 and UTL4 are connected by use of one three-port parallel adaptor with port 2 being dependent. Models of the cascaded transmission line UTL2 and the open stub UTL3 are connected by use of one two-port series adaptor. The cascade-connected segments are connected to adaptor's dependent port 2.

Adaptors are memoryless devices whose task is to perform transformations between pairs of wave variables that are referred to different levels of port resistance.

A network of three-port parallel adaptor with port 2 being dependent is depicted in Fig. 5, [1-4]. The adaptor coefficients α_1 and α_2 are shown explicitly next to the ports 1 and 3, respectively. The coefficients are

$$\alpha_j = 2 \cdot G_j / (G_1 + G_2 + G_3), \quad j = 1,2, \quad (11)$$

where the port conductances are $G_k = 1/Z_{ck}$, $k = 1,2,3$.

In the symbolic representation of a two-port series adaptor [1-4] given in Fig. 6, it is shown explicitly the parameter α next to the port 1.

Block diagram is a representation of physical structure using blocks. Individual blocks can be put together to represent the structure in block diagram form. Individual blocks can be the basic blocks or they can be subsystems. They are considered in the text given below.

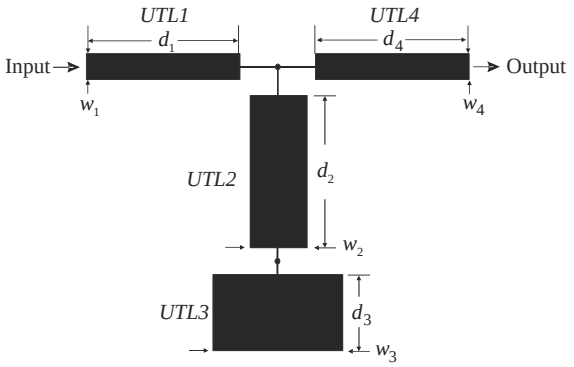


Fig. 3. Segment connection in microstrip structure

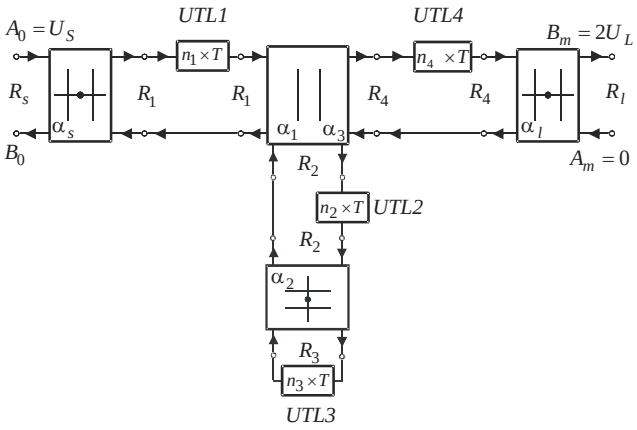


Fig. 4. WDN of the structure given in Fig. 1

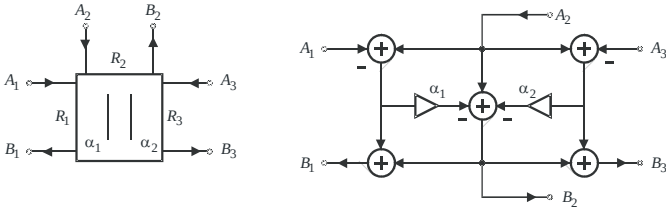


Fig. 5. Three-port parallel adaptor with port 2 being dependent: (a) symbol, and (b) WDE

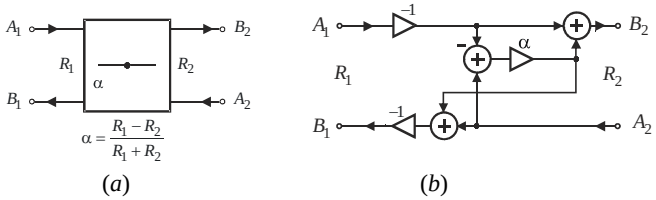


Fig. 6. Two-port series adaptor: (a) symbol, and (b) WDE

The formed Simulink model of WDN is depicted in Fig. 7. The blocks **TLine_1**, **TLine_2**, **Stub_3** and **TLine_4** represent uniform segments. The blocks **ADP-S**, **ADP-S23** and **ADP-L** represent two-port series adaptors, and block **ADP_T1S23T4** three-port adaptor. The two-port adaptors at the ends are used for matching source and load resistances to the rest of the WDN. The adaptor coefficients for the blocks **ADP-S** and **ADP-L** in the WDN are

$$\alpha_s = (R_s - Z_{c1})/(R_s + Z_{c1}), \quad (12)$$
$$\alpha_l = (Z_{c2} - R_l)/(Z_{c2} + R_l).$$

For the blocks **TLine_1**, **TLine_2**, **Stub_3**, and **TLine_4**, the Integer Delay blocks from Simulink/Discrete Library are used. The Integer Delay block delays its input by N sample periods. The block accepts one input and generates one output. Its mask is given in Fig. 8.

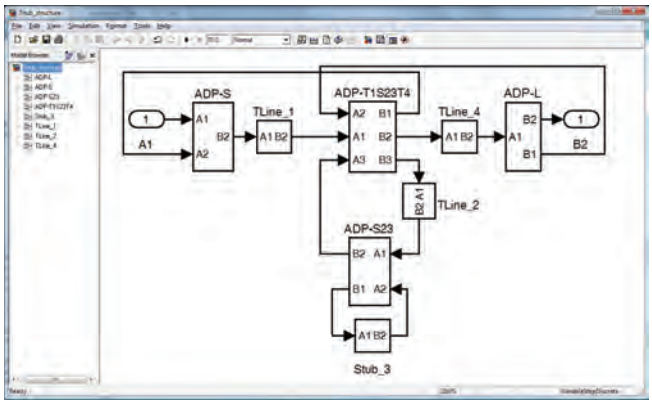


Fig. 7. Simulink model of WDN given in Fig. 4

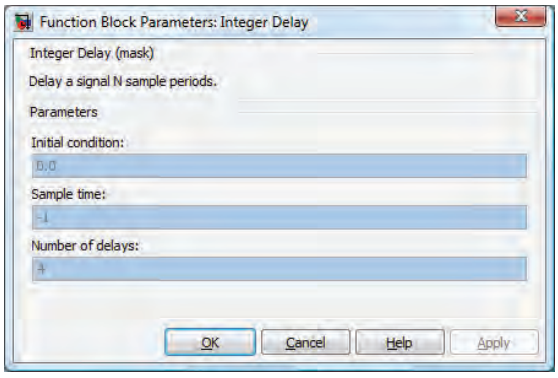


Fig. 8. A mask of the Integer Delay block

The study of discontinuities associated with the circuit (step, T-junction and open stub) is a very important. The simplest modeling approaches of the equivalent discontinuity networks involve changing line lengths. Simulink model given in Fig. 7 can be used in both cases where discontinuity effects aren't callculated and where they are modeled as described in the papers [7], [12], and [14].

A very simple method of analysis of the WDN is a block-diagram method. Each Simulink model represents a series of MATLAB and Simulink commands and functions which are used for its creation. Response is obtained directly in the time domain, and Fourier transformation is used for frequency response calculation.

A major part here is formed Simulink model (mdl-file), whereby the mdl-file is run by m-file that is provided for initialization, response calculation and plotting.

MATLAB built-in functions *dlinmod.m*, *dimpulse.m* and *fft.m* are employed to find a response of observed WDN. *Dlinmod.m* function obtains linear model from discrete system described by block-diagram. *Dimpulse.m* function gives impulse response of discrete-time linear systems. Function *fft.m* is the discrete Fourier transform, i.e. it gives response in the frequency domain.

IV. Approach Verification - An Elliptic Filter

The objective of this section is to prove the accuracy of the proposed modeling and analyzing approaches.

To demonstrate the main idea and approach, a microstrip Elliptic lowpass filter with a bandpass frequency of 1100 *MHz* [15] is depicted. The layout is shown in Fig. 9. The substrate dielectric constant is $\varepsilon_r = 2.55$, and the board thickness is $h = 558.80\,\mu m$. Metalisation is cooper and the metal thickness is $t = 61.4680\,\mu m$. Simulink model of WDN for Elliptic filter is depicted in Fig. 10.

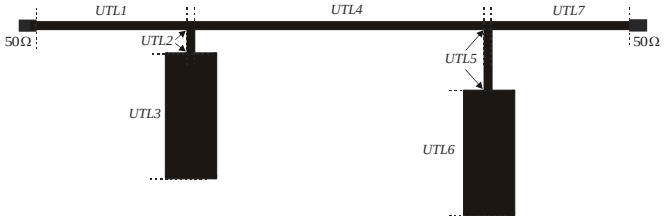


Fig. 9. Layout of lowpass Elliptic filter

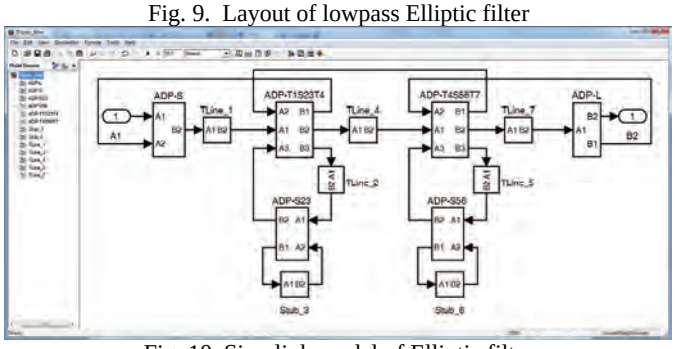


Fig. 10. Simulink model of Elliptic filter

The Elliptic filter is approximated by connection of seven uniform segments with parameters given in the Tables I and II. In order to absorb discontinuity effects new line lengths are counted. T-junction discontinuities are modeled by decreasing lengths of the lines in junctions (UTL1, UTL2, UTL4 in the first junction, and UTL4, UTL5, UTL7 in the other one) [12]. The effects of the open stubs are compensated by increasing lengths of the segments UTL3 and UTL6, [14]. According to the parameters given in the Tables, it can be concluded that the number of transmission line is the same in both cases, but their physical lengths differ one another and beacuse of that their delays differ either. The step discontinuities are modeled by increasing lengths of the lines in junctions (UTL2, UTL3 and UTL5, UTL6 segments are modeled). The characteristic impedances are chosen to be: for the transmission line used for approximation of the series inductance $150\,\Omega$, and for the line used for approximation of the parallel capacitance $5\,\Omega$ [7].

TABLE I. Segment parameters without modeled discontinuities

nv	d [mm]	w [mm]	Zc [Ohm]	Tv [ps]
1	15.2554	0.4229	92.9455	71.6036
2	2.2298	0.4229	92.9455	10.4659
3	11.7203	3.9569	24.9115	58.6098
4	30.8920	0.4229	92.9455	144.9967
5	6.2308	0.4229	92.9455	29.2454
6	10.0432	3.9569	24.9115	50.2232
7	12.6924	0.4229	92.9455	59.5738

TABLE II. SEGMENT PARAMETERS WITH MODELED DISCONTINUITIES

nv	d [mm]	Zc [Ohm]	Tv [ps]
1	15.0439	92.9455	70.6112
2	2.2403	92.9455	10.5151
3	12.0681	24.9115	60.3492
4	30.4691	92.9455	143.0119
5	6.2413	92.9455	29.2946
6	10.3910	24.9115	51.9625
7	12.4809	92.9455	58.5814

For given error of $n\_er = 0.001\%$, a total minimal number of sections in *WDN* is $n_t = \sum_{k=1}^{7} n_k = 3309$. The numbers of sections in individual segments $n_k = round\left[q \cdot T_k / T_{\min}\right]$ are 551, 82, 471, 1115, 228, 405, and 457, respectively. A total delay for the digital model of the structure is $T_t = n_t \cdot T_{\min} / q = 424.3260\ ps$ where $q = 82$ is a multiple factor and $T_{\min} = \min\{T_1, T_2, ..., T_7\} = 10.5151\ ps$ is a minimum delay. A total real delay of the structure is $T_\Sigma = \sum_{k=1}^{7} T_k = 424.3235\ ps$. A sampling frequency of the digital model of the planar structure for the chosen minimal number of sections is $F_s = n_t / T_t = 7798.2945\ GHz$. In this case, a relative error of delay is $er = \frac{T_\Sigma - T_t}{T_\Sigma} \cdot 100\,\% = 0.0005685\,\%$. According to the relation (11), the three-port adaptor coefficients are $\alpha_1 = 0.6667$ and $\alpha_2 = 0.5773$. The two-port adaptor coefficients are $\alpha = 0.6667$, and $\alpha_S = -\alpha_L = -0.3004$.

The results obtained by direct analysis of WDN, by linear simulations in the programs GENESYS and ADS (Advanced Design System), and by electromagnetic simulation done in ADS, are depicted in Fig. 11. It can be concluded that WDN and ADS curves are practically identical.

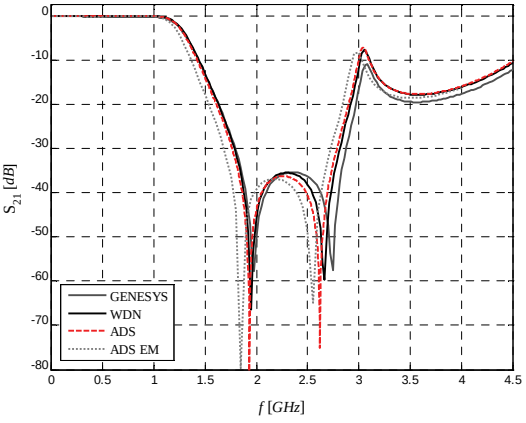


Fig. 11. Frequency response comparison

V. CONCLUSION

After an extensive search in the literature it was found an application of ADS to simulations of different microstrip structures based on their wave digital network representations [13]. The equivalent circuit model is solved through a circuit simulator. The adaptors with many multiply elements were used there.

The main objective of this paper is to give an original and general method to characterize the behaviour of microstrip structures in both the frequency and time domains by use of the wave digital networks in MATLAB/Simulink.

A wave digital approach into MATLAB provides easy simulation of microwave layouts such as microstrip structures of different geometries. Also, allows the user to get the fastest return out of 3D electromagnetic tool investments.

The broadband accuracy of the suggested procedure is validated by an Elliptic filter realized in the microstrip line technique. The analysis result obtained by WDN has shown a very good agreement with those obtained by other programs mentioned above.

REFERENCES

[1] A. Fettweis, "Digital Circuits and Systems", *IEEE Transactions on Circuits and Systems*, Vol. CAS-31, No. 1, January, 1984, pp. 31-48.

[2] A. Fettweis, "Wave Digital Filters: Theory and Practice", *Proc. IEEE*, Vol. 74, 1986, pp. 270-327.

[3] W.K. Chen, *The Circuits and Filters Handbook*, CRC Press, 1995 (Wave Digital Filters, pp. 2634-2661).

[4] M.V. Gmitrović, *Microwave and Wave Digital Filters*, Faculty of Electronic Engineering, Niš, 2007 (in Serbian).

[5] B.P. Stošić, *Analysis of Planar Microwave Structures Modeled by Wave Digital Elements*, Doctoral thesis, Faculty of Electronic Engineering, University of Niš, September 2008 (in Serbian).

[6] B.P. Stošić and M.V. Gmitrović, "Implementation of Wave Digital Model in Analysis of Arbitrary Nonuniform Transmission Lines", *Microwave and Optical Technology Letters*, Vol. 49, No. 9, September 2007, pp. 2150-2153.

[7] B.P. Stošić and M.V. Gmitrović, "Wave Digital Approach - Different Procedures for Modeling of Microstrip Step Discontinuities", *International Journal of Circuits, Systems and Signal Processing*, Issue 3, Volume 2, 2008, pp. 209-218.

[8] Advanced Design Software 2002, Agilent Technologies 1983-2002, 395 Page Mill Road, Palo Alto, CA 94304, USA.

[9] B.P. Stošić and M.V. Gmitrović, "Block-based Analysis of Microstrip Structures with Stubs by use of 1D Wave Digital Approach", *XLIV Intern. Scientific Conference on Information, Communication and Energy Systems and Technologies - ICEST 2009*, Bulgaria, Veliko Tarnovo, June 25-27, 2009, pp. 23-26.

[10] B.P. Stošić and M.V. Gmitrović, "A Wave Digital Approach in Obtaining *z*-domain Functions for Microstrip Stub-line Structures", *The 9th Intern. Conference on Telecomm. in Modern Cable, Satellite and Broadcasting Services - TELSIKS 2009*, Serbia, Niš, October 7-9, 2009, Vol. 1, pp. 193-197.

[11] B.P. Stošić and M.V. Gmitrović, "Wave-based Modeling and Analysis of Microstrip Stub-line Structures", *17th Telecommunications forum TELFOR 2009*, Serbia, Belgrade, November 24-26, 2009, pp. 533-539.

[12] B.P. Stošić, "Using z-variable Functions for the Analysis of Wave-based Model of Microstrip Stub-line Structure", *Microwave Review*, December 2009, Serbia, pp. 6-11.

[13] F. Maggioni, "Time Domain Electrical Simulation using Equivalent Digital Wave Networks in ADS", *ADS User Group Meeting*, Rome, May 13, 2009.

[14] P.F. Combes, J. Graffeuil and J.-F. Sautereau, *Microwave Components, Devices and Active Circuits*, John Wiley & Sons, New York, 1987.

[15] R.W. Rhea, *HF Filter Design and Computer Simulation*, Noble Publishing Corporation, USA, 1994.

Investigation on EHF Radio Link Availability in Bulgaria

Boncho G. Bonev[1], Metodi P. Yankov[2], Kliment N. Angelov[3]

***Abstract* – The attenuation of the EHF radio waves in hydrometeors, especially in rain, and in the earth's atmosphere severely affects the quality of a radio link, working in that frequency range, and limits its path length. In the present article a study on the quality of a radio link, working on the 30-100GHz range, is made, using the ITU models for rain and atmospheric attenuation, and the Crane global model for rain attenuation. The results show, that in the 67-100GHz range the attenuation does not increase significantly, and therefore that range can be successfully used for independent radio link, as well as a part of a FSO/RF communication systems. The research is made based on actual statistical rain rate data for the Bulgarian region.**

***Keywords* – EHF, rain attenuation, atmospheric gaseous attenuation, link availability.**

I. INTRODUCTION

The constant demand of higher data rates requires the use of higher frequencies, such as the EHF (30-300GHz) range, for radio transmission. Radio links, working in that range can be used independently, as well as a part of a hybrid FSO/RF communication system. The EHF diapason is very practical for use in those systems, because of its atmospheric propagation characteristics. They can successfully compensate the disadvantages of the optical wave propagation, used in the FSO systems.

The imminent problems that occur when using millimeter waves are the high losses, caused by absorption and scattering by the atmospheric gases, as well as by the hydrometeors, especially rain [1 – 4]. The atmosphere is completely transparent for frequencies below 3 GHz. Above that frequency the radio waves are significantly absorbed by hydrometeors, and above 20 GHz – by the atmospheric gases [5, 6, 9].

In our previous researches, the link availability was investigated for a number of typical geographic regions in Bulgaria – Sofia, Smolyan etc., using the ITU model [11 – 13] for rain attenuation. In the present article that study is done for Bulgaria overall, using the Crane global model. The difference in this model is that it takes into account the rainfall length in depending on rain rate, when the total losses in the rain are calculated.

[1]Boncho G. Bonev is with the Faculty of Telecommunications, Technical University of Sofia, Kliment Ohridski Blvd. No 8, 1000 Sofia, Bulgaria, E-mail: bbonev@tu-sofia.bg

[2]Metidi P. Yankov is with the Faculty of Telecommunications, Technical University of Sofia, Kliment Ohridski Blvd. No 8, 1000 Sofia, Bulgaria, E-mail: m_yankov@abv.bg

[3]Kliment N. Angelov is with the Faculty of Telecommunications, Technical University of Sofia, Kliment Ohridski Blvd. No 8, 1000 Sofia, Bulgaria, E-mail: kna@tu-sofia.bg

II. THEORETICAL ANALYZIS

Lets analyze a radio link with a length d[m], working on frequency f [GHz], respectively wavelength λ [m]. The received power, P_r [dBm] can be derived from the following equation:

$$P_r = P_t + G_t - L + G_r , \quad (1)$$

where P_r is received power in dBm, P_t – transmitted power in dBm, G_t,G_r are gains of the transmitter/receiver antenna in dBi, L - total propagation losses in dB.

The total losses are due to three major factors – free space loss, attenuation in the atmosphere, and attenuation in rain:

$$L = L_{FS} + L_{atm} + L_{rain} . \quad (2)$$

The free space losses are calculated by the well known expression [6]

$$L_{FS} = 20\lg\frac{4\pi d}{\lambda} = 20\lg\frac{4\pi df}{c} , \quad (3)$$

where c=3.10^8 m/s is the light speed in vacuum.

The losses in the atmosphere are given by the formula:

$$L_{atm} = L_{sp\_atm}.d\,/1000 , \quad (4)$$

where L_{sp_atm} is the specific attenuation in the atmosphere, dB/km [9].

The rain attenuation L_{rain} in dB as defined by the Crane global model [5, 7] is:

$$L_{rain} = k.RR^{\alpha}\left(\exp(y.\delta)-1\right)/\,y , \text{ when } 0 < d < \delta \quad (5)$$

and

$$L_{rain} = k.RR^{\alpha} r , \quad (6)$$

where

$$r = \frac{\exp(y.\delta)-1}{y} + \frac{\left(\exp(z.d)-\exp(z.\delta)\right).\exp\left(0.83-0.17\ln RR\right)}{z} , \quad (7)$$

when $\delta < d < 22.5\,km$.

In Eq. (5) and (7) δ in km is

$$\delta = 3.8 - 0.6 \ln RR \,. \tag{8}$$

And *y* and *z* are defined as follow

$$y = \alpha \left[\frac{0.83 - 0.17 \ln RR}{\delta} + 0.26 - 0.03 \ln RR \right] \tag{9}$$

$$z = \alpha . \left(0.026 - 0.03 \ln RR\right). \tag{10}$$

In equations (5-9) *RR* is the rain rate, *k* and α are frequency and polarization depending coefficients and can be taken from [8].

We can see from equations (2) – (10), that the total propagation losses of the radio link are frequency, distance, and rain rate dependant – $L(f, d, I)$.

From expressions (1) – (10), using numerical methods, we can calculate the maximum rain rate, at which we still have connectivity - $RR_{\max}$.

If for the given radio link the transmitted power P_t, antenna gains G_t and G_r, receiver sensitivity for a given *BER* - P_{r_min} and the frequency *f*, are constants, $RR_{\max}$ will depend only on the link distance *d* - $RR_{\max}$=$RR_{\max}$ *(d)*. Then, when we have the statistical rain rates data for the geographical region, and more important, the percentage of time, when the rain rate exceeds a certain point - *p(RR)=p(RR'>RR)* based on the function $RR_{\max}$ *(d)*, we can calculate the exact percentage of time, when the connectivity will be lost - *p(d)* [10]

$$p(RR) = p\left(RR(d)\right) = p(d) \tag{11}$$

III. NUMERICAL RESULTS

The above equations are used for a theoretical calculation of the link breakage probability for radio links working in the 30-100 GHz frequency range. In order to do that, the rain rates statistical data for Bulgaria is used. The percentage of time, when the rain rate exceeds a certain point $p(RR)$ is shown on Fig. 1.

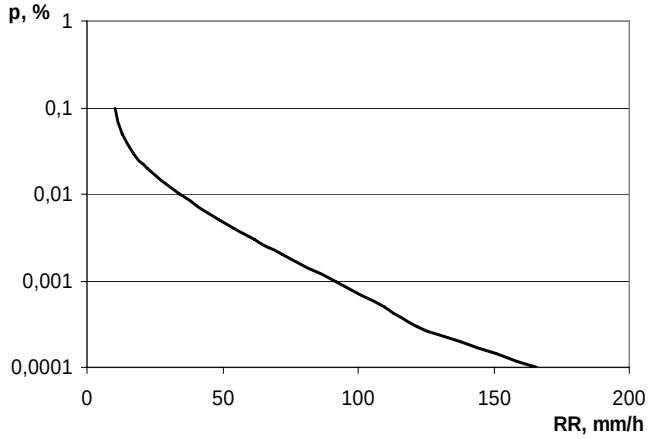


Fig.1. Rain rate statistic for Bulgaria

The function from Fig. 1 has been interpolated with the following expression

$$p,\% = 10^{A.RR^6 + B.RR^5 + C.RR^4 + D.RR^3 + E.RR^2 + F.RR + H}, \tag{12}$$

where the values of the constants *A, B, C, D, E, F* and *H* are given in Table 1. The error of this approximation is calculated to be 0.15%.

TABLE I. VALUES OF APPROXIMATION CONSTANTS

A	B	C	D	E	F	H
4.87E-12	***-2.79E-9***	***6.36E-7***	***-7.3E-5***	***4.426E-3***	***-0.1507***	***0.1082***

The link constants are as follows P_t = 10 dBm, antennas diameter is fixed and for *f*=60 GHz their gain is $G_t = G_r$ = 43 dBi and receiver's sensibility is P_{r_min} = –60 dBm.

The coefficients *k* and α can be determined from [8] and the specific attenuation in the atmosphere from [9].

In Fig. 2 - Fig. 5 are presented the values of maximum rain rate and the values of link breaking probability depending on frequency *f* and radio link distance *d*.

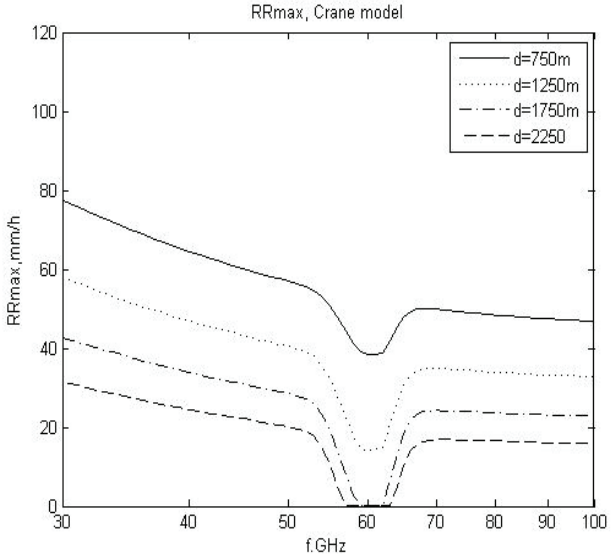


Fig. 2. Maximum rain rate depending on frequency for different values of link distance.

Fig. 2 shows that in frequencies of 67 – 100 GHz the maximum admissible rain rate $RR_{\max}$ is varying a little and it is slightly lower than in frequencies 40 – 55 GHz.

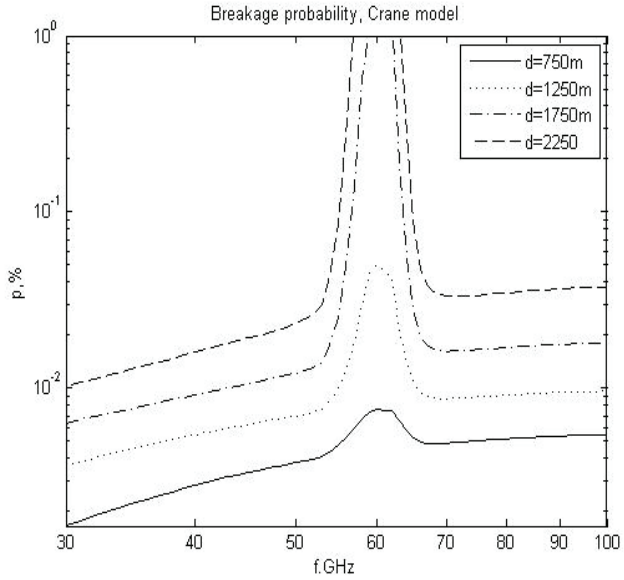


Fig. 3. Link breaking probability in %, depending on frequency for different link distance.

Fig.3 show that the minimum breaking probability is achieved at frequencies from 30 to 55 GHz but for these frequencies the data rate will be smaller than in frequencies of 67 – 100 GHz where the breaking probability is greater, but remain of the same order. Therefore we can say that the frequency range 67 – 100 GHz is better for this type of links because it provide greater data transmission speed on approximately same distances as the range 30 – 55 GHz.

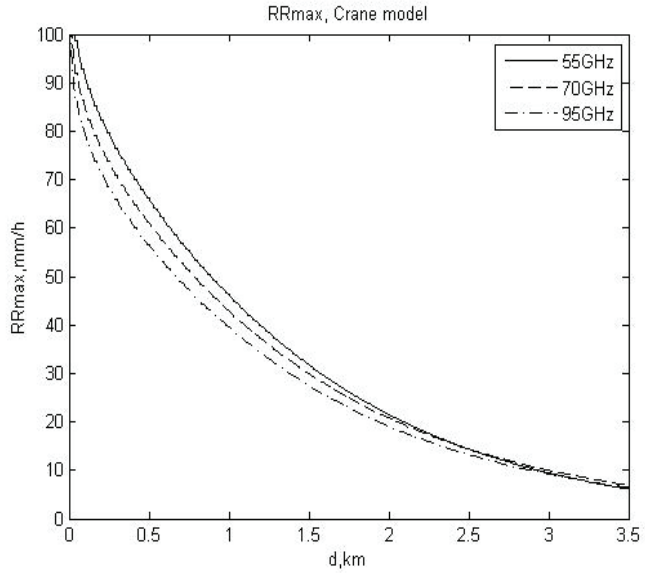


Fig. 4. Maximum rain rate depending on link distance for different values of frequency.

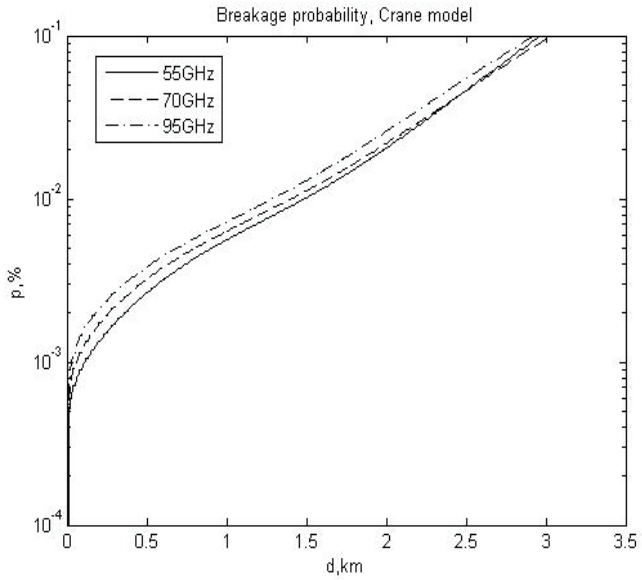


Fig. 5. Link breaking probability in %, depending on link distance for different frequencies.

Fig. 4 and Fig. 5 show that if a link breaking smaller than 0.01% is required, the link distance is limited to 1,2 – 1,4 km for the studied frequencies.

IV. Conclusion

The frequencies in the 67 GHz – 100 GHz diapason can be used for radio links on distances of about 1 km with link availability greater than 99.99%, which is sufficient enough for some type of radio links. The rain attenuation at higher frequencies is greater, but in this case a narrower beam, and respectively the higher antenna gain can be produced with the same sized antenna as for the lower frequencies. Furthermore, in this case, the link distance is almost the same as in lower frequencies if the same sized antennas are used.

If the greater than 1 – 1.5 km distances or greater than 99.99% link availability have to be achieved, the hybrid Free-Space Optics/Radio Frequency (FSO/RF) communication links can be used.

In our previous works [11 – 13], a similar study for different Bulgarian regions was done, using the ITU model. These studies show that the Crane model calculates significantly higher losses than calculated by the ITU model for the same rain rates. Therefore, the connections break probability is about one order higher, when calculating by the Crane global model. This significant difference in the results demands a practice realization and experimental researches on EHF radio link in studied frequency diapason for a more accurate analysis of the studied problem.

Acknowledgement

Scientific researches whose results are presented in this paper are financed by Internal competition of Technical University of Sofia.

REFERENCES

[1] V. Kvicera, M. Grabner, M. Hlavaty, "*Rain Intensity Statistical Processing and Comparison with ITU-R Recommendations*", Radioengineering, Vol. 13, No. 2, pp. 1-2, June 2004.

[2] G. Timms, V. Kvicera, M. Grabner, "*60 GHz Band Propagation Experiments on Terrestrial Paths in Sydney and Praha*", Radioengineering, Vol. 14, No. 4, pp. 27 – 32, December 2005.

[3] W. Asen, C. Gibbins, "*A comparison of rain attenuation and drop size distributions measured in Chilbolton and Singapore*", *Radio Science*, 2002, Vol. 37, No. 3, p. 6-1 to 6-14.

[4] R. Crane, *"Electromagnetic Wave Propagation Through Rain"*, New York: Wiley-Interscience, 1996, pp 245-259.

[5] J. Seybold, *"Introduction to RF propagation"*, John Wiley & Sons Inc., Hoboken, New Jersey, 2005.

[6] R. L. Freeman, *"Radio System Design for Telecommunications"*, John Wiley & Sons Inc., Hoboken, New Jersey, 2007.

[7] R. Crane, "*Propagation Handbook for Wireless Communication System Design"*, CRC Press, 2003.

[8] Rec. ITU-R P.838-3, *"Specific attenuation model for rain for use in prediction methods"*.

[9] Rec. ITU-R P.676-6, "*Approximate estimation of gaseous attenuation in the frequency range 1 – 350 GHz*".

[10] E. Ferdinandov, B Pachedjieva, *"Probability and statistic methods in communications" – Part 1*, Siela, Sofia, 2005. (in Bulgarian)

[11] B. Bonev and Kl. Angelov, "*Investigation on Millimeter Waves Usage in Hybrid FSO/RF Communication Systems"*, Proc. ICEST'2009, Veliko Tarnovo, Bulgaria, 25-26 June, 2009.

[12] B. Bonev, Kl. Angelov, *"Study on link availability for Radio Lines Working in Milimeter Waves Frequency Diapason"*, Proc. TELECOM'2009, Varna, Bulgaria, October 2009. (in Bulgarian)

[13] B. Bonev, Kl. Angelov, M. Yankov, Ts. Mitsev, "*Investigation on influence of attenuation in hydrometeors and atmospheric gases on link range at millimeter waves radio lines"*, Proc. "Smolyan-2009", Smolyan, Bulgaria, September 2009. (in Bulgarian)

Design and Analysis of Ultra-Wideband Low Noise Amplifier in 0.13μm CMOS Technology

Alena Djugova[1], Jelena Radic[2] and Mirjana Videnovic-Misic[3]

***Abstract* – In this paper design of ultra-wideband (UWB) low noise amplifier (LNA) in 0.13 μm CMOS technology is presented. As a first block in a receiver chain LNA needs to provide adequate gain, low noise figure, input and output matching, low power consumption, small chip area and very wide bandwidth. For proposed LNA topology simulation results for basic figures of merit are given. Optimized LNA has maximum gain (S_{21}) of 15.77 dB with the 3-dB band from 2.5 to 5.4 GHz and return losses (S_{11} and S_{22}) of less than –8.4 dB and –18.9 dB, respectively. In addition, average noise figure of 1.5 dB is achieved, with power consumption of only 3.2 mA at 1.2 V.**

***Keywords* – Ultra-wideband (UWB), Low noise amplifier (LNA), Inductive degeneration, Quality factor, Resistive shunt-feedback, Inductive shunt peaking.**

I. INTRODUCTION

Ultra-wideband (UWB) system presents an emerging, low power, wireless technology offering a high data rates at small distances. UWB allocated band (IEEE 802.15.3a) is between 3.1–10 GHz (low-frequency band from 3.1–5 GHz and high-frequency band from 6–10.6 GHz) [1]. Two major solutions, Multi-band Orthogonal Frequency Division Multiplexing (MB-OFDM) based on frequency hopping and Direct Sequence UWB (DS-UWB) are proposed to transmit the data rate up to 480 Mbps by using only the low frequency band. This low frequency band has been allocated for development of the first generation of UWB systems.

Design of low noise amplifier (LNA) maintains to be one of the challenging tasks in up-to-date receiver design. Being the first circuit in the receiver's chain it must meet several stringent requirements. It needs to provide sufficient gain to amplify received weak signal and to overcome the noise of the subsequent stages, low noise figure (NF) to improve sensitivity and adequate input and output matching to improve reflection coefficients. All this requirements have to be fulfilled within defined (wide) bandwidth and with low power consumption. In Section II several circuit techniques, proposed in literature to target these goals, are described.

In this paper broadband LNA topology, which merge narrowband LNA architecture with inductive source degeneration and resistive shunt-feedback technique, is proposed. Analysis of chosen topology is given in Section III. Simulated figures of merit (FOMs) are given in Section IV followed by discussion of achieved trade-offs and advantages of presented topology. The Section V concludes the paper.

[1]Alena M. Djugova is with the Faculty of Technical Sciences, Trg D. Obradovica 6, 21000 Novi Sad, Serbia, E-mail: alenad@uns.ac.rs

[2]Jelena B. Radic is with the Faculty of Technical Sciences, Trg D. Obradovica 6, 21000 Novi Sad, Serbia, E-mail: jelenar_@uns.ac.rs

[3]Mirjana S. Videnovic-Misic is with the Faculty of Technical Sciences, Trg D. Obradovica 6, 21000 Novi Sad, Serbia, E-mail: mirjam@uns.ac.rs

II. WIDEBAND AMPLIFIERS TOPOLOGIES

Two major UWB LNA design goals are wide band and low NF. There are four basic techniques used to obtain wide frequency band: distributed amplifier, common gate circuit, inductively degenerated common source amplifier incorporated with additional input band-pass LC-filter and common source amplifier with resistive shunt-feedback.

Distributed amplifiers can provide flat gain over the wide frequency band. To achieve necessary gain several amplifying stages are needed. Consequently, this approach is power hungry and area consuming. Moreover, difficult NF optimization for this technique is still an issue [2].

Second approach utilizes common gate circuit that provides wideband input matching. Main disadvantage of common gate amplifiers is their relatively low transconductance value that can not provide low noise and high gain in a whole frequency range. This type of amplifier is usually used as the first stage of multi cascade amplifiers where the next stages enhance the amplification bandwidth [3].

The inductively degenerated common source amplifier incorporated with additional input band-pass LC-filter represents another technique for achieving broadband input matching. In addition gain flatness and low power consumption are achieved. However, large number of on-chip reactive elements demand high chip area and degrade NF at higher frequencies [4].

Common source amplifier with resistive shunt-feedback provides wideband input matching and flat gain [5, 6]. Input resistance is determined by the feedback resistance divided by the voltage gain of the common source amplifier [7]. To obtain 50Ω input match few hundred Ohms for feedback resistance is needed, which results in NF degradation. Moreover, due to strong dependence of voltage gain on the amplifying transistor transconductance, the amplifier requires a large amount of current to achieve high gain.

III. CIRCUIT DESIGN

In this paper, an inductive degenerated amplifier circuit is modified with resistive feedback. This method is selected to achieve high gain, low noise and low power consumption LNA for the low-frequency band from 3.1–5 GHz. Proposed method is based on extending the operational bandwidth of

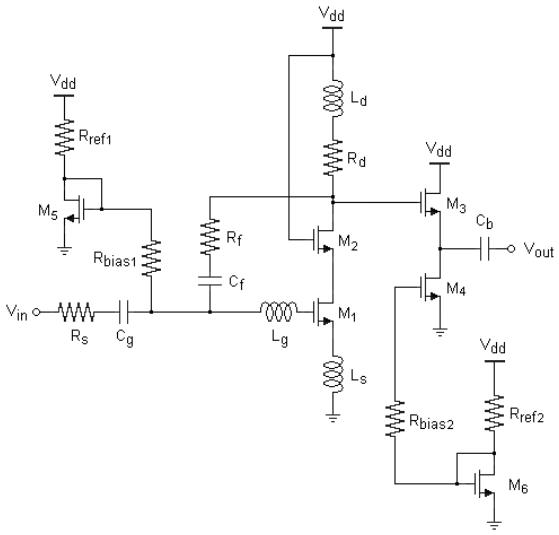


Fig. 1. Cascode feedback LNA topology

existing narrowband systems in order to cover UWB frequency band.

Schematic of a one stage LNA is shown in Fig.1. Transistors M_1 and M_2 are connected in cascode to improve the reverse isolation of the circuit. Additionally, cascode topology reduces effective M_1 input capacitance and helps to obtain higher gain and broader bandwidth. The source follower composed of transistors M_3 and M_4 form the output buffer and provides output wideband matching. Bias circuit is composed of resistors R_{ref1}, R_{bias1} and transistor M_5, where transistor M_5 forms current mirror with M_1. To decrease the overall power consumption (P_D), transistor M_5 width, W_5, should be small fraction of transistor M_1 width, W_1. The value of R_{bias1} is chosen large enough to provide high impedance path to RF signal while giving at the same time small contribution to the circuit noise. For biasing of output matching circuit the same bias circuit topology is used, composed of transistor M_6 and resistors R_{ref2} and R_{bias2}.

LNA architecture is input matched using inductive source degeneration (L_s in the schematics). As this method do not use additional noisy components (resistors) lower noise figure (*NF*) is expected. The source degenerated LNA first order input impedance Z_{in} (assuming feedback effect negligible and ignoring C_{gd}) is equal to:

$$Z_{in} = j\omega(L_s + L_g) + \frac{1}{j\omega C_{gs}} + \frac{g_m}{C_{gs}} \cdot L_s, \qquad (1)$$

where C_{gs} is total gate-source capacitance and g_m is transconductance of the input RF transistor M_1. At resonant frequency ω_0, where imaginary parts cancel, input impedance becomes:

$$imag(Z_{in}) \mid \omega_0 = 0 \rightarrow Z_{in} = \frac{g_m}{C_{gs}} \cdot L_s = \omega_T \cdot L_s \ . \qquad (2)$$

In case of the perfect input LNA match with source degeneration inductor L_s, input impedance $Z_{in} = \omega_T \cdot L_s$ is equal to source resistance $R_s = 50\Omega$.

An additional degree of freedom is provided by the gate inductance L_g that is used to set the resonance frequency ω_0 given by:

$$\omega_0 = \left[(L_g + L_s) \cdot \frac{C_{gs} \cdot C_g}{C_{gs} + C_g} \right]^{-\frac{1}{2}}. \qquad (3)$$

Quality factor of the series resonating input circuit is:

$$Q_{nb} = \frac{1}{(R_s + \omega_T \cdot L_s) \cdot \omega_0 \cdot C_{gs}}. \qquad (4)$$

The (–3-dB) bandwidth of RLC resonant circuit is inversely proportional to its Q-factor ($BW_{-3dB} = \omega_0/Q_{nb}$) [8], therefore high Q value is not desirable in the wideband design. To achieve broad band, resistor R_f is added as a shunt-feedback element to reduce Q-factor of the resonating narrowband LNA circuit. Resistance R_f can be much larger that in the conventional resistive shunt-feedback circuits where the size of the R_f is limited as it determines low input impedance. Miller equivalent input resistance of R_f is given by [7]:

$$R_{feq} = \frac{R_f}{(1 - A_v)}, \qquad (5)$$

where A_v is the open-loop voltage gain of an LNA. With this transformation Q-factor of the input circuit shown in Fig. 1 is given by:

$$Q_{wb} = \frac{1}{(R_s \,||\, R_{feq} + \omega_T \cdot L_s) \cdot \omega_0 \cdot C_{gs}}. \qquad (6)$$

It can be seen from (6) that by proper R_{feq} selection narrowband LNA can be transformed into wideband LNA.

In Fig. 1 resistive and capacitive shunt-feedback (R_f and C_f) are employed to produce better stability, gain flatness and bandwidth. Capacitor C_f is used for the ac coupling purpose while large values of feedback resistor also improve input impedance matching without affecting the noise figure significantly [9]. Additionally, technique called inductive shunt-peaking, applied by series connection of L_d and load resistance R_d, is used to increase bandwidth and improve gain flatness [8].

Capacitors C_g and C_b are input and output DC blocking capacitors. Their values are chosen large so they do not influence the resonant frequency of the input and output circuit, respectively.

IV. Simulation Results

The designed circuit has been simulated using Spectre Simulator from Cadence Design System. To achieve wideband LNA that covers frequency band from 3.1–5 GHz narrowband amplifier (without feedback) was optimized at 4GHz central frequency. To obtain initial device dimensions and component values in UMC 0.13µm CMOS eight-metal

technology the optimization technique with constant power consumption is used [10]. After shunt-feedback resistor introduction frequency band is extended to desired boundaries. In order to get simulation results as close as possible to measurement results, BSIM3V3 models for all circuit components were introduced. Due to increased complexity of the circuit it was necessary to carefully examine dependence of the LNA FOMs on circuit parameters. This has been carried out by changing values of one by one circuit parameter analyzing at the same time the LNA performance behavior.

LNA topology was optimized with the main aim to minimize *NF* and P_D, improve voltage gain while still keeping acceptable values for remaining FOMs. Simulation results for the cascode LNA topology are given in Figs. 2–4 (S_{11} and S_{22}, S_{12} and S_{21}, *NF* and NF_{min}).

Parameter S_{11} is below –8 dB in the band of interest (–8.812 dB at 3.1 GHz and –8.336 dB at 5 GHz) and is less than –7 dB from 3–5.4 GHz. This demonstrates the effectiveness of wideband matching realized by matching network (L_s and L_g), shunt-feedback and shunt-peaking load. The lowest S_{11} value can be seen at 4 GHz. This result is in accordance with expectations as initial LNA topology is optimized at this frequency. Increase in R_f value results in flatter characteristics, with S_{11} minimum shifted to the higher frequencies and higher maximum S_{11} value (S_{11} closer to zero). The same effect can be seen with increase in L_d inductor value. The first block of the proposed LNA topology can be roughly seen as basic common source amplifier with resistive R_f shunt-feedback and $Z_D = R_d + j\omega L_d$ drain impedance. Though there are some difference, such as source degeneration and cascode part of topology, some R_{in} and A_v properties can be more easily seen in simplified structure. In the basic common source amplifier with resistive R_f shunt-feedback and Z_D drain resistance input resistance and voltage gain at lower frequencies are:

$$R_{feq} = \frac{R_f + Z_D}{1 + g_m \cdot Z_D}, \tag{7a}$$

$$A_v \approx -\frac{g_m \cdot R_f \cdot Z_D}{R_f + g_m \cdot R_s \cdot Z_D} = -\frac{1}{\frac{1}{g_m \cdot Z_D} + \frac{R_s}{R_f}}. \tag{7b}$$

From (7a) can be seen that R_f increase results in R_{feq} increase. According to (6) Q_{wb} is lower and BW wider, which is in accordance with simulation results. To shift the position of the flat and minimum part of the S_{11} spectrum larger inductor L_g values need to be used. As L_g increase narrow the S_{11} characteristic trade-off between its value and wanted bandwidth has to be achieved.

Parameter S_{22} is less than –18.9 dB for the frequency band of interest and below –16.45 dB for the whole simulated range. The excellent reverse isolation (S_{12}) below –43 dB is also obtained.

Regarding the gain, maximum S_{21} parameter value of 15.77dB is achieved with 3-dB band from 2.48GHz to 5.4GHz. Size of the amplifying transistor M_1 is chosen to be 230.4 μm for optimum S_{21} and noise performance. Width of M_1 determines the gain at the lower frequencies while width

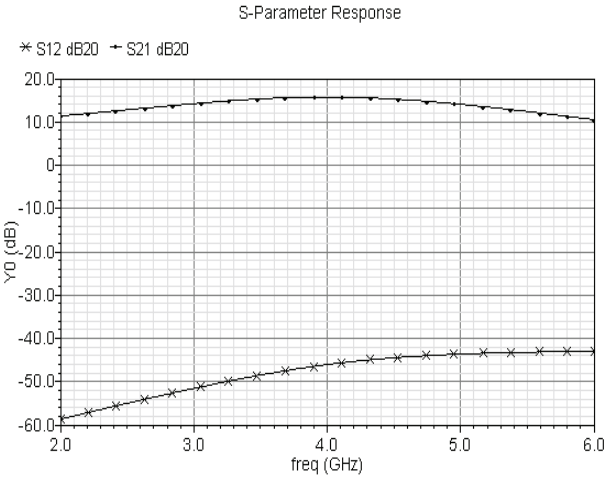


Fig. 3. LNA input-output isolation (S_{12}) and voltage gain (S_{21})

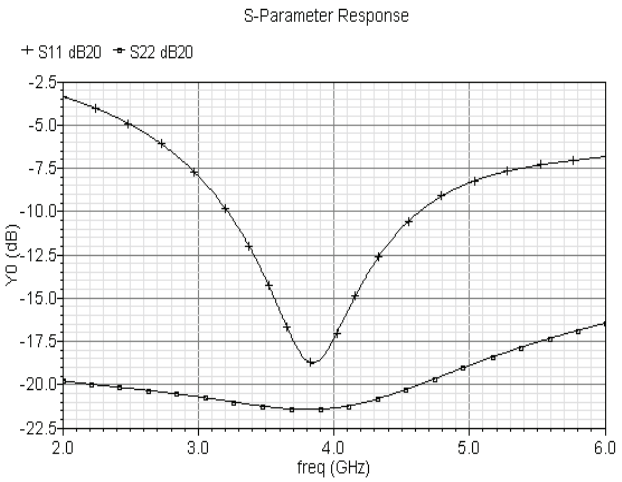


Fig. 2. LNA input (S_{11}) and output (S_{22}) return loss

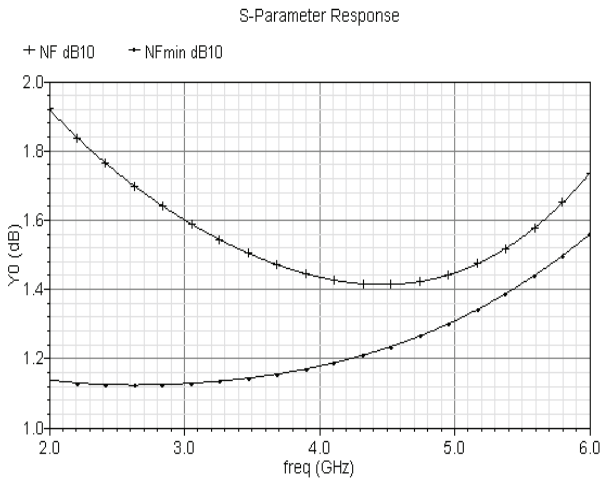


Fig. 4. LNA noise figure (*NF*) and minimum noise figure (NF_{min})

of M_2 (57.6 μm) represents the trade-off between gain and –3dB bandwidth. These values satisfy the gain requirement, while keeping at the same time P_D at reasonable level. Parameter R_f increase results in S_{21} increase, that can also be seen from (7b). For voltage supply V_{DD}=1.2V, a total current including the biasing circuits and output buffer is 3.2mA. The core consumes 1.9mA, while the rest is consumed by the buffer (1mA) and biasing circuits (0.260mA).

NF is flat over the simulated range with the average value of 1.5 dB. Minimum NF value of 1.414 dB is obtained at 4.47GHz. This value is close to NF_{min}, the minimum possible NF value.

Stability parameters simulations were performed from 2GHz to 6GHz. Unconditional stability requirements for the whole simulated range is satisfied. Minimum values for stability factors K_f (Rollet stability factor) and B_{1f} (alternate stability factor) are equal to 12 and 1, respectively, which are values much higher than 1 and 0 [11, 12].

The optimized FOMs of designed LNA and FOMs values of similar LNA topologies in the frequency band of interest [13, 14, 15] are summarized in Table I. In comparison with LNAs FOMs found in literature, design presented in this paper shows higher S_{21}, lower power consumption and *NF*. This result is achieved for simple cascode topology, without additional improvement techniques such as current reuse and cascade. While comparing amplifier design in this work and [13], simulated in the same technology, it can be seen that proposed design have better values for almost all FOMs. The cascade source degenerated LNA design, presented in [14], achieves the best S_{11} value using band-pass filter at a price of large number of reactive elements that causes higher *NF*, power consumption and larger chip area. The same gain value is obtained in [15] where current reuse topology is used. The design [15] drawback is high power consumption and higher *NF* since more circuit elements are needed.

TABLE I
PERFORMANCE COMPARISON

	This work	[13]	[14]	[15]
BW [GHz]	3–5	3–5	3–5	3–5
S_{11} [dB]	<–8.4	<–10	<–10	<–9
S_{22} [dB]	<–18.9	N/A	<–10	<–8.5
S_{21} [dB]	15.77	11	14	14
NF [dB]	1.41	1.8	2	2.4
Power [mW]	3.8	5.2	9	12
Technology	0.13μm	0.13μm	0.18μm	0.18μm

V. CONCLUSION

In this paper low-power approach for ultra-wideband low noise amplifier design in the 3.1–5 GHz band is demonstrated. Using single stage cascode amplifier topology, high gain and low noise can be achieved in the desired frequency range consuming only 3.8mW. Due to the LNA architecture simplicity, input and output is strongly coupled which makes impedance input/output matching more difficult. To obtain required S_{11} parameter value typical source degenerated amplifier is expanded using shunt-feedback resistor. This technique gives adequate input match without introducing large number of additionally components (inductors and capacitors). Obtained simulation results prove that matching method used in this paper together with shunt-load peaking is good solution for UWB LNA design.

ACKNOWLEDGEMENT

This research was supported by the Serbian Ministry of Science and Technological Development (contract TR-11023 and TR-11006) and FP6-INCO project (contract number 043669).

REFERENCES

[1] G. R. Aiello, G. D. Rogerson, “Ultra-Wideband Wireless Systems”, IEEE Microwave Magazine, vol. 4, pp. 36–47, 2003.

[2] P. Heydari, D. Lin, "A Performance Optimized CMOS Distributed LNA for UWB Receivers", CICC'05, Proceedings of the IEEE 2005, pp. 337-340, San Jose, California, USA, 2005.

[3] K. Chen, J. Lu, B. Chen, S. Liu, “An Ultra-Wide-Band 0.4–10-GHz LNA in 0.18-μm CMOS”, IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 54, no. 3, pp. 217-221, 2007.

[4] A. Bevilacqua, A. M. Niknejad, “An Ultrawideband CMOS Low Noise Amplifier for 3.1–10.6-GHz Wireless Receivers”, IEEE Journal of Solid-State Circuits, vol. 39, pp. 2259-2268, 2004.

[5] F. Bruccoleri, E. A. M. Klumperink, B. Nauta, "Noise Canceling in Wideband CMOS LNA’s", ISSCC'02, Digest of Technical Papers, pp. 406–407, San Fransisco CA, USA, 2002.

[6] S. Andersson, C. Svensson, O. Drugge, "Wideband LNA for a Multistandard Wireless Receiver in 0.18 μm CMOS", ESSCIRC'03, Conference Proceedings, pp. 655–658, Estoril, Portugal, 2003.

[7] B. Razavi, *Design of Analog CMOS Integrated Circuits*, New York, McGraw-Hill, 2001.

[8] T. H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*, Cambridge University Press, 1998.

[9] F. Ali, C. Hutchinson, A. Podell, “A Novel Cascode Feedback GaAs MMIC LNA with Transformer-Coupled Output using Multiple Fabrication Process”, IEEE Microwave and Guided Wave Letters, vol. 2, pp. 70–72, 1992.

[10] D. K. Shaeffer, T. H. Lee, “A 1.5-V, 1.5-GHz CMOS Low Noise Amplifier”, IEEE Journal of Solid-State Circuits, vol. 32, no. 5, pp.745-759, 2005.

[11] SpectreRF simulator, Cadence Design System, www.cadence.com

[12] G. Gonzalez, *Microwave Transistor Amplifiers: Analysis and Design*, Prentice Hall, 1984.

[13] A. Mirvakili, M Yavari, "A Linear Wideband CMOS LNA for 3–5 GHz UWB systems", ISOCC'08, Conference Proceedings, pp. 150–153, Busan, Korea, Portugal, 2008.

[14] H.-Jin Lee, Dong Sam Ha, S.S. Choi, "A Systematic Approach to CMOS Low Noise Amplifier Design for Ultrawideband Applications", ISCAS'05, Conference Proceedings, pp. 3962–3965, Kobe, Japan, 2005.

[15] B. Ansari, H. Shamsi, A. Shahhoseini, "Analysis of a 3–5 GHz UWB CMOS Low-Noise Amplifier for Wireless Applications", MWSCAS'09, Conference Proceedings, pp. 979–982, Cancun, Mexico, 2009.

SESSION RCMTA II

Radio Communications, Microwave Technique and Antennas II

Synthesis of Microstrip Filters Using Triangular Open-Loop Resonators

Marin V. Nedelchev[1], Ilia G. Iliev[2]

Abstract: **This paper presents a design of microstrip triangular open loop resonator filter. Because of their structure, the triangular open loop coupled resonator offers wide variety of coupling schemes and different in value and sign coupling coefficient. The coupling schemes are electrical, magnetic and mixed. The paper presents three designs of Chebyshev filter with tapped input/output lines based on triangular open loop resonators.. There is a good agreement between the theoretical and EM simulation results.**

Keywords: coupling coefficient, triangular open loop resonator, cross coupled filters.

I. INTRODUCTION

The fast development of the mobile communication systems stimulates the research of microwave filters with specific symmetrical response. Microstrip filters are preferred for these systems, because of their compact size, low weight, easy integration in integrated circuits, fine adjustment.

Most of the microwave filters are of Chebyshev type. They are equiripple in the passband and maximally flat in the stopband. Such filters can be realized by cascading resonators in series. High filter selectivity requires higher filter order and more resonators. Because of the low unloaded Q factor of the microstrip resonators, the passband loss increases. Alternative way is to use cross-coupled filters with coupling between non-adjacent resonators. Non-adjacent couplings cause transmission zeroes in the stopband or equalization in the group delay.

Among the variety of filter topologies, the classic half-wavelength and hairpin resonator filters are commonly used. Miniaturization is an important requirement for the used resonators. In order to reduce the size of the half wavelength it is possible to fold back the ends of the resonator into a "U" shape. The further miniaturization of half wavelength is achieved by the square open loop filters [1]. The resonator is bent in square form.

The authors of [2] propose to use triangular open loop resonators in order to achieve compact cascaded quadruplet filter with two symmetrical transmission zeros. The geometrical form of triangular resonators allows coupling in different coupling schemes, which exhibits various coupling types. It is possible to form the resonator as an isosceles right-angled, isosceles or equilateral triangle, according to the desired topology. The form of the resonator used in simulations is isosceles right-angled triangle.

[1]Marin Veselinov Nedelchev – Assistant, PhD in Dept. of Radiocommunication and Videotechnology in Faculty of Telecommunication in TU –Sofia E-mail mnedelchev@tu-sofia.bg

[2]Ilia Georgiev Iliev – Assoc. Professor, PhD in Dept. of Radiocommunication and Videotechnology in Faculty of Telecommunication in TU –Sofia E-mail: igiliev@tu-sofia.bg

The synthesis of cross coupled filters is based on the works from Atia and Williams [3], Cameron and Rhodes [4,5] considering waveguide cavity filter design. It is based on the deriving the coupling matrix from the transfer function and its reduction to the corresponding topology form. This technique is found to be useful in the design of microstrip cross-coupled filters. Hong and Lancaster proposed in several papers [6,7] numerical method for cross coupled filter design based on approximation of the low pass filter prototype elements.

This paper presents a design procedure of all pole Chebyshev type and cross-coupled filters based on isosceles right angled triangular resonators. A full wave EM simulator is used for obtaining the coupling coefficients between the resonators. Numerical results for the values of the coupling coefficients are presented. Three design examples are carried out. The frequency responses from full wave EM simulator are presented in order to validate the procedure.

II. TRIANGULAR OPEN LOOP RESONATOR COUPLING STRUCTURES

The triangular open loop resonator is half wavelength long for the central frequency of the filter.

The coupling mechanism is based on the fringe fields of closely situated resonators. The nature of coupling depends on the resonator configuration. It is clear that half wavelength triangular resonator is symmetrical along the center of the hypotenuse. The coupling coefficient for synchronously tuned resonators can be calculated easily by the resonance frequencies of even and odd mode [1]:

$$k = \frac{f_e^2 - f_o^2}{f_e^2 + f_o^2} \qquad (1)$$

A full wave EM simulator based on the Method of the Moments (MoM) is used to identify the resonance frequencies. Most of the coupling structures are simulated using the symmetry in their topology for electrical and magnetic wall introducing in-between. When symmetry does not present, the whole structure is simulated. This does not constrain in any way the obtained results. The simulation are carried out for standard FR-4 substrate with $\varepsilon_r = 4.4$, $h = 1.5mm$, $tg\delta = 0.02$.

The coupling structures are formed by different orientation in the plane of identical resonators. They should be closely situated and separated by distance ***d.***

Both resonators in the coupling structure may have or not an offset in the direction of the coupled lines- denotes as ***delta*** (Fig.1b). This offset is made for two major reasons- to achieve loose coupling or to make possible connection of another resonator.

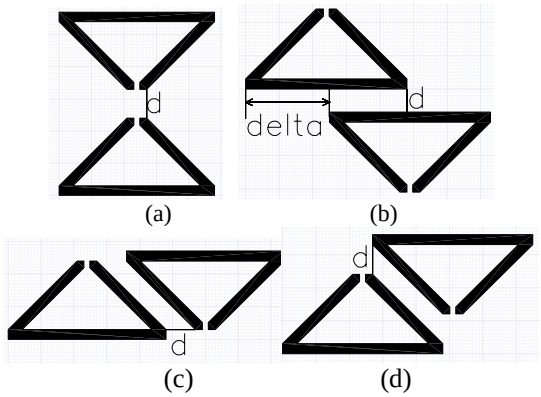


Fig.1 Topologies of triangular open loop resonators (a) electrical coupling, (b) magnetic coupling, (c), (d) and (e) mixed coupling

Figure 2 shows the topology of the microstrip triangular open loop resonator with the following dimensions: strip width $w = 2.77mm$, corresponding to 50Ω line on FR-4, hypotenuse length $c = 45mm$, cathetus length $a = 29mm$, gap between the cathetuses $s = 2mm$.

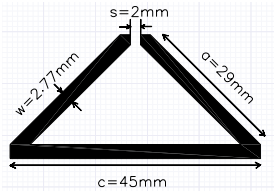


Fig.2 Triangular open loop resonator used in the simulations tuned on 811MHz

The first resonance frequency of the resonator is 811MHz. A number of full wave EM simulations are carried out in order to obtain the resonance frequencies in the frequency response. Following the above mentioned considerations, an electrical coupling presents in the structure shown on Fig.1a. Both resonators are close each other with their open ends. The coupling coefficient is calculated according to Eq.(1).

The numerical results for the coupling coefficient are presented on Fig.3. The coupling between the non-adjacent resonators should be out-of-phase the other couplings. In the case of triangular coupled resonators, the coupling is not strong enough. This leads to transmission zeros in the frequency response away from the passband.

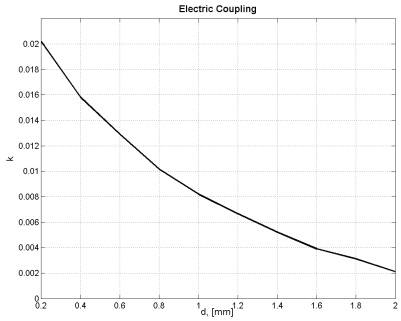


Fig.3 Coupling coefficient for electrically coupled resonators

Magnetic nature of coupling presents in the coupling structure shown on Fig.1b, when the offset distance ***delta*** is zero. Both resonators are coupled through their hypotenuses, where short circuit is present for the resonance frequency.

The magnetic field at this point is predominant over the electric. The results for the coupling coefficient for different offsets ***delta*** are shown on Fig.4.

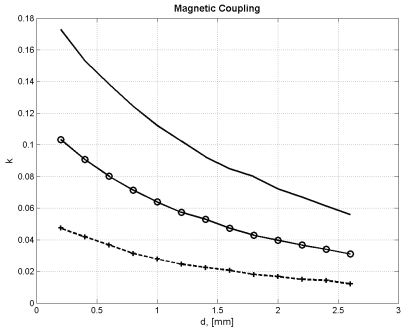


Fig.4. Coupling coefficient for magnetic coupling. Solid line- $delta = 0$ (no offset), circles- $delta = 0.5c = 22.5mm$, dashed- $delta = 0.25c = 11.25mm$.

More flexible form of predominant magnetic coupling is when an offset is presented in the structure (Fig.1b). The coupling is not purely magnetic in nature because of the offset. When *delta* is one half of the hypotenuses, the one more side (the other side of the hypotenuses) is present for additional coupling. In this way the coupling coefficient lowers its values, but remains the monotonically decreasing behavior (Fig.4 circles). When the offset *delta* is a quarter of the hypotenuses, the coupling becomes weaker (Fig.4 dashed). The character of the coupling is not longer magnetic, but mixed.

Mixed coupling presents, when both triangular resonators are close with their cathetuses (Fig.1c,d).

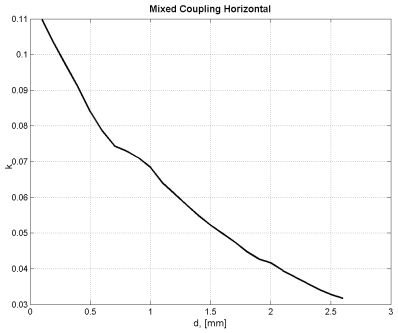


Fig.5 Mixed coupling with spacing in horizontal direction

The coupling resonators are opposite situated in the plane. The currents in the coupled lines are equal in amplitude and in-phase. The value of the coupling coefficient is with positive sign. The approach chosen in this paper for obtaining the coupling coefficient is to fix one of the resonators and to move the other one either in horizontal (Fig.1c), or in vertical (Fig.1d) direction. When horizontal spacing is introduced, the graphical results are presented on Fig.5. The graphical results for vertical spacing between the resonators (Fig.1d) are presented on Fig.6. The coupling coefficient has the same behavior as the coupling coefficient for horizontal spacing.

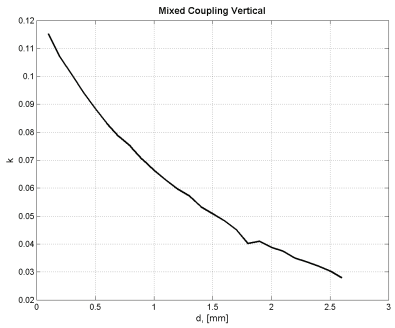


Fig.6 Mixed coupling with spacing in vertical direction

In order to determine the exact tapping position of the input/output lines, it is necessary to derive the external quality factor Q_e. The mathematical expression for it is [1]:

$$Q_e = \frac{f_0}{\Delta f_{\pm\frac{\pi}{2}}} \quad (2),$$

where f_0 is the resonance frequency of the resonator, and $\Delta f_{\pm\frac{\pi}{2}}$ is the bandwidth at which the phase of the s_{11} shifts to $\pm\frac{\pi}{2}, rad$ with respect to the phase at the resonance frequency. Using Ansoft Designer SV, it is extracted the external quality factor. The results are shown on Fig.7.

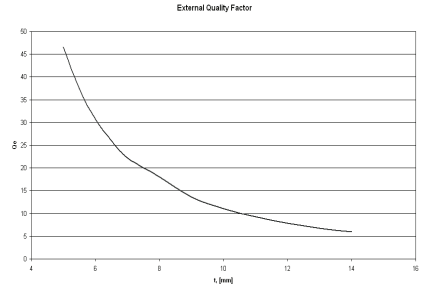


Fig.7 Coupling coefficient versus the distance between the resonators.

The external quality factor decreases with the distance from the center of the resonator. For narrow bandwidth filters Q_e is bigger and he tapping position is closer to the center of the resonator. This leads to bigger sensitivity of the filter's response against the tapping position. Consequently the passband reflection coefficient will degrade.

III. DESIGN EXAMPLES

Three microstrip filters are synthesized and simulated in order to prove the design procedure. The design procedure starts with computation of the elements of the coupling matrix corresponding to the filter topology. The next step is to define the coupling structures for realizing the coupling coefficients. From the presented figures (3-6), the distance between the resonators is found. For magnetic coupling the offset should be previously known. The tapping position can be found by Fig.7.

III.A. Third order Chebyshev microstrip filter

The filter specification is as follows:

Order: *3*; **Approximation:** *Chebyshev type*; **Center frequency:** $f_0 = 810MHz$**;** **Bandwidth:** $\Delta f = 70MHz$**;** **Ripple:** $0.1dB$

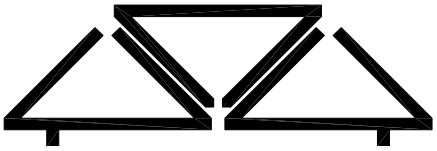


Fig.7 Third order Chebyshev filter topology

The coupling matrix elements are derived from the Chebyshev approximation. They are as follows: $M_{12} = M_{23} = 0.089$ and theexternal quality factor is $Q_e = 12.53$. The coupling is realized by mixed coupling with vertical spacing. The spacing between the resonators is found by means of Fig.6 $s = 0.32mm$. The tap position is $t = 9.2mm$ from the center of the resonator. The simulated frequency response is shown on Fig.8.

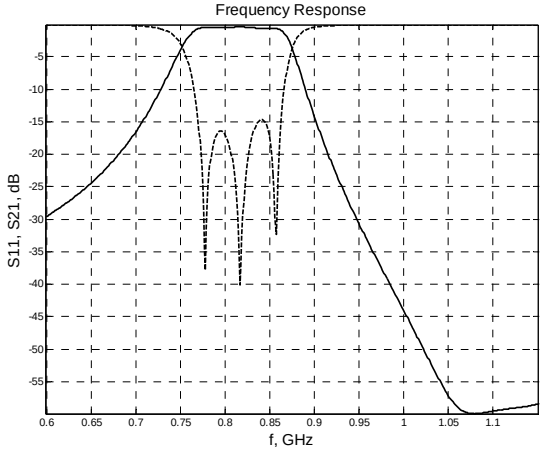


Fig.8. Frequency response of third order microstrip filter. S_{21} - solid line, S_{11} - dashed line

As it is clearly seen the transmission coefficient follows exactly the equiripple response of the Chebyshev approximation. In the transmission coefficient is seen a transmission zero on $1070MHz$ because of the parasitic coupling between the first and the third resonator. It is positive in sign and the transmission zero is above the passband. The realized bandwidth is 80MHz and the maximum level of the reflection coefficient is 15dB. The passband loss is less than 2dB, mainly due to the high dielectric loss of the FR-4 substrate.

III.B. Forth order Chebyshev microstrip filter

The filter specification is as follows:

Center frequency: $f_0 = 810MHz$**;** **Bandwidth:** $\Delta f = 80MHz$**;** **Ripple:** $0.1dB$

The topology of the filter is shown on Fig. 9.

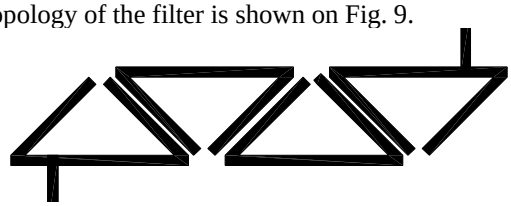


Fig.9. Forth order mictrostrip triangular filter

The coupling coefficients are found to be $M_{12} = M_{34} = 0.091$, $M_{23} = 0.0691$. The external quality

factor is $Q_e = 10.48$. The couplings are realized with mixed coupled structure with horizontal spacing. The spacing is found by means of Fig.5. The spacing between the resonators realizing the M_{12} and M_{34} is $s_{12} = s_{34} = 0.41mm$ and the spacing for M_{23} is $s_{23} = 0.8mm$. The tap position is $t = 10.4mm$. The results of the full wave EM simulation are shown on Fig.10.

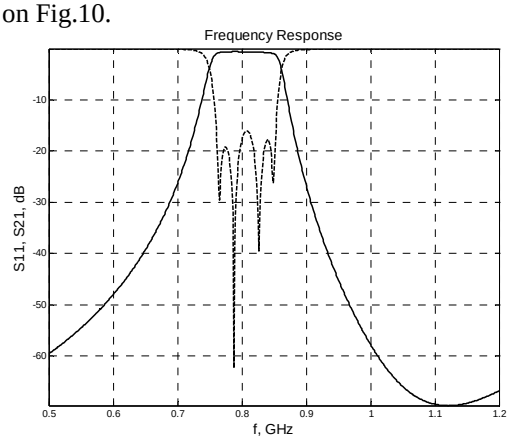


Fig.10. Frequency response of forth order microstrip filter. S_{21} - solid line, S_{11} - dashed line

The realized passband is 90MHz, due to the higher strength of the couplings. The insertion loss is less than 1.5dB, due to the high dielectric loss of the FR-4 substrate and the conductor loss. In the S_{11}, it is seen the Chebyshev type response with four easily distinguished resonances. The maximum reflection coefficient is 16dB.

III.C. Forth order cross-coupled Chebyshev microstrip filter

The filter specification is as follows:

Center frequency: $f_0 = 810MHz$**; Bandwidth:** $\Delta f = 80MHz$**; Ripple:** $0.1dB$ **Frequencies of symmetric transmission zeros:** $f_0 = 670MHz$ **and** $f_0 = 950MHz$**.**

The topology of the simulated filter is shown on Fig.11.

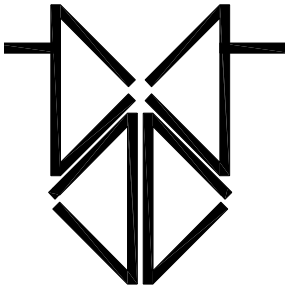


Fig.11. Forth ordercross coupled mictrostrip triangular filter

The coupling coefficients from the approximation are: $M_{12} = M_{34} = 0.089$, $M_{23} = 0.0707$, $M_{14} = -0.004$. The couplings M_{12} and M_{34} are realized with horizontal spaced coupled resonators and the spacing is $s_{12} = s_{34} = 0.38mm$ (Fig.5). The coupling M_{23} is realized by magnetic coupling with no offset. The spacing is found by Fig.4 $s_{23} = 2mm$. The negative coupling coefficient M_{41} is realized by electrically coupled resonators. The distance between the electrically coupled resonators is $s_{14} = 1.6mm$. The frequency response is shown on Fig.12.

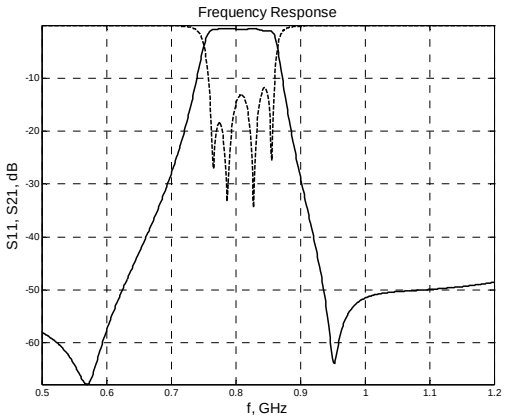


Fig.11 Frequency response of forth order cross coupled microstrip filter. S_{21} -solid line, S_{11} - dashed line

The realized passband is 90MHz and the insertion loss is lower than 1.5dB. The transmission zeros are not symmetrical, the upper one is placed at 950MHz, but the lower one is at 580MHz. This is because of the frequency dependence of the electrical coupling. The maximum value of the reflection coefficient in the passband is better than 13dB.

V. CONCLUSION

This paper presents a research of the microstrip filters based on triangular open loop coupled resonators. The form of the resonator used is isosceles right-angled triangle. There are described the four basic coupling topologies with electric, magnetic and mixed coupling. In order to characterize the coupling coefficient, a full wave EM simulator is used for obtaining the resonance frequencies when introducing electrical and magnetic walls. Three design examples are synthesized and simulated in order to verify the procedure. The frequency responses from full wave EM simulator are presented in order to validate the procedure.

REFERENCES

[1] Hong J.-S. , M.J. Lancaster. "Couplings of Microstrip Square Open-Loop Resonators for Cross-Coupled Planar Microwave Filters." 1996 Transactions on Microwave Theory and Techniques 44.11 (Nov. 1996 [T-MTT]): 2099-2109.

[2] A. GÖRÜR, and C. KARPUZ, “Cross-coupled bandpass filter using microstrip triangular open-loop resonators“, Proceedings of 31th The European Microwave Conference, CD-ROM, Sep 24-28, 2001, Excel, London, ENGLAND

[3] A.E. Atia and A.E. Williams. "Narrow-Bandpass Waveguide Filters." 1972 Trans. on Microwave Theory and Techniques 20.4 (Apr. 1972 [T-MTT]): 258-265

[4] Cameron, Rhodes, J.D., Asymmetric Realization for Dual-Mode Bandpass Filters, IEEE Trans on MTT 1981 Jan., pp.51-58

[5] Cameron, R., Advanced Coupling Matrix Synthesis Techniques for Microwave Filters, IEEE Trans on MTT-50, Jan.2003, pp.1-10

[6] J.S.Hong, Lancaster, M, Theory and Experiment of Novel Microstrip Slow-Wave Open Loop Resonator Filters, IEEE Trans. On MTT-45, Dec.1997, pp.2358-2365

[7] J.S.Hong, Lancaster, M, Cross-Coupled Microstrip Hairpin-Resonator Filters, IEEE Trans on MTT-46, Jan.1998, pp.118-122

Optimal Multi-Antenna Reception in Presence of Three-Dimensional Directional Multipath

Dimitar G. Valchev

***Abstract* – In this paper the directional aspects of a spatial diversity reception in multipath wireless channels are analysed. The analysis implies that there is a three-dimensional orientation of the receiver multi-antenna array at which the highest possible signal-to-noise ratio (SNR) gain is achieved for a given number of antennas and a given distance between two adjacent antennas.**

***Keywords* – Fading channels, Multipath channels.**

I. INTRODUCTION

A very important phenomenon in wireless communications is the small-scale (or multipath) fading channel formed by the interference of numerous waves incident to the receiver in a local volume. The fading causes a fluctuation in the received signal-to-noise ratio (SNR). This degradation is usually overcome by using a spatial diversity system at the receiver [1]. Studies show [2] that capacity increases linearly with the increase of the number of receive branches provided that the fading at the different branches is uncorrelated. However, in realistic channels there is a non-zero cross-correlation between two receiver branches placed at a limited distance. If the fading at two branches is highly correlated, a less diversity gain is achieved. Therefore it is important to study the impact on the resulting SNR using receiver branches with non-zero cross-correlation of the received faded signal envelope.

The multipath propagation channel has a generally non-uniform multipath angular power density (APD) function at the receiver due to the non-isotropic scattering within the wireless environment. In such channels the multipath is characterized by its *directivity*. The directional channel determines different width of the signal cross-correlation function between two branches in a multi-antenna system in different directions [3, 4]. Therefore, for achieving a high SNR gain by the multi-antenna system, while keeping a minimum distance between the antennas, the axis of the antenna array should be aligned with the direction with the most narrow cross-correlation function determined by the particular APD. Thus, two more degrees of freedom are introduced in designing spatial dicersity reception systems – the *azimuth* and *elevation* of the receiver array axis.

Dimitar G. Valchev is with the Department of Computer Systems and Technologies, Asen Zlatarov University, Burgas 8010, Bulgaria, E-mail: dvalchev@ece.neu.edu

This work is supported by Asen Zlatarov University, Faculty of Technical Sciences, project # 213/2010.

II. APD AND RECEIVER ARRAY TOPOLOGY

The spatial diversity reception is achieved by using a multi-antenna array. For the present analysis a uniform linear array is assumed with uniform angular gain for each antenna with polarization mismatch being neglected. The distance between any two adjacent antennas in the array is d. The axis of the array has an azimuth θ and an elevation ϕ.

The APD function around the receiver may be estimated by using the same multi-antenna array, configured in a direction-of-arrival estimation mode. It may be rotated in the horizontal plane and in a vertical plane for each azimuth. Alternatively, the APD function may be estimated using other array topologies [5]. Once created, the APD function $p(\theta,\phi)$ shows the power that comes to the receiver at a particular azimuth-elevation pair. The geometrical configuration of the receiver multi-antenna array is shown in Fig. 1. The optimal direction of the array axis is the direction of the most narrow cross-correlation function of the received signal given the APD around the receiver as shown in the next section.

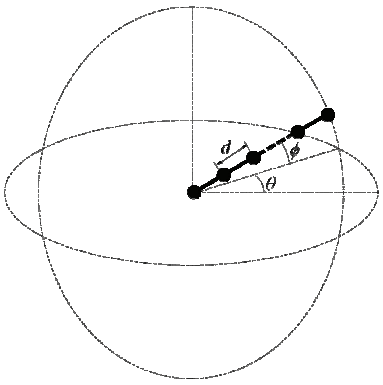


Fig. 1. Multiantenna receiver array geometry

III. SPATIAL CROSS-CORRELATION OF THE RECEIVED SIGNAL ENVELOPE

For a three-dimensional propagation the following approximate and mathematically tractable cross-correlation function has been derived in [4]:

$$\rho(d,\theta,\phi) \approx e^{-15.33\sigma(\theta,\phi)\left(\frac{d}{\lambda}\right)^2} \quad (1)$$

where the quantity $\sigma(\theta,\phi)$ is termed *fading rate variance* of the received signal. The increase of the fading rate variance as a measure of the fluctuation rate leads to a decreased width of

the fading signal cross-correlation function. The fading rate variance can be expressed as [4]

$$\sigma(\theta,\phi)=\mathrm{Y}^2\left[\begin{array}{l}\dfrac{2}{3}+\xi\left(2\sin^2\phi-\dfrac{2}{3}\right)\\+\chi\sin 2\phi\cos\left(\theta-\theta_{\phi 45^\circ}^{\max}\right)\\+\zeta\cos^2\cos\left(\theta-\theta_{\phi 0}^{\max}\right)\end{array}\right] \quad (2)$$

where the parameters Y, ξ, χ, ζ, $\theta_{\phi 45^\circ}^{\max}$ and $\theta_{\phi 0}^{\max}$ are termed *multipath shape factors* [4] defined in terms of the multipath APD around the mobile receiver $p(\theta,\phi)$ and its spherical harmonic coefficients S_l^m of *l*-th degree and *m*-th order [6]:

Angular Spread:

$$\mathrm{Y}=\sqrt{1-\frac{S_1^{0^2}+\left|S_1^1\right|^2}{S_0^{0^2}}};\quad 0\le \mathrm{Y}\le 1 \quad (3)$$

Elevational Constriction:

$$\xi=\frac{1.5S_2^0S_0^0-\left(S_1^{0^2}-0.5\left|S_1^1\right|^2\right)}{S_0^{0^2}-S_1^{0^2}-\left|S_1^1\right|^2};\ -0.5\le\xi\le 1 \quad (4)$$

45˚-*Inclined Constriction*:

$$\xi=\frac{2\left|S_2^1S_0^0-S_1^0S_1^1\right|^2}{S_0^{0^2}-S_1^{0^2}-\left|S_1^1\right|^2};\quad 0\le\chi\le 1 \quad (5)$$

Azimuthal Constriction:

$$\zeta=\frac{2\left|S_2^2S_0^0-S_1^{1^2}\right|}{S_0^{0^2}-S_1^{0^2}-\left|S_1^1\right|^2};\quad 0\le\zeta\le 1 \quad (6)$$

Azimuth of Maximum Fading at 45˚ *Elevation*:

$$\theta_{\phi 45^\circ}^{\max}=\arg\left\{S_2^1S_0^0-S_1^0S_1^1\right\};\quad 0\le\theta_{\phi 45^\circ}^{\max}\le 2\pi \quad (7)$$

Azimuth of Maximum Fading at Zero Elevation:

$$\theta_{\phi 0}^{\max}=\arg\left\{S_2^2S_0^0-S_1^{1^2}\right\};\quad 0\le\theta_{\phi 0}^{\max}\le\pi \quad (8)$$

The expressions for the necessary particular spherical harmonic coefficients S_l^m are

$$S_0^0=\int_0^{2\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}p(\theta,\phi)\cos\phi\,d\phi\,d\theta, \quad (9)$$

$$S_1^0=\int_0^{2\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}p(\theta,\phi)\sin\phi\cos\phi\,d\phi\,d\theta, \quad (10)$$

$$S_1^1=\int_0^{2\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}p(\theta,\phi)\cos\phi e^{j\theta}\cos\phi\,d\phi\,d\theta \quad (11)$$

$$S_2^0=\int_0^{2\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}p(\theta,\phi)\left(\sin^2\phi-\frac{1}{3}\right)\cos\phi\,d\phi\,d\theta \quad (12)$$

$$S_2^1=\int_0^{2\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}p(\theta,\phi)\cos\phi\sin\phi e^{j\theta}\cos\phi\,d\phi\,d\theta \quad (13)$$

$$S_2^2=\int_0^{2\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}p(\theta,\phi)\cos^2\phi e^{j2\theta}\cos\phi\,d\phi\,d\theta \quad (14)$$

It is seen that there is an explicit angular dependence in the expression for the autocorrelation of the received fading signal envelope in (1). This determines an explicit angular dependence for the SNR gain.

For the angular spread Y=0 there is only one path incoming to the mobile receiver, the expression in (1) reduces to a constant one, implying no capacity gain when using a multi-antenna system compared to a single antenna. For a positive angular spread the correlation and hence the SNR gain depends on the angular orientation of the multi-element antenna. The combined action of the multipath shape factors determines an azimuth-elevation three-dimensional direction of maximum fading rate variance. This is the direction of the most narrow cross-correlation function given by (1). Therefore, the multi-antenna array should be aligned with this direction in order to achieve maximum SNR gain.

IV. Conclusion

This paper addresses the spatial dependence of the SNR gain achieved by a multi-antenna receiver in multipath fading channels. The results show that as the multipath power incoming to the receiver becomes more concentrated around a single direction, less SNR gain is obtained when using receive diversity techniques. For a single faded path the receiver spatial diversity is useless. For a non-zero angular spread of the received multipath power, and for a desired SNR gain, the receiver elements spacing is determined by the pair wise spatial cross-correlation functions of the received signal envelope which depend on the angular distribution of the multipath power, and which influence the resulting received SNR in the spatially diversified channel.

References

[1] W. C. Jakes, ed., *Microwave Mobile Communications*, New York, IEEE Press, 1974.

[2] E. Telatar, "Capacity of multi-antenna Gaussian channels", AT&T Bell Labs, 1995. Available online: http://mars.bell-labs.com/cm/ms/what/mars/papers/proof.

[3] D. G. Valchev and D. Brady, "Multpath directivity and spatial selectivity in three-dimensional wireless channels", *IEEE Trans. Antennas Propag.*, vol. 57(7), pp. 2147–2154, 2009.

[4] D. G. Valchev and D. Brady, "Three-dimensional multpath shape factors for spatial modeling of wireless channels", *IEEE Trans. Wireless Commun.*, vol. 8(11), pp. 5542–5551, 2009.

[5] J. Fuhl, J.-P. Rossi, and E. Bonek, "High-resolution 3-D direction-ofarrival determination for urban mobile radio," *IEEE Trans. Antennas Propagat.*, vol. 45, pp. 672–682, Apr. 1997.

[6] W. W. Bell, *Special Functions for Engineers*, Dover, 2004.

Secret Keys Based on Multipath Directivity in Wireless Channels

Dimitar G. Valchev

***Abstract* – This paper proposes a new method for determining the secret key for encrypting the information exchanged through the multipath channel between two wireless communicating nodes. Based on the multipath structure of the wireless environment, this method is site-specific, making it stable against various kinds of systematic attacks.**

***Keywords* – Cryptography, multipath channels.**

I. INTRODUCTION

The shared medium of the wireless channel poses a challenge in implementing secure communications. The problem is addressed mainly by various cryptographic methods usually performed at the application layer of wireless networking. The so implemented cryptographic algorithms are vulnerable to attacks similarly to every kind of encrypting algorithm. A usual way to achieve resistance to attacks is to regularly change the algorithm according to some predetermined sequence. Another way is to exploit the physical and geometrical parameters of the multipath wireless channel to determine the secret key shared between the communicating nodes [1, 2]. This can be used especially in multiple-input multiple-output (MIMO) communications where the transmission parameters of the distinct propagation paths may be used to determine the secret keys used for the encrypted signal through them.

To ensure unique symmetric secret key through a given propagation path, the channel has to be reciprocal, meaning that the propagation path between two wireless link ends is one and the same in both directions of transmission. The reciprocal multipath wireless channel having *directivity* at both link ends is characterized by a set of spatial parameters [4] that can be used to determine the parameters of the encryption algorithm. Those spatial parameters are derived from the bidirectional joint angular power density (APD) function showing how much multipath power is exchanged at a particular azimuth pair at both wireless link ends. The joint APD depends on the scattering in the wireless channel. Thus, the encryption would be based on a particular placement of the scattering objects within the environment of the wireless networks. The spatial parameters determining the secret key can be properly quantized to map a particular secret key to a quantization level for the spatial parameter.

Dimitar G. Valchev is with the Department of Computer Systems and Technologies, Asen Zlatarov University, Burgas 8010, Bulgaria, E-mail: dvalchev@ece.neu.edu

This work is supported by Asen Zlatarov University, Faculty of Technical Sciences, project # 213/2010.

This approach may be used in both infrastructure-based and ad hoc wireless networks. In the infrastructure-based wireless networks a reciprocal channel is formed between the access point and each wireless user to which the corresponding channel is assigned. In ad hoc wireless networks there is a reciprocal channel between each pair of communicating nodes. In both cases the spatial parameters of the multipath channel may determine the secret key used to encrypt the information between the two communicating devices.

II. JOINT MULTIPATH SHAPE FACTORS

In [4] the concept of spatial parameters termed joint multipath shape factors has been developed for modeling wireless channels characterized by multipath at both link ends. Those spatial parameters are extensions to the multipath shape factors developed in an earlier work for a single wireless node [3]. The joint multipath shape factors are derived in terms of the double Fourier coefficients of the joint multipath APD at both the communicating nodes. A particular interest is paid to the following joint multipath shape factors:

Positive joint angular constriction between link ends a and b:

$$\Gamma_{+} = \frac{2\left|F_{1,-1} - F_{1,0}F_{0,-1}\right|}{2F_{0,0}^2 - F_{1,0}^2 - F_{0,1}^2} \tag{1}$$

Negative joint angular constriction between link ends a and b:

$$\Gamma_{+} = \frac{2\left|F_{1,1} - F_{1,0}F_{0,1}\right|}{2F_{0,0}^2 - F_{1,0}^2 - F_{0,1}^2} \tag{2}$$

with $F_{m,n}$ being the m^{th} order, n^{th} order double Fourier coefficient of the joint APD $p(\theta_a, \theta_b)$ given by

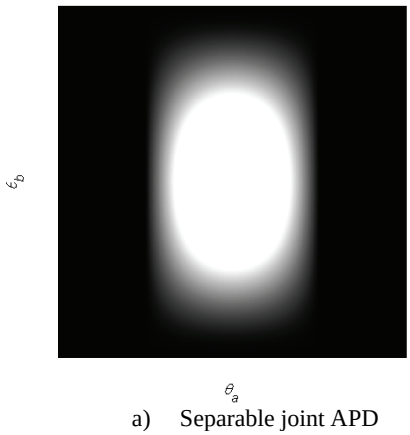
$$F_{m,n} = \int_0^{2\pi}\int_0^{2\pi} p(\theta_a, \theta_b) e^{j(m\theta_a + n\theta_b)} d\theta_a d\theta_b \,. \tag{3}$$

with θ_a and θ_b being the azimuths at link ends *a* and *b*, respectively. The joint APD function may be estimated by channel sounding techniques between the *a* and *b* link ends before the actual communication process begins. Then the corresponding spatial parameters in (1)–(2) are derived for determining the secret key.

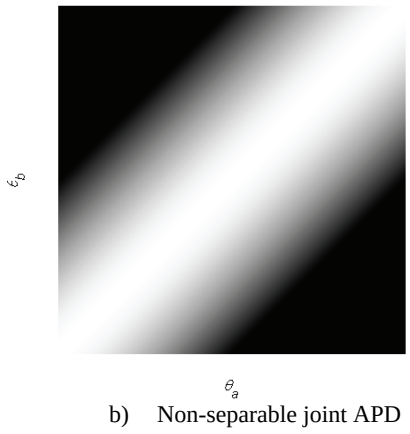
The positive and negative joint angular constrictions between the *a* and *b* link ends, Γ_+ and Γ_-, both range from 0 to 1 and are measures of the angular separability of the joint APD. Indeed, if the joint APD is separable, that is, $p(\theta_a,\theta_b)=p(\theta_a)p(\theta_b)$, then after substitution in (3) it can be easily shown that both (1) and (2) become zero. Note that the terms positive and negative a relative, depending on the chosen orientation at each link end. For $\Gamma_+=\Gamma_-=0$, the joint APD is separable in the azimuth angles at the two link ends,

or $p(\theta_a,\theta_b) = p(\theta_a)p(\theta_b)$. The positive joint angular constriction Γ_+ is equal to one when the joint APD is non-zero only on the line $\theta_a - \theta_b = \theta_0^+$ for some $0 \le \theta_0^+ \le 2\pi$. The negative joint angular constriction Γ_- is equal to one when the joint APD is non-zero only on the line $\theta_a - \theta_b = \theta_0^-$ for some $0 \le \theta_0^- \le 2\pi$.

The so defined joint multipath shape factors have certain significance in expressing the second-order fading statistics in mobile-to-mobile wireless channels [4]. Here, these multipath



a) Separable joint APD



b) Non-separable joint APD

Fig. 2. Joint APD between link ends *a* and *b*

shape factors are used only to determine the secret keys for the encryption algorithm.

Two possible joint APDs are shown in Fig. 1 with the darker regions corresponding to lower power and the brighter regions corresponding to higher power at a given angular pair θ_a, θ_b. Fig.1a shows a separable joint APD and Fig.1b shows a non-separable joint APD. The separable joint APD determines zero joint angular constrictions while the non-separable joint APD determines non-zero joint angular constrictions. Thus, the joint APD with its spatial parameters given by (1) and (2) determines the secret key to be used for encryption.

III. Secret Key Determination

The joint multipath shape factors are suitable for determining a symmetric secret key generation because they are directly associated with the reciprocity of the channel. The single node multipath shape factors are characteristic only to the node for which they are derived and therefore are not appropriate for determining a symmetric secret key to be used to transmit information to the other node.

It is possible to derive the secret key using the multipath shape factors at a single node only [3] but then the secret key would no longer be symmetric, since the multipath structure at the one node is generally different than the multipath structure at the other node. However, such an approach would allow for a larger variety of methods for determining the secret key since more spatial parameters would be used either alone or in a combination. In an infrastructure-based wireless network the information about the multipath structure at each node could be sent from the node to the access point so the controller at the access point determines the secret key to be used for that wireless node. In an ad hoc wireless network the information about the multipath structure at each node is exchanged between the communicating devices during a handshaking phase of the communication initialization process. Taking this idea further, the secret keys can be based on the three-dimensional multipath shape factors at each link end [5] which would give more possibilities for combining the increased number of spatial parameters and hence for making the encrypting algorithms less vulnerable to various attacks.

IV. Conclusion and Future Work

This paper proposes a new method for generating secret keys for encrypting the information transmitted between two communicating nodes. The method is based on the joint multipath directivity characterizing the wireless channel between the two devices. There are two joint spatial parameters, termed joint angular constrictions, which characterize the joint multipath directivity. This gives freedom to choose various approaches for determining the corresponding secret key for communication between the two nodes. If the placement of the communicating nodes or the scattering objects within the wireless environment changes, the secret key would also be changed through the new spatial parameters. This further stabilizes the method against attacks. Future work will be focused on designing non-symmetric secret keys based on the single multipath shape factors at each wireless communicating node.

References

[1] A. Sayeed and A. Perrig, “Secure wireless communications: secret keys through multipath”, ICASSP 2008, 2008.

[2] T.-H. Chou, A. Sayeed and S. Draper, “Impact of channel sparsity and correlated eavesdropping on secret key generation from multipath channel randomness”, ISIT 2010, 2010.

[3] G. D. Durgin and T. S. Rappaport, “Theory of multipath shape factors for small-scale fading wireless channels”, *IEEE Trans. Antennas Propag.*, vol. 48(5), pp. 682–693, 2000.

[4] D. G. Valchev, "Spatial modeling of three-dimensional multipath wireless channels", PhD dissertation, Boston, 2008.

[5] D. G. Valchev and D. Brady, "Three-dimensional multipath shape factors for spatial modeling of wireless channels", *IEEE Trans. Wireless Commun.*, vol. 8(11), pp. 5542–5551, 2009.

Comparative Performance Studies of Equipments in Laboratory Wi-Fi IEEE 802.11a Point-to-Point Links

José A. R. Pacheco de Carvalho[1], Hugo Veiga[2], Nuno Marques[3]
Cláudia F. F. P. Ribeiro Pacheco[4], António D. Reis[5]

***Abstract* –Wi-Fi IEEE 802.11a has been receiving increasing interest, as it works in the 5 GHz band and the growing use of 2.4 GHz band is leading to higher interferences. Performance is a very important issue, resulting in more reliable and efficient communications. Laboratory measurements are made about several performance aspects of Wi-Fi IEEE 802.11a open point-to-point links. A contribution is given to performance evaluation of this technology, using two types of access points from Enterasys Networks (RBTR2 and RBT-4102). Detailed results are presented and discussed, namely at OSI levels 4 and 7, from TCP, UDP and FTP experiments, permitting measurements of TCP throughput, jitter, percentage datagram loss and FTP transfer rate. Conclusions are drawn about the comparative performance of the links.**

Keywords – **WLAN, Wi-Fi, Point-to-Point Links, IEEE 802.11a, Wireless Network Laboratory Performance.**

I. INTRODUCTION

Wireless communications are increasingly important for their versatility, mobility and favorable prices. It is the case of microwave based technologies, e.g. Wi-Fi. The importance and utilization of Wi-Fi have been growing for complementing traditional wired networks. Wi-Fi has been used both in ad hoc mode and infrastructure mode. In this case an access point, AP, is used to permit communications of Wi-Fi devices with a wired based LAN through a switch/router. In this way a WLAN, based on the AP, is formed. Wi-Fi has penetrated the personal home, forming a WPAN, allowing personal electronic devices to communicate. Point-to-point and point-to-multipoint configurations are used both indoors and outdoors, requiring specific directional and omnidirectional antennas. Wi-Fi uses microwaves in the 2.4 and 5 GHz frequency bands and IEEE 802.11a, 802.11b and 802.11g standards [1]. As the 2.4 GHz band becomes increasingly used and interferences increase, the 5 GHz band has received considerable interest, although absorption increases and ranges are shorter.

Nominal transfer rates up to 11 (802.11b) and 54 Mbps (802.11 a, g) are permitted. CSMA/CA is the medium access control. Wireless communications, wave propagation [2,3] and WLAN practical implementations [4] have been studied. Detailed information is available about the 802.11 architecture, including performance analysis of the effective transfer rate. An optimum factor of 0.42 was presented for 11 Mbps point-to-point links [5]. Wi-Fi (802.11b) performance measurements are available for crowded indoor environments [6]. Performance has been a very important issue, resulting in more reliable and efficient communications. Telematic applications have specific performance requirements, depending on application. New telematic applications present special sensitivities to performances, when compared to traditional applications. Application characterization and requirements have been discussed [7]. Requirements have been quoted such as: for video on demand/moving images, 1-10 ms jitter and 1-10 Mbps throughput; for Hi Fi stereo audio, jitter less than 1 ms and 0.1-1 Mbps throughputs.

Several performance measurements have been made for 2.4 and 5 GHz Wi-Fi [8,10], as well as high speed FSO [11]. In the present work further Wi-Fi (IEEE 802.11 a) results arise, through OSI levels 4 and 7. Performance is evaluated in laboratory measurements of open point-to-point links using different types of access point equipments.

The rest of the paper is structured as follows: Chapter II presents the experimental details i.e. the measurement setup and procedure. Results and discussion are presented in Chapter III. Conclusions are drawn in Chapter IV.

II. EXPERIMENTAL DETAILS

Two types of experiments were carried out, which are referred as Expa and Expb. The measurements of Expa used (Fig. 1-(A), (C)) Enterasys RoamAbout RBTR2 level 2/3/4 access points (mentioned as APa), equipped with 15 dBm IEEE 802.11 a/b/g cards [12], and 100-Base-TX/10-Base-T Allied Telesis AT-8000S/16 level 2 switches [13]. The access points had RBTBH-R2W radio cards similar to the Agere-Systems model 0118 type, and firmware version 6.08.03. The configuration was for minimum transmitted power i.e. micro cell, point-to-point, LAN to LAN mode, using the antenna which was built in the card. Expb used (Fig. 1-(B), (C))

[1]José Pacheco de Carvalho is with the Remote Detection Unit and the Physics Department at the University of Beira Interior, R. Marquês d'Ávila e Bolama, 6201-001 Covilhã, Portugal, E-mail: pacheco@ubi.pt.

[2]Hugo Veiga, and [3]Nuno Marques are with the Remote Detection Unit and the Informatics Centre at the University of Beira Interior, R. Marquês d'Ávila e Bolama, 6201-001 Covilhã, Portugal, E-mails: hveiga@ubi.pt, nmarques@ubi.pt.

[4]Cláudia Pacheco is with the Remote Detection Unit at the University of Beira Interior, R. Marquês d'Ávila e Bolama, 6201-001 Covilhã, Portugal, E-mail: a17597@ubi.pt.

[5]António Reis is with the Remote Detection Unit and the Physics Department at the University of Beira Interior, and with the Department of Electronics and Telecommunications/Institute of Telecommunications, at the University of Aveiro, 3810 Aveiro, Portugal, E-mail: adreis@ubi.pt.

Enterasys RoamAbout RBT-4102 level 2/3/4 access points (mentioned as APb), equipped with 16-20 dBm IEEE 802.11 a/b/g transceivers based on the Atheros 5213A chipset, internal dual-band diversity antennas [12], and the same type of level 2 switch [13]. The configuration was for minimum transmitted power and equivalent to point-to-point, LAN to LAN mode, using the internal antenna. In each type of experiment, interference free communication channels were used. WEP encryption was not activated. No power levels above the minimum were required as the access points were very close.

A laboratory setup was planned and implemented for the measurements (Fig. 2). TCP and UDP experiments at OSI level 4, were as mentioned in [11], permitting network performance results to be recorded. For a TCP connection, TCP throughput was obtained. For a UDP communication with a given bandwidth parameter, UDP throughput, jitter and percentage loss of datagrams were obtained. TCP packets and UDP datagrams of 1470 bytes size were used. A window size of 8 kbytes and a buffer size of the same value were used for TCP and UDP, respectively. One PC, with IP 192.168.0.2 was the Iperf server and the other, with IP 192.168.0.6, was the Iperf client. Jitter, which represents the smooth mean of differences between consecutive transit times, was continuously computed by the server, as specified by RTP in RFC 1889 [14]. This scheme was also used for FTP measurements, where FTP server and client applications were installed in the PCs with IPs 192.168.0.2 and 192.168.0.6, respectively.

The PCs were portable computers running Windows XP. They were configured to maximize the resources allocated to the present work. Batch command files were written to enable the TCP, UDP and FTP tests. The results were obtained in batch mode and written as data files to the client PC disk. Each PC had a second network adapter, to permit remote control from the 193.136.64.0 network, via switch.

III. Results and Discussion

In each type of experiment, Expa and Expb, the corresponding access points APa and APb were configured for the IEEE 802.11a standard, with typical fixed transfer rates. For every fixed transfer rate, measurements were made for every experiment type. In this way, for each AP type, data were obtained for comparison of the laboratory performance of IEEE 802.11a (at 6, 9, 12, 24, 36, 48 and 54 Mbps) links, measured namely at OSI levels 4 and 7 using the setup of Fig. 2.

At OSI level 1, the signal to noise ratios SNR of the point-to-point links were monitored.

For Expa and Expb, and for every nominal fixed transfer rate, an average TCP throughput was determined from several experiments. This value was used as the bandwidth parameter for every corresponding UDP test, giving average jitter and average percentage datagram loss. The main results are shown in Figs. 3-8. In Figs. 3-4, polynomial fits were made to the TCP throughput data, where R^2 is the coefficient of determination. It is seen that both APa and APb have similar TCP throughput performances, within the statistical error. In Figs. 5-8, the data points representing jitter and percentage datagram loss were joined by smoothed lines. From Figs. 5-6 it follows that the lowest jitter values are for APb. The average values of the jitter are 1.3+-0.1 ms for APb and 2.1+-0.1 ms for APa. This means that, on average, APb has a better jitter performance than APa. Figs. 7-8, where there are significant statistical errors, show that on average the percentage datagram loss is lower for APa (1.1+-0.1%) than for APb (1.4+-0.2%).

At OSI level 7 FTP transfer rates were measured versus nominal transfer rates configured in the access points for the IEEE 802.11a standard. Every measurement was the average for a single FTP transfer, using a binary file size of 100 Mbytes. The results thus obtained are represented in Figs. 9-10. Polynomial fits were made to data of each experiment. It was found that APa and APb have similar FTP performances, within the statistical error. These results show the same trends found for TCP throughput.

Fig. 1- Access Points (A), (B) [12] and switch (C) [13].

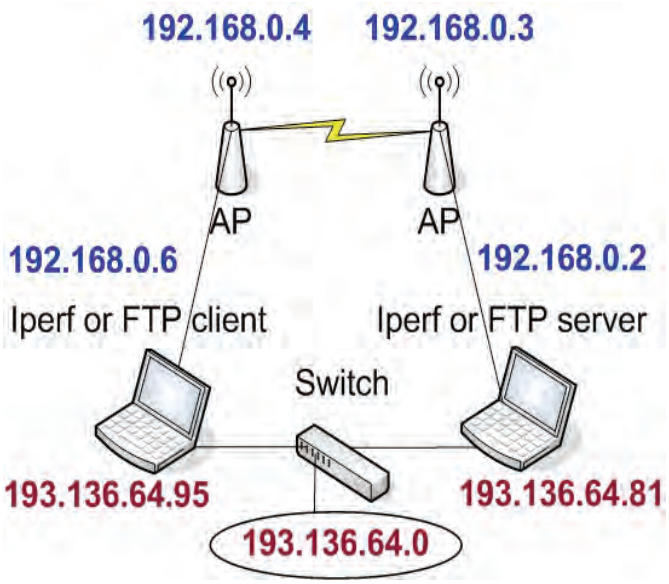


Fig. 2- Laboratory setup scheme.

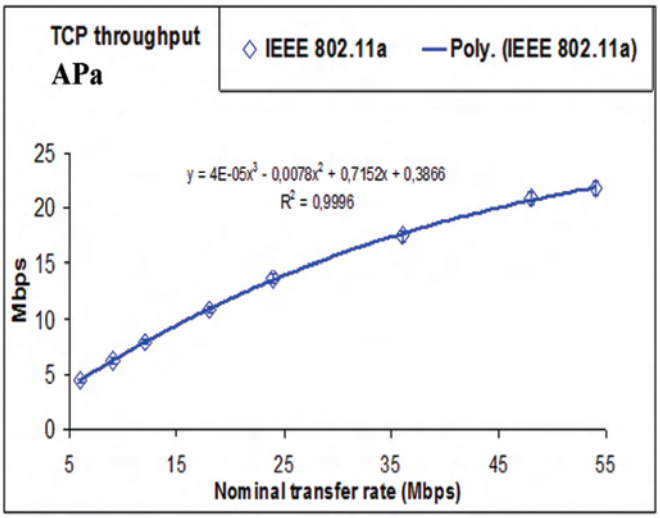


Fig. 3- TCP throughput versus technology and nominal transfer rate; Expa.

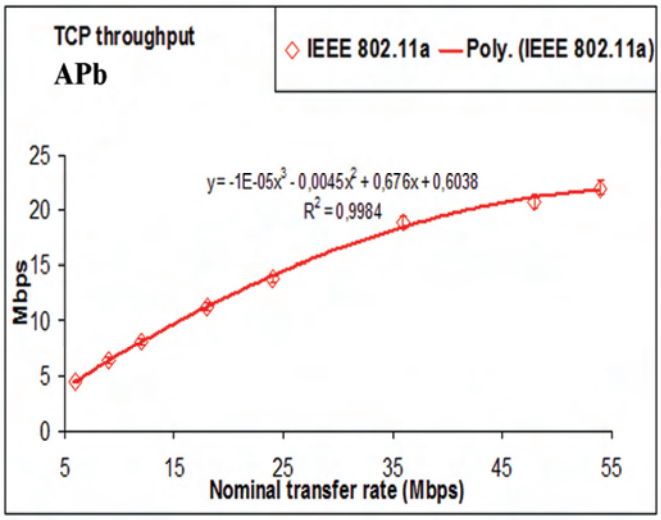


Fig. 4- TCP throughput versus technology and nominal transfer rate; Expb.

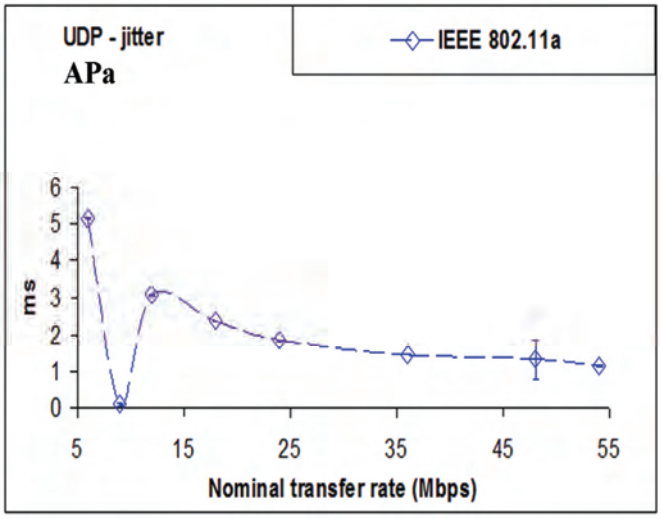


Fig. 5- UDP – jitter results versus technology and nominal transfer rate; Expa.

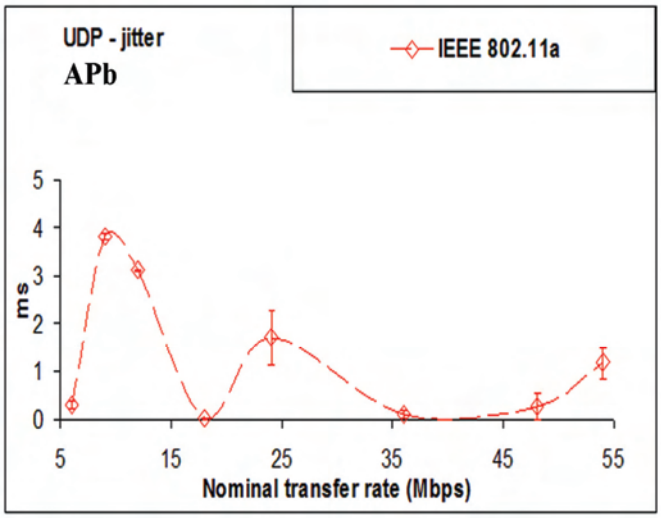


Fig. 6- UDP – jitter results versus technology and nominal transfer rate; Expb.

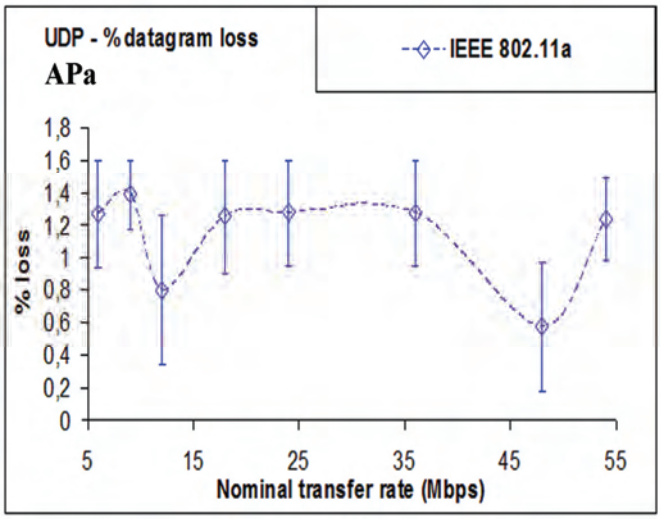


Fig. 7- UDP – percentage datagram loss results versus technology and nominal transfer rate; Expa.

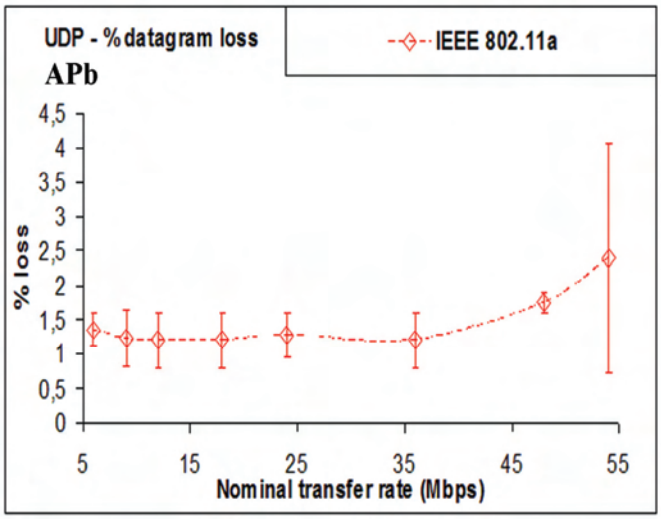


Fig. 8- UDP – percentage datagram loss results versus technology and nominal transfer rate; Expb.

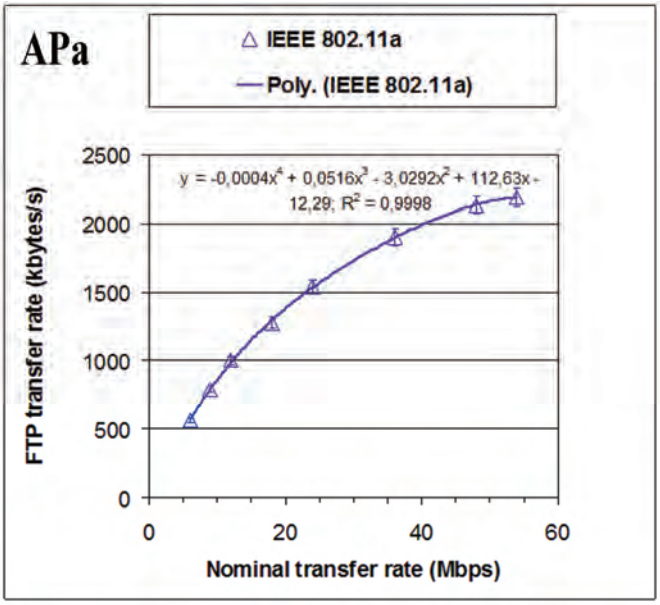


Fig. 9- FTP transfer rates versus technology and nominal transfer rate; Expa.

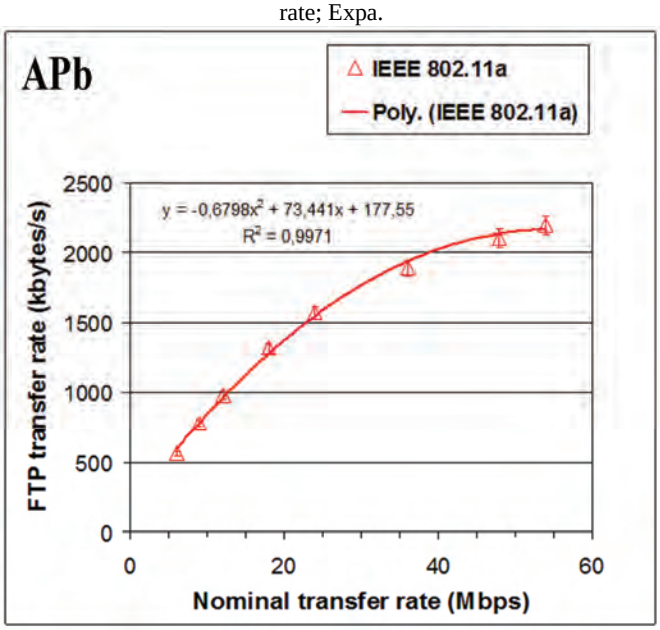


Fig. 10- FTP transfer rates versus technology and nominal transfer rate; Expb.

IV. Conclusion

In the present work a laboratory setup was planned and implemented that permitted systematic performance measurements of available access points equipments (RBTR2 and RBT-4102) from Enterasys in Wi-Fi IEEE 802.11a open point-to-point links. Through OSI level 4, TCP throughput, jitter and percentage datagram loss were measured and compared. It was found that both APs have similar TCP throughput performances. Concerning jitter, the results were found sensitive to AP type. For the percentage datagram loss, the results were not very significantly dependent on AP type. At OSI level 7 it was found that both APs have similar FTP performances. This result shows the same trends found for TCP throughput.

Additional performance measurements either started or are planned using several equipments, not only in laboratory but also in outdoor environments involving, mainly, medium range links.

Acknowledgement

Supports from Universidade da Beira Interior and FCT (Fundação para a Ciência e a Tecnologia)/POCI2010 (Programa Operacional Ciência e Inovação) are acknowledged. We acknowledge Enterasys Networks for their availability and collaboration.

References

[1] Web site http://standards.ieee.org
[2] J. W. Mark, W. Zhuang, *Wireless Communications and Networking*, Upper Saddle River, NJ: Prentice-Hall, Inc., 2003.
[3] T. S. Rappaport, *Wireless Communications Principles and Practice*, 2nd ed., Upper Saddle River, NJ: Prentice-Hall, Inc., 2002.
[4] W. R. Bruce III, R. Gilster, *Wireless LANs End to End*, NY: Hungry Minds, Inc., 2002.
[5] M. Schwartz, *Mobile Wireless Communications*, Cambridge University Press, 2005.
[6] N. I. Sarkar, K. W. Sowerby, "High Performance Measurements in the Crowded Office Environment: a Case Study", Proc. ICCT'06-International Conference on Communication Technology, Guilin, China, 27-30 November 2006, pp. 1-4.
[7] E. Monteiro, F. Boavida, Engineering of Informatics Networks, 4th ed., FCA-Editor of Informatics Ld., Lisbon, 2000.
[8] J. A. R. Pacheco de Carvalho, P. A. J. Gomes, H. Veiga, A. D. Reis, "Development of a University Networking Project", in Encyclopedia of Networked and Virtual Organizations, Goran D. Putnik, Maria Manuela Cunha, Eds. Hershey, PA (Pennsylvania): IGI Global, 2008, pp. 409-422.
[9] José A. R. Pacheco de Carvalho, Paulo A. J. Gomes, Hugo Veiga, Cláudia F. F. P. Ribeiro Pacheco, Nuno Marques, António D. Reis, "Performance Measurements of IEEE 802.11 a, g Laboratory Point-to-Point Links", Proc. ICEST 2009 – Proc. XLIV International Scientific Conference on Information, Communication and Energy Systems and Technologies, pp. 99-102, Veliko Tarnovo, Bulgaria, 7-10 July, 2009.
[10] J. A. R. Pacheco de Carvalho, P. A. J. Gomes, N. Marques, C. F. F. Ribeiro Pacheco, H. Veiga, A. D. Reis, "Laboratory Performance Comparison of IEEE 802.11 b, g Point-to-Point Links: a Case Study", Proc. Applied Electronics 2009 - 14th International Conference, pp.73-78, Pilsen, Czech, 9-10 September 2009.
[11] J. A. R. Pacheco de Carvalho, H. Veiga, P. A. J. Gomes, C. F. F. P. Ribeiro Pacheco, A. D. Reis "Experimental Performance Study of a Very High Speed Free Space Optics Link at the University of Beira Interior Campus: a Case Study", Proc. ISSPIT 2008-8th IEEE International Symposium on Signal Processing and Information Technology, Sarajevo, Bosnia and Herzegovina, December 16-19, 2008, pp. 154-157.
[12] Web site http://www.enterasys.com
[13] Web site http://www.alliedtelesis.com
[14] Network Working Group. "RFC 1889-RTP: A Transport Protocol for Real Time Applications", http://www.rfc-archive.org

TCP, UDP and FTP Equipment Performances in Laboratory Wi-Fi IEEE 802.11a WEP Point-to-Point Links

José A. R. Pacheco de Carvalho[1], Hugo Veiga[2], Nuno Marques[3]
Cláudia F. F. P. Ribeiro Pacheco[4], António D. Reis[5]

***Abstract* –Increasing attention has been focused on Wi-Fi IEEE 802.11a, as it works in the 5 GHz band and the growing use of 2.4 GHz band is producing higher interferences. Performance is a crucial issue, resulting in more reliable and efficient communications. Security is equally important. Laboratory measurements are made about several performance aspects of Wi-Fi IEEE 802.11a point-to-point links, under WEP encryption. A contribution is given to performance evaluation of this technology, using two types of access points from Enterasys Networks (RBTR2 and RBT-4102). Detailed results are presented and discussed, namely at OSI levels 4 and 7, from TCP, UDP and FTP experiments, resulting in determinations of TCP throughput, jitter, percentage datagram loss and FTP transfer rate. Conclusions are drawn about the comparative performance of the links.**

Keywords – **WLAN, Wi-Fi, WEP Point-to-Point Links, IEEE 802.11a, Wireless Network Laboratory Performance.**

I. INTRODUCTION

Wireless communications are increasingly important for their versatility, mobility and favorable prices. It is the case of microwave based technologies, e.g. Wi-Fi. The importance and utilization of Wi-Fi have been growing for complementing traditional wired networks. Wi-Fi has been used both in ad hoc mode and infrastructure mode. In this case an access point, AP, is used to permit communications of Wi-Fi devices with a wired based LAN through a switch/router. In this way a WLAN, based on the AP, is formed. Wi-Fi has penetrated the personal home, forming a WPAN, allowing personal electronic devices to communicate. Point-to-point and point-to-multipoint configurations are used both indoors and outdoors, requiring specific directional and omnidirectional antennas. Wi-Fi uses microwaves in the 2.4 and 5 GHz frequency bands and IEEE 802.11a, 802.11b and 802.11g standards [1]. As the 2.4 GHz band becomes increasingly used and interferences increase, the 5 GHz band has received considerable interest, although absorption increases and ranges are shorter.

Nominal transfer rates up to 11 (802.11b) and 54 Mbps (802.11 a, g) are permitted. CSMA/CA is the medium access control. Wireless communications, wave propagation [2,3] and WLAN practical implementations [4] have been studied. Detailed information is available about the 802.11 architecture, including performance analysis of the effective transfer rate. An optimum factor of 0.42 was presented for 11 Mbps point-to-point links [5]. Wi-Fi (802.11b) performance measurements are available for crowded indoor environments [6]. Performance has been a very important issue, resulting in more reliable and efficient communications. Telematic applications have specific performance requirements, depending on application. New telematic applications present special sensitivities to performances, when compared to traditional applications. Application characterization and requirements have been discussed [7]. Requirements have been quoted such as: for video on demand/moving images, 1-10 ms jitter and 1-10 Mbps throughput; for Hi Fi stereo audio, jitter less than 1 ms and 0.1-1 Mbps throughputs.

Wi-Fi security is most important as microwave radio signals can be easily captured by everyone. WEP was initially intended to provide confidentiality comparable to that of a traditional wired network. In spite of its weaknesses, WEP is still widely used in Wi-Fi communications for security reasons. A shared key for data encryption is involved. In WEP, the communicating devices use the same key to encrypt and decrypt radio signals.

Several performance measurements have been made for 2.4 and 5 GHz Wi-Fi [8-10], as well as high speed FSO [11]. In the present work further Wi-Fi (IEEE 802.11 a) results arise, using WEP, through OSI levels 4 and 7. Performance is evaluated in laboratory measurements of WEP point-to-point links using different types of access point equipments.

The rest of the paper is structured as follows: Chapter II presents the experimental details i.e. the measurement setup and procedure. Results and discussion are presented in Chapter III. Conclusions are drawn in Chapter IV.

[1]José Pacheco de Carvalho is with the Remote Detection Unit and the Physics Department at the University of Beira Interior, R. Marquês d'Ávila e Bolama, 6201-001 Covilhã, Portugal, E-mail: pacheco@ubi.pt.

[2]Hugo Veiga, and [3]Nuno Marques are with the Remote Detection Unit and the Informatics Centre at the University of Beira Interior, R. Marquês d'Ávila e Bolama, 6201-001 Covilhã, Portugal, E-mails: hveiga@ubi.pt, nmarques@ubi.pt.

[4]Cláudia Pacheco is with the Remote Detection Unit at the University of Beira Interior, R. Marquês d'Ávila e Bolama, 6201-001 Covilhã, Portugal, E-mail: a17597@ubi.pt.

[5]António Reis is with the Remote Detection Unit and the Physics Department at the University of Beira Interior, and with the Department of Electronics and Telecommunications/Institute of Telecommunications, at the University of Aveiro, 3810 Aveiro, Portugal, E-mail: adreis@ubi.pt.

II. EXPERIMENTAL DETAILS

Two types of experiments were carried out, which are referred as Expa and Expb. The measurements of Expa used (Fig. 1-(A), (C)) Enterasys RoamAbout RBTR2 level 2/3/4 access points (mentioned as APa), equipped with 15 dBm IEEE 802.11 a/b/g cards [12], and 100-Base-TX/10-Base-T Allied Telesis AT-8000S/16 level 2 switches [13]. The access points had RBTBH-R2W radio cards similar to the Agere-Systems model 0118 type, and firmware version 6.08.03. The configuration was for minimum transmitted power i.e. micro cell, point-to-point, LAN to LAN mode, using the antenna which was built in the card. Expb used (Fig. 1-(B), (C)) Enterasys RoamAbout RBT-4102 level 2/3/4 access points (mentioned as APb), equipped with 16-20 dBm IEEE 802.11 a/b/g transceivers based on the Atheros 5213A chipset, internal dual-band diversity antennas [12], and the same type of level 2 switch [13]. The configuration was for minimum transmitted power and equivalent to point-to-point, LAN to LAN mode, using the internal antenna. In each type of experiment, interference free communication channels were used. WEP encryption was activated, using 128 bit encryption and a shared key for data encryption composed of 13 ASCII characters. No power levels above the minimum were required, as the access points were very close.

A laboratory setup was planned and implemented for the measurements (Fig. 2). TCP and UDP experiments at OSI level 4, were as mentioned in [11], permitting network performance results to be recorded. For a TCP connection, TCP throughput was obtained. For a UDP communication with a given bandwidth parameter, UDP throughput, jitter and percentage loss of datagrams were obtained. TCP packets and UDP datagrams of 1470 bytes size were used. A window size of 8 kbytes and a buffer size of the same value were used for TCP and UDP, respectively. One PC, with IP 192.168.0.2 was the Iperf server and the other, with IP 192.168.0.6, was the Iperf client. Jitter, which represents the smooth mean of differences between consecutive transit times, was continuously computed by the server, as specified by RTP in RFC 1889 [14]. This scheme was also used for FTP measurements, where FTP server and client applications were installed in the PCs with IPs 192.168.0.2 and 192.168.0.6, respectively.

The PCs were portable computers running Windows XP. They were configured to maximize the resources allocated to the present work. Batch command files were written to enable the TCP, UDP and FTP tests. The results were obtained in batch mode and written as data files to the client PC disk. Each PC had a second network adapter, to permit remote control from the 193.136.64.0 network, via switch.

III. RESULTS AND DISCUSSION

In each type of experiment, Expa and Expb, the corresponding access points APa and APb were configured for the IEEE 802.11a standard, with typical fixed transfer rates. For every fixed transfer rate, measurements were made for every experiment type. In this way, for each AP type, data were obtained for comparison of the laboratory performance of IEEE 802.11a (at 6, 9, 12, 24, 36, 48 and 54 Mbps) links, measured namely at OSI levels 4 and 7 using the setup of Fig. 2.

At OSI level 1, the signal to noise ratios SNR of the point-to-point links were monitored.

For Expa and Expb, and for every nominal fixed transfer rate, an average TCP throughput was determined from several experiments. This value was used as the bandwidth parameter for every corresponding UDP test, giving average jitter and average percentage datagram loss. The main results are shown in Figs. 3-8. In Figs. 3-4, polynomial fits were made to the TCP throughput data, where R^2 is the coefficient of determination. It is seen that both APa and APb have similar TCP throughput performances, within the statistical error. The same trends were observed for open links. In Figs. 5-8, the data points representing jitter and percentage datagram loss were joined by smoothed lines. From Figs. 5-6 it follows that the lowest jitter values are for APa. The average jitter values are 1.8+-0.1 ms for APa and 1.9+- 0.1 ms for APb. This means that, on average, APa has a slightly better jitter performance than APb. This is the reverse situation when compared to open links. Figs. 7-8, where there are significant statistical errors, show that on average the percentage datagram loss is slightly lower for APb (1.2+-0.1%) than for APa (1.3+-0.1%).

At OSI level 7 FTP transfer rates were measured versus nominal transfer rates configured in the access points for the IEEE 802.11a standard. Every measurement was the average for a single FTP transfer, using a binary file size of 100 Mbytes. The results thus obtained are represented in Figs. 9-10. Polynomial fits were made to data of each experiment. It was found that APa and APb have similar FTP performances, within the statistical error. These results show the same trends found for TCP throughput and for open links.

Generally, the WEP throughput and FTP transfer rate results are consistent and agree, within the experimental errors, with open link results. Jitter and percentage datagram loss WEP results are, as for open links, sensitive to AP type; but APs behave in a reverse way, as described above. This is, probably, due to the WEP implementation of 802.11a for each AP type.

Fig. 1- Access Points (A), (B) [12] and switch (C) [13].

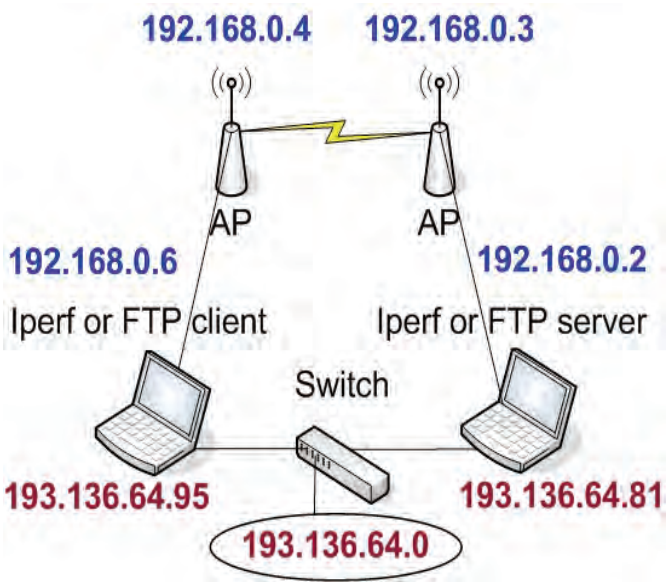


Fig. 2- Laboratory setup scheme.

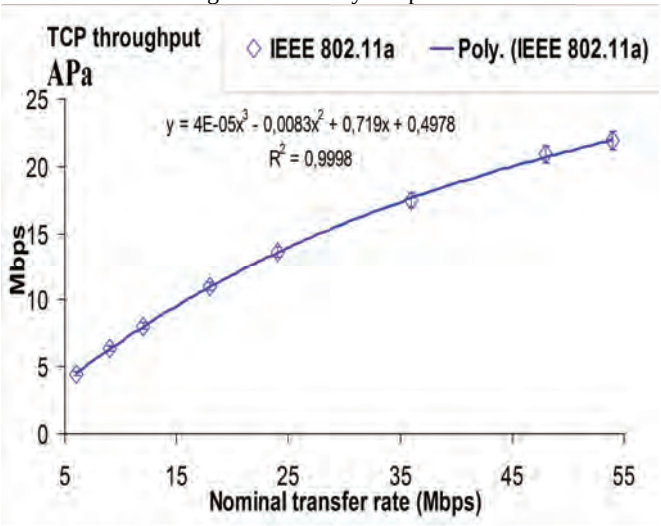


Fig. 3- TCP throughput versus technology and nominal transfer rate; Expa.

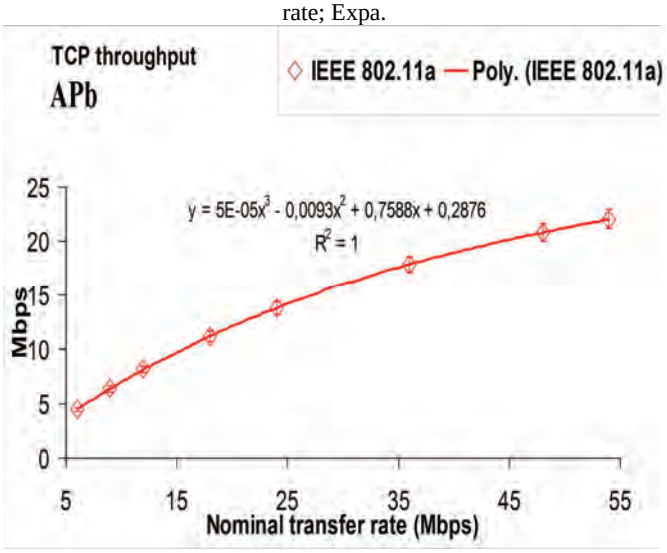


Fig. 4- TCP throughput versus technology and nominal transfer rate; Expb.

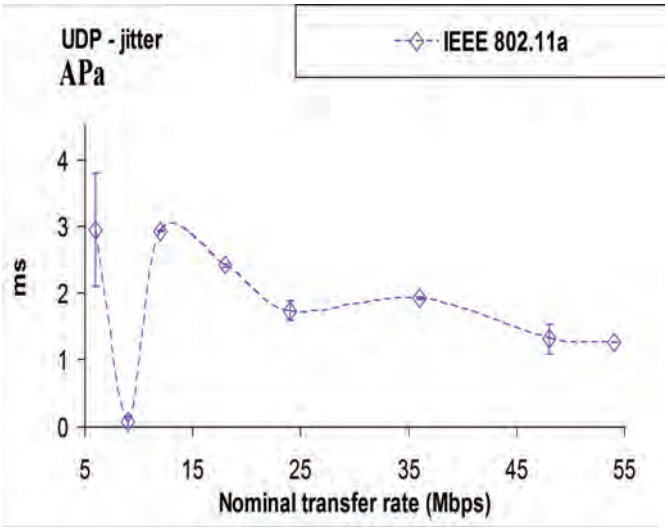


Fig. 5- UDP – jitter results versus technology and nominal transfer rate; Expa.

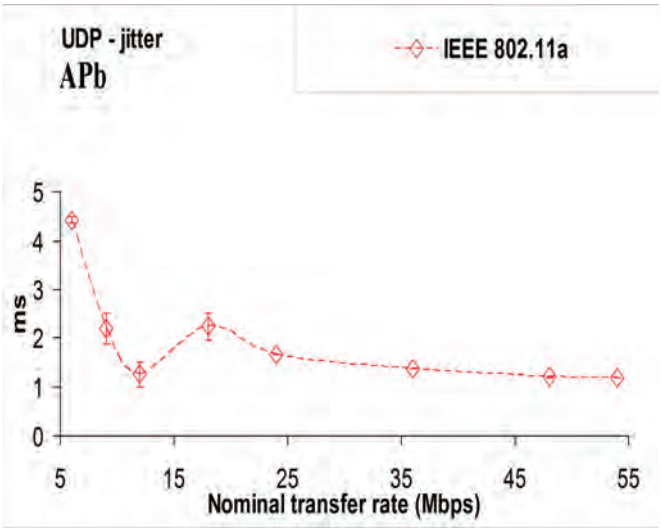


Fig. 6- UDP – jitter results versus technology and nominal transfer rate; Expb.

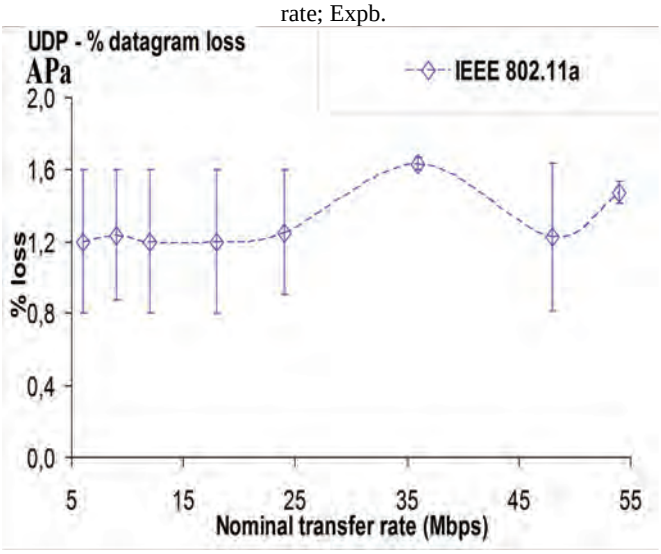


Fig. 7- UDP – percentage datagram loss results versus technology and nominal transfer rate; Expa.

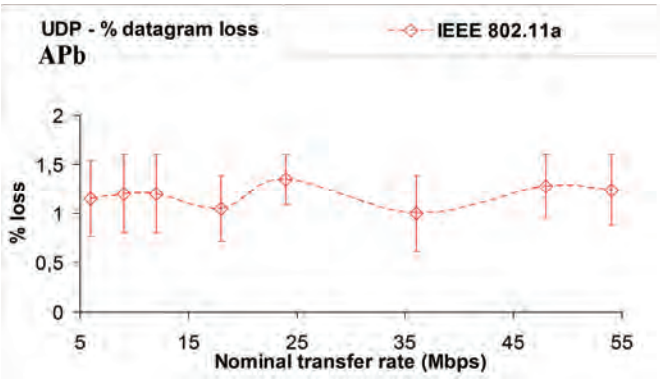


Fig. 8- UDP – percentage datagram loss results versus technology and nominal transfer rate; Expb.

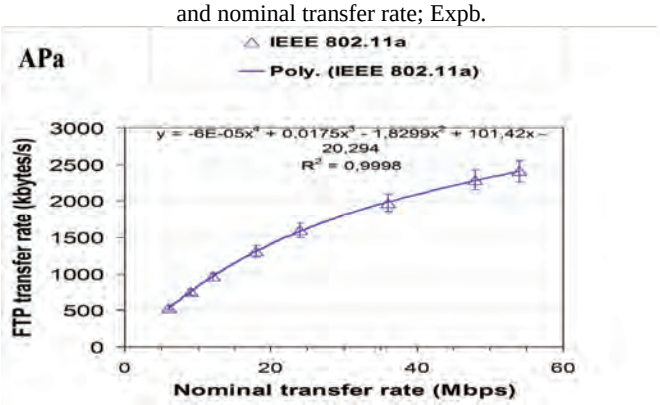


Fig. 9- FTP transfer rates versus technology and nominal transfer rate; Expa.

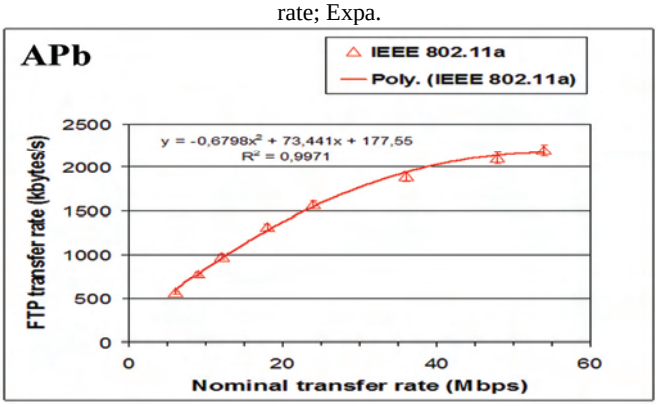


Fig. 10- FTP transfer rates versus technology and nominal transfer rate; Expb.

IV. Conclusion

In the present work a laboratory setup was planned and implemented that permitted systematic performance measurements of available access points equipments (RBTR2 and RBT-4102) from Enterasys in WEP Wi-Fi IEEE 802.11a point-to-point links. Through OSI level 4, TCP throughput, jitter and percentage datagram loss were measured and compared. It was found that both APs have similar TCP throughput performances. Concerning jitter and percentage datagram loss, results were found, as for open links, sensitive to AP type; but APs behave in a reverse way. At OSI level 7 it was found that both APs have similar FTP performances. This result shows the same trends found for TCP throughput and for open links.

Additional measurements either have been planned or are under way using several equipments, not only in laboratory but also in outdoor medium range links.

Acknowledgement

Supports from Universidade da Beira Interior and FCT (Fundação para a Ciência e a Tecnologia)/POCI2010 (Programa Operacional Ciência e Inovação) are acknowledged. We acknowledge Enterasys Networks for their availability and collaboration.

References

[1] Web site http://standards.ieee.org
[2] J. W. Mark, W. Zhuang, *Wireless Communications and Networking*, Upper Saddle River, NJ: Prentice-Hall, Inc., 2003.
[3] T. S. Rappaport, *Wireless Communications Principles and Practice*, 2nd ed., Upper Saddle River, NJ: Prentice-Hall, Inc., 2002.
[4] W. R. Bruce III, R. Gilster, *Wireless LANs End to End*, NY: Hungry Minds, Inc., 2002.
[5] M. Schwartz, *Mobile Wireless Communications*, Cambridge University Press, 2005.
[6] N. I. Sarkar, K. W. Sowerby, "High Performance Measurements in the Crowded Office Environment: a Case Study", Proc. ICCT'06-International Conference on Communication Technology, Guilin, China, 27-30 November 2006, pp. 1-4.
[7] E. Monteiro, F. Boavida, Engineering of Informatics Networks, 4th ed., FCA-Editor of Informatics Ld., Lisbon, 2000.
[8] J. A. R. Pacheco de Carvalho, P. A. J. Gomes, H. Veiga, A. D. Reis, "Development of a University Networking Project", in Encyclopedia of Networked and Virtual Organizations, Goran D. Putnik, Maria Manuela Cunha, Eds. Hershey, PA (Pennsylvania): IGI Global, 2008, pp. 409-422.
[9] José A. R. Pacheco de Carvalho, Paulo A. J. Gomes, Hugo Veiga, Cláudia F. F. P. Ribeiro Pacheco, Nuno Marques, António D. Reis, "Performance Measurements of IEEE 802.11 a, g Laboratory Point-to-Point Links", Proc. ICEST 2009 – Proc. XLIV International Scientific Conference on Information, Communication and Energy Systems and Technologies, pp. 99-102, Veliko Tarnovo, Bulgaria, 7-10 July, 2009.
[10] J. A. R. Pacheco de Carvalho, P. A. J. Gomes, N. Marques, C. F. F. Ribeiro Pacheco, H. Veiga, A. D. Reis, "Laboratory Performance Comparison of IEEE 802.11 b, g Point-to-Point Links: a Case Study", Proc. Applied Electronics 2009 - 14th International Conference, pp.73-78, Pilsen, Czech, 9-10 September 2009.
[11] J. A. R. Pacheco de Carvalho, H. Veiga, P. A. J. Gomes, C. F. F. P. Ribeiro Pacheco, A. D. Reis "Experimental Performance Study of a Very High Speed Free Space Optics Link at the University of Beira Interior Campus: a Case Study", Proc. ISSPIT 2008-8th IEEE International Symposium on Signal Processing and Information Technology, Sarajevo, Bosnia and Herzegovina, December 16-19, 2008, pp. 154-157.
[12] Web site http://www.enterasys.com
[13] Web site http://www.alliedtelesis.com
[14] Network Working Group. "RFC 1889-RTP: A Transport Protocol for Real Time Applications", http://www.rfc-archive.org

Modeling and Study of Broadband Transmission Line Transformers

Boyan D. Karapenev[1]

***Abstract* – This report presents modelling and study of broadband transmission line transformers in frequency response 0,5÷30 [MHz] with coefficient of resistant ratio 1:1, 4:1 and 1:4. Two models are shown as a lumped equivalent circuit of the broadband transmission line transformer and their model parameters have been calculated. Some transformers with 3 twists of copper enamelled wires have been made in practice and studied. Their simulation and experimental qualitative parameters are given which allows model parameter values to be optimized.**

***Keywords* – Transmission Line Transformers, Ferrite Toroidal Cores of Amidon, Modeling, Model Parameters, Study.**

I. Introduction

Broadband co-ordination of input and output resistance of a high-frequency amplifier and between two adjacent amplifier stages can be carried out by transmission line transformers employing an electromagnetic connection between the primary and secondary windings. They ensure the required resistance transformation ratio and minimum deco-ordination in a broad frequency band though they have small sizes. They have high efficiency and reliability and through them can be made: galvanic dissociation between nodes and units of the equipment, transition from asymmetric to symmetric I/O and vice versa, broadband power aggregation and division, etc.

II. Modeling and Study of Transmission Line Transformers

Broadband Transmission Line Transformers (BTLT) are constructed employing appropriately interconnected transmission lines, positioned on a ferromagnetic core which is mostly of toroidal shape [3]. The input signal excites electromagnetic waves whose linear combinations depending on the type of line connection, determine the output signal voltage. The operating principle of this type of transformer is illustrated in Fig. 1.

Fig. 1. Transmission Line Transformer

[1] Boyan D. Karapenev is with the Department of Communications Technology and Engineering of Technical University, 4 Hadji Dimitar Str., 5300 Gabrovo, Bulgaria, E-mail: bkarapenev@tugab.bg

The main component in transmission line transformers is the ferrite toroidal core used. The most important ferrite core parameters are:

- geometric dimensions: outside diameter D_{dim}, inside diameter d_{dim} and height h_{dim} in [mm];
- average length of magnetic line of force l_e, [cm];
- cross-section area A_e, [cm^2];
- volume of ferrite V_e, [cm^3];
- frequency response Δf, [MHz];
- permeability of material μ_r;
- specific volume resistance ρ, [Ω.cm];
- induction factor A_l, which is given by the expression [6]

$$A_L = \frac{L}{\omega^2} = \frac{\mu.\mu_r}{C}, \text{ [mH/1000winding]} \qquad (1)$$

Table 1 gives the main catalogue parameters of ferrite toroidal cores manufactured by the firm Amidon [1].

TABLE I
CATALOG PARAMETERS OF AMIDON FERRITE TOROIDAL CORES

Parameter	Ferrite core grade, Amidon			
	FT82-43	FT82-61	FT82-77	FT114-77
D_{dim}, [mm]	21,000	21,000	21,000	29,000
d_{dim}, [mm]	13,100	13,100	13,100	19,000
h_{dim}, [mm]	6,350	6,350	6,350	7,000
l_e, [cm]	5,260	5,260	5,260	7,420
A_e, [cm^2]	0,246	0,246	0,246	0,375
V_e, [cm^3]	1,290	1,290	1,290	2,790
Δf, [MHz]	1÷50	10÷200	0,5÷30	0,5÷30
A_l,[mH/1000w]	557	73,3	1170	1270
μ_r	850	125	2000	2000
ρ, [Ω.cm]	1.10^5	1.10^8	1.10^2	1.10^2

A. Model Parameter Calculation of Ferrite Toroidal Cores

The equivalent diagram (model) of a ferrite toroidal core is presented in Fig. 2. The model parameter values R, L and C can be determined by the following dependencies:

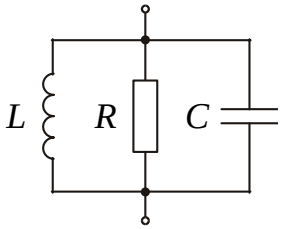


Fig. 2. Model Parameters of Ferrite Toroidal Core

- ***inductivity L*** is derived from the condition $B = \frac{\mu_r N.I}{2\pi r}$, hence

$$L_1 \approx \frac{\mu_r N^2 A_e[cm^2]}{2\pi r_{av}[cm]}.10^{-2} = \frac{\mu_r N^2 A_e[cm^2]}{l_e[cm]}.10^{-2}, [\mu H] \quad (2)$$

where B [Tesla] or [Gauss] is the magnetic field produced, A_e,[cm^2]=(r_2-r_1).h_{dim}, r_{av},[cm] is the average radius of the ferrite toroidal core as r_{av}= [(r_2-r_1)/2]+r_1.

To determine inductivity L of a ferrite toroidal core, it is assumed that the number of windings N is equal to 1.

If the catalogue parameter l_e, [cm] is given, L can be determined by [6]

$$L_2 = 0{,}4\pi N^2 \mu_r \frac{A_e[cm^2]}{l_e[cm]}.10^{-8}\,[H]. \quad (3)$$

Owing to the existing roundness in the geometric shape of the ferrite toroidal cores, when calculating the average diameter r_{av} (through r_2 and r_1), an error appears and for FT82$_{AMIDON}$ it is +1,78 %, and for FT114$_{AMIDON}$ +1,56 %. To avoid this error r_{cp} should be determined as r_{av}=l_e/(2π);

- ***capacitance C*** is determined by the geometric dimensions of the ferrite following:

$$C = 2{,}8\left(1{,}2781 - \frac{h_{dim}}{D_{dim}}\right).\sqrt{\frac{2\pi^2\left(D_{dim} - h_{dim}\right)\frac{h_{dim}}{2}}{4\pi}}, \text{ [pF]} \quad (4)$$

where D_{dim} and h_{dim} are in inches (1 [inch] = 25,4 [mm]). To calculate C in geometric dimensions in millimeters [mm], use (4) divided by 25,4;

- ***ohmic resistance R*** is determined from the catalogue parameter ρ by:

$$R = \frac{\rho.l_e}{S} = \frac{\rho.(2\pi r_{av})}{\pi.\left(r_2 - r_1\right)^2}, [\Omega] \quad (5)$$

for dimensions of r_2, r_1 and r_{av} (l_e) in [m].

On the basis of (2)÷(5) and catalogue parameters of ferrite toroidal cores of the firm AMIDON given in Table 1, their model parameters have been calculated and presented in Table 2.

TABLE 2
MODEL PARAMETERS OF AMIDON FERRITE TOROIDAL CORES

Trade Ferrite	L_1 [μH]	L_2 [μH]	C [pF]	R [kΩ]
FT82-43	0,398	0,499	0,919	$1{,}074.10^3$
FT82-61	0,058	0,073	0,919	$1{,}074.10^6$
FT82-77	0,935	1,175	0,919	1,074
FT114-77	1,011	1,270	1,256	0,945

Deductions: From the calculated model parameters of AMIDON ferrite toroidal cores presented in Table 2 it has been found that:

- for each overall dimension the ferrite inductivity value L varies within relatively close limits from a few tenths of μH to 1,3 μH, and it is directly proportional both to its dimensions and to its permeability value μ_r;
- since $tg\delta$ has very small values with ferrites, the capacitance C of the ferrite toroidal core depends only on its geometric dimensions - (4). From the obtained values of the model parameter C of the ferrite toroidal core it has been found that its values vary within (0,9÷1,3) pF, which allows in the general case its averaged value 1 [pF] to be adopted;
- due to great differences in the value of catalogue parameter "specific volume resistance" ρ of various grades of ferrite toroidal cores of the order of 10^6, its ohmic resistance value varies from several kΩ to several hundreds of MΩ and it is strictly individual.

B. Model Parameters Calculating of Broadband Transmission Line Transformer

The complete model of broadband transmission line transformer, with secondary side parameters reduced to the primary side is shown in Fig. 3. Resistors R_1 and R_2 connected in series, present the existing losses, respectively in the conductors of the primary and secondary windings. When transmission lines are used – small-length conductors in broadband transformers with ferromagnetic cores, the active loss share in the total losses is negligibly small.

C_{11} and C_{22} are the distributed shunt capacitances of the primary and secondary windings respectively, and C_{12} is the distributed capacitance between transmission lines which provides the electromagnetic connection at high frequencies. Capacitance C_{12} can form a transmission line together with distributed inductances L_{S1} and L_{S2}. The characteristic impedance of the transmission line can be regulated by changing the conductor length, number of windings, angle and pitch of twist, location of windings on ferrite core, etc.

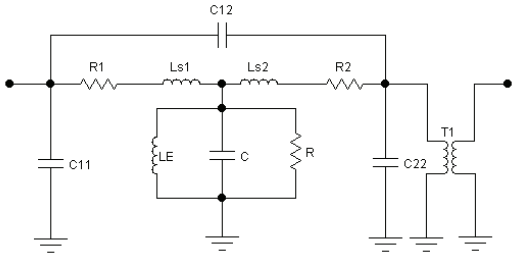


Fig. 3. Full Model of Broadband Transmission Line Transformer with Ideal Transformer

Model parameters of BTLT are determined as follows:

- ohmic resistance values R_1 and R_2 of the primary and secondary windings according to (5) are

$$R_1 = R_2 = N.\frac{\rho_{Cu}(2\pi r_{av})}{\pi\left(r_2 - r_1\right)^2}, [\Omega] \quad (6)$$

when ρ_{Cu}=1,66 [μΩ/cm];

- dissipation inductances (leakage) L_{S1} and L_{S2}, whose typical value is 1÷2 %, is assumed to be 1,3 % of the ferrite core inductance value L when N windings of the primary (L_{11}) and secondary (L_{22}) sides are available, calculated by (2) – L_{S1}=L_{S2}=1,3%.L. Model parameter LE=L_{11}+L_{22}+L_1 (L_{11}=L_{22}=L);

- intrinsic capacitances C_{11} and C_{22} have been measured in practice using a passive bridge of BTLT constructed with the preset output data and C_{11}=C_{22}=5 [pF], and depending on the twist pitch of the transmission lines used (enamelled copper conductors) the inter-winding capacitance value C_{12} is adopted within (0,1÷10) pF.

Model parameters for the presented complete model of BTLT in Fig. 3 have been calculated using the following output data: ferrite core grades AMIDON FT82-77 and FT114-77, nz=1:1, number of windings N=7, diameter of the copper enamelled conductor used d_{Cu}=0,62 [mm], given in Table 3.

TABLE 3
MODEL PARAMETERS OF BROADBAND TRANSMISSION LINE TRANSFORMERS

BTLT	Model parameter values		
	R, [Ω]	L, [μH]	C, [pF]
Ferrite FT82-77	R=1074	L=0,935	C=0,919
Primary side	R_1=1,25	L_{11}=45,82 L_{S1}=0,6	C_{11}=5
Secondary side	R_2=1,25	L_{22}=45,82 L_{S2}=0,6	C_{22}=5
Model parameter LE	-	LE=92,58	-
Inter-winding capacitance C_{12}=2 [pF]			

BTLT	Model parameter values		
	R, [Ω]	L, [μH]	C, [pF]
Ferrite FT114-77	R=945	L=1,011	C=1,256
Primary side	R_1=1,1	L_{11}=49,54 L_{S1}=0,64	C_{11}=5
Secondary side	R_2=1,1	L_{22}=49,54 L_{S2}=0,64	C_{22}=5
Model parameter LE	-	LE=100,1	-
Inter-winding capacitance C_{12}=2 [pF]			

Fig. 4 presents a complete model of BTLT with Amidon FT82-77 ferrite core, containing a dependent voltage-controlled voltage source, in relation to which the calculated model parameters are symmetrical. The dependent voltage source V with transmission coefficient 1 when nz=1:1 is connected between the model parameter values R_1, L_{S1} and R_2, L_{S2} divided by two. Similar modeling of BTLT is closer to its real construction.

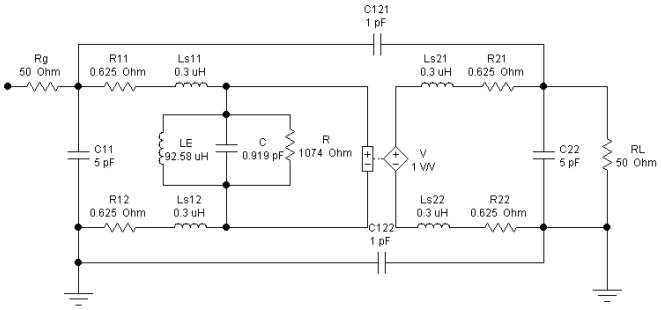


Fig. 4. Full Model of Broadband Transmission Line Transformer with Voltage-Controlled Voltage Source

The advantage of the proposed BTLT model with dependent voltage-controlled voltage source is that signal attenuation at its output can be compensated and thus to be able to specify and provide its high performance and efficiency.

When modeling BTLT with resistance transformation ratios nz≠1:1, it is necessary to preset the respective voltage transformation ratio. When nz=1:4 U_2=2.U_1 or n_U=1:2, and therefore the model of BTLT with a dependent voltage-controlled voltage source should be used. It is presented in Fig. 4 with voltage transmission coefficient, in this case V=2 (1V/2V). For BTLT with nz=4:1 n_U=2:1 and V=0,5.

C. Simulation Results of BTLT with nz=1:1, nz=4:1 and nz=1:4 for the model in Fig. 4

Using Multisim software, simulation study of amplitude-frequency responses of BTLT model with ferrite cores FT82-77 and FT114-77 with resistance transformation ratios nz=1:1, nz=4:1 and nz=1:4 (Fig. 4), presented in Fig. 5, has been conducted.

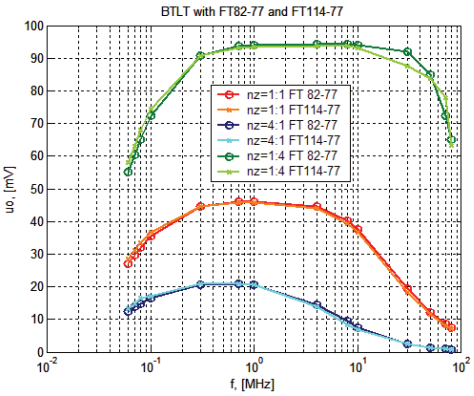


Fig. 5. Amplitude-frequency Responses of Broadband Transmission Line Transformer with Voltage-Controlled Voltage Source

Table 4 presents BTLT performance characteristics obtained: lower fb and upper fh cut-off frequency and bandwidth Δf.

TABLE 4
QUALITATIVE PARAMETERS OF BROADBAND TRANSMISSION LINE TRANSFORMERS

nz	BTLT with Amidon FT82-77		
	fb, [kHz]	fh, [MHz]	Δf, [MHz]
1:1	82,60	14,07	13,99
4:1	81,40	3,72	3,64
1:4	83,85	77,85	77,77
nz	BTLT with Amidon FT114-77		
	fb, [kHz]	fh, [MHz]	Δf, [MHz]
1:1	76,00	13,18	13,10
4:1	73,96	3,51	3,44
1:4	76,88	75,62	75,54

TABLE 5
QUALITATIVE PARAMETERS OF BROADBAND TRANSMISSION LINE TRANSFORMERS

nz	BTLT with Amidon FT82-77		
	fb, [kHz]	fh, [MHz]	Δf, [MHz]
1:1	1	120	119,99
4:1	1	78	77,99
1:4	50	14	13,95
nz	BTLT with Amidon FT114-77		
	fb, [kHz]	fh, [MHz]	Δf, [MHz]
1:1	1	73	72,99
4:1	10	125	124,99
1:4	80	13	12,92

After a comparative evaluation of the obtained bandwidth Δf for various values of *nz* it has been found that it has the largest width when *nz*=1:4 (n_U=1:2), which exceeds the preset operating frequency range of the employed ferrite toroidal core more than twice.

Amplitude-frequency responses of BTLT with ferrite cores FT82-77 and FT114-77 with identical *nz* match. The greatest difference is in the values of fh and Δf when *nz*=1:4, which decrease with the larger ferrite core compared to the smaller one by about 3 %.

D. Experimental Results of BTLT with nz=1:1, nz=4:1 and nz=1:4

Experimental studies have been carried out to find the performance characteristics fb, fh and Δf of BTLT with ferrite toroidal cores manufactured by Amidon: FT82-77 and FT114-77 for *nz*=1:1, *nz*=4:1 and *nz*=1:4. BTLT were made in practice with 7 windings from a twisted couple of copper enameled conductors with diameter d=0,62 [mm] and 3 twists per 1 [cm].

Fig. 6 shows the amplitude-frequency responses obtained experimentally at input voltage Ui=48,76 [mV], and Table 5 lists the achieved performance characteristics.

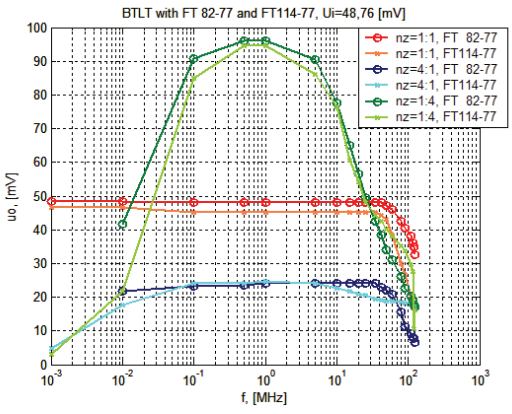


Fig. 6. Amplitude-frequency Responses of Broadband Transmission Line Transformer – Experimental Results

On the basis of the experimentally obtained results, it has been found that:

- the transmission coefficients in the bandwidth for various resistance transformation ratios *nz* are different and they almost reach their respective theoretical values;
- for *nz*=1:1 the area with an even voltage transmission coefficient in the bandwidth is the largest and in this case no drop can be reported in the low frequency range;
- when *nz*=1:4 a drop in amplitude-frequency responses has been observed both at low and high frequencies, hence a clearly formed bandwidth exists.

III. CONCLUSION

This paper presents two models of BTLT with ferrite toroidal cores made by Amidon and a technique for calculating their model parameter values. Changes in their values could enable us to optimize performance characteristics: *f*b, *f*h and Δ*f*. This allows the proposed models to be used in modeling BTLT, in conducting simulation studies of broadband amplifier modules containing BTLT, powerful broadband amplifiers, bridge circuits for summing and division of signals, and other devices used in radio-transmission engineering.

REFERENCES

[1] Amidon Associates, „Iron-powder and Ferrite Coil Forms", Catalogue, 2008.

[2] Breed, G.A., "Transmission Line Transformer Basics," Applied Microwave & Wireless, Vol. 10, No. 4, May 1998.

[3] Dobrev D., L. Jordanova, "Radiocommunications", Part One, Sofia, Siela, 2001.

[4] Horn, J. and G. Boeck, "Design and Modeling of Transmission Line Transformers," IEEE 2003 Interna- tional Symposium on Microwave Theory and Techniques.

[5] Kuo C.C., M.Y. Kuo, and M.S. Kuo, "Novel Model of 2:1 Balance-to-Unbalance Transmission-Line Transformer," Electronics Letters, Vol. 31, No. 23, 9 November 1995.

[6] Sevick J., "Transmission Line Transformers", New Jersey, Noble Publishing Corp., 4th Edition, 2001.

[7] Trask Chris, "Transmission Line transformers: Theory, Design and Application", Part I and II, 2006.

Structure Related FD BPM Design of Photonic Devices

Dušan Ž. Djurdjević[1]

***Abstract* – The structure related finite difference beam propagation method (SR-FD-BPM) is recently developed simulation technique for photonic and optoelectronic design. This paper reviews some key issues related to this conceptually straightforward and flexible numerical simulation technique. SR-FD-BPM is based on the numerical solution of the 3D vector Helmholtz's equation subject to open boundary conditions. Structure related co-ordinate transformation approach allows the comfortable analysis of wide variety of geometrically complex light-wired photonic structures. Some illustrative design examples are presented in the paper.**

***Keywords* – Beam propagation method, Finite difference, Photonics, Optoelectronics, CAD design, Numerical simulations.**

I. INTRODUCTION

Advances in photonics and optoelectronics in the 21st century have been mostly driven by the demands of the telecom and datacom boom. Stringent demands in ever-changing photonics industry are to enable production of high-quality components, to increase efficiencies and reduce costs. The research in modelling techniques and the developments of computer-aided design (CAD) software for modelling photonic and optoelectronic components and systems play the crucial and extremely significant role behind the industrial and commercial scene.

The most of integrated photonics and optoelectronic devices and components are built in dielectric waveguide technology. The numerical simulation has become a mandatory approach providing excellent results in photonic design. Amongst several developed numerical techniques, the beam propagation method (BPM) is the most widely used numerical simulation technique for modelling photonic devices. BPM is a particular approach for numerical solving of appropriate approximation of an exact vector Helmholtz's equation (known as the vector paraxial or Fresnel's equation).

FE-BPM, MoL-BPM, FDTD-BPM algorithm, etc., have been developed in recent two decades [1]. One of the most commonly used BPM simulation algorithm is the frequency-domain based finite difference beam propagation method (FD-BPM), [1-4]. FD-BPM gives a numerical solution of vector (or scalar) Fresnel's equation for monochromatic waves (fundamental or higher order mode fields) propagating in the waveguide based photonic structure. Most commercial BPM software is based on FD-BPM.

[1]Dušan Ž. Djurdjević is with the Faculty of Technical Sciences, Knjaza Miloša 7, 38220 Kosovska Mitrovica, Serbia, E-mail: drdusan.djurdjevic@gmail.com

FD-BPM is usually implemented in a rectangular co-ordinate system (standard FD-BPM), [2]. If the structure under analysis contains oblique or curved interfaces or when the structure is changing in the direction of the propagation, the inevitable staircase approximation of the boundaries causes serious problems and certain restrictions of the method. To avoid using the fine meshes and small propagation steps, FD-BPM has been reformulated in non-orthogonal structure related (SR) co-ordinate systems, [5]. In SR co-ordinate schemes for FD-BPM the discretisation procedure exactly matches the local geometry of the structure, thus eliminating non-physical scattering due to the staircasing effect.

SR approach is used nowadays to design tapered, oblique and bi-oblique shaped waveguide-based devices, such as z-variant directional waveguide couplers, y-branches, optical interconnects, waveguide polarizers, optical modulators and similar components that include waveguide bends. Some illustrative examples of SR design, including BPM analysis of a sloped walled rib waveguide and 3D SR-BPM analysis of an "S"-curved directional waveguide coupler, are given.

II. STRUCTURE RELATED FD-BPM FORMULATION

Attempts to overcome the presence of the staircasing in the FD related numerical methods have been resulted, during the last two decades, in the developing of the so-called improved FD-BPM schemes. The improved FD approach takes into account the boundary conditions for the field and its derivatives near the dielectric interfaces. Contrary to the improved FD approach, SR co-ordinate systems naturally follow sloped or curved geometry of dielectric interfaces. In SR FD-BPM algorithm coarser mesh sizes can be used for the same accuracy, in comparison to the rectangular FD schemes.

A. SR-FD-BPM in the transverse plane

For z-invariant structures (constant cross-section in the transverse plane), the transverse plane (x, y) is transformed in appropriate non-orthogonal SR co-ordinates. As an example of the SR-FD-BPM approach in the transverse plane, some results of the sloped walled rib waveguide analysis are given. The staircasing approximation of the sloped waveguide wall, obtained with the standard FD-meshing procedure in the rectangular co-ordinate system, is shown in Fig. 1.

The simplest case of the scalar Helmholtz's equation in the rectangular co-ordinate system (x, y, z), reads

$$\left[\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}+k^2n^2\right]\mathbf{\Phi_t}=0, \qquad (1)$$

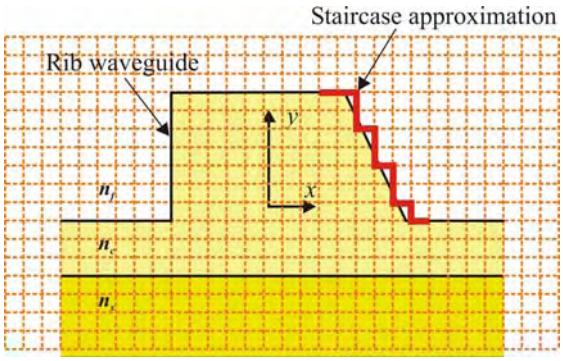


Fig. 1. Staircase approximation forced onto a rectangular grid in the transverse plane, causes non-physical numerical noise in simulation.

where $k = 2\pi/\lambda$ is the free space wave number, λ is the operating wavelength, $n = n(x, y)$ is the z invariant refractive index and $\mathbf{\Phi_t}(x, y, z)$ represents the scalar transverse electric, $\mathbf{E_t}$, or magnetic field, $\mathbf{H_t}$. Eq. (1) is not suitable for obtaining stable numerical algorithms and it is usually replaced with its first order paraxial approximation by assuming a slowly-varying envelope approximation of the transverse field,

$$\mathbf{\Phi_t}(x, y, z) = \mathbf{F_t} e^{-j\beta z}, \tag{2}$$

where $\beta = kn_0$ is an imposed propagation constant for the scalar field envelope $\mathbf{F_t}$. By neglecting the $\partial^2/\partial z^2$ term, the one-way scalar paraxial wave equation, or Fresnel's equation, in the Cartesian orthogonal co-ordinate system is obtained [2],

$$j2kn_0 \frac{\partial \mathbf{F_t}}{\partial z} = \frac{\partial^2 \mathbf{F_t}}{\partial x^2} + \frac{\partial^2 \mathbf{F_t}}{\partial y^2} + k^2(n^2 - n_0^2)\mathbf{F_t} = 0, \tag{3}$$

where n_0 denotes a reference refractive index. Eq. (3) can be rewritten in any non-orthogonal transverse co-ordinate system. If we choose the non-orthogonal oblique co-ordinate system (u, v) in the transverse plane, in which $u = x \mp y \tan\theta$, $v = y$, one can derive the scalar paraxial wave equation in the oblique co-ordinate system as, [5-7],

$$\sigma \frac{\partial \mathbf{F_t}}{\partial z} = \sec^2\theta \frac{\partial^2 \mathbf{F_t}}{\partial u^2} \pm 2\tan\theta \frac{\partial^2 \mathbf{F_t}}{\partial u \partial v} + \frac{\partial^2 \mathbf{F_t}}{\partial v^2} + \kappa^2 \mathbf{F_t}, \tag{4}$$

where $\sigma = j2kn_0$, $\kappa^2 = k^2(n^2 - n_0^2)$, the envelope of the field $\mathbf{F_t} = \mathbf{F_t}(u, v, z)$ and the refractive index $n = n(u, v)$ are functions of the oblique co-ordinates u and v. A sloped angle θ is measured with respect to the negative direction of the y axis. In Eq. (4) the sign "+" stands for the right-hand side type of the oblique co-ordinate system.

In the case of a tapered co-ordinate system (t, v) in the transverse plane, in which $t = \tan\theta$, $x = t(y - y_0)$, $v = y$, the scalar paraxial wave equation can be expressed as, [5-7],

$$\sigma \frac{\partial \mathbf{F_t}}{\partial z} = \frac{\partial^2 \mathbf{F_t}}{\partial v^2} - \frac{2t}{v - v_0} \frac{\partial^2 \mathbf{F_t}}{\partial v \partial t} + \frac{1}{(v - v_0)^2} \frac{\partial}{\partial t} \left[(1 + t^2) \frac{\partial \mathbf{F_t}}{\partial t} \right] + \kappa^2 \mathbf{F_t}, \tag{5}$$

where the envelope of the scalar field $\mathbf{F_t} = \mathbf{F_t}(t, v, z)$ and the refractive index $n = n(t, v)$ are functions of the tapered co-ordinates t and v. The origin of tapered co-ordinate system (x_0, y_0) is given parametrically as $v_0 = y_0 = x_0 \cot\theta$.

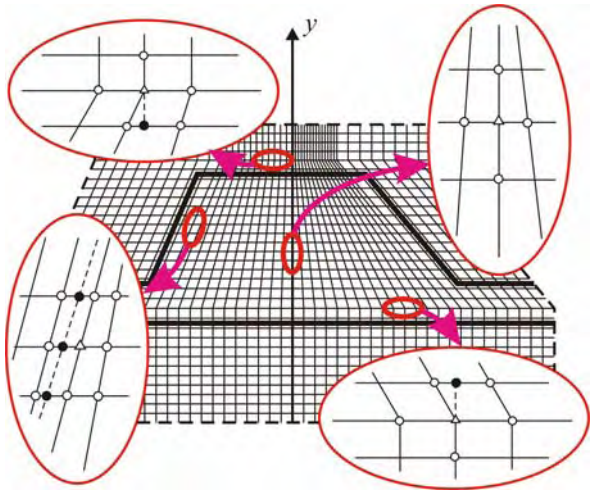


Fig. 2. SR "ROTOR" scheme of FD meshing, [6]. Interpolated field regions are shown in the small ellipses and interpolation procedures are shown in the zoomed ellipses (solid circles denote interpolated and empty circles denote directly sampled field values).

Each discretisation scheme offers potential benefits. By using hybrid SR "ROTOR" (**r**ectangular-**o**blique-**t**apered-**o**blique-**r**ectangular) FD-scheme, where non-orthogonal systems are coupled with rectangular one, FD-discretisation, where oblique interfaces are modelled exactly, can be obtained (co-ordinate SR lines are parallel with dielectric interfaces). Fig. 2 shows FD-discretisation with "ROTOR"-type scheme of an asymmetrical sloped walled rib waveguide. The deficiency of the hybrid FD-schemes is the necessity to join different co-ordinate systems. However, this can be done by using a simple linear approximation. Interpolation procedures are illustrated in Fig. 2 and explained in [6].

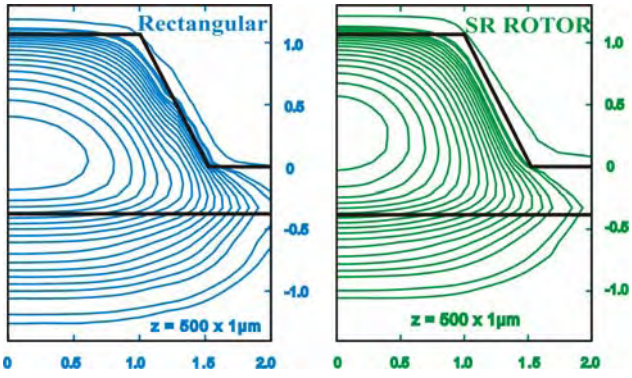


Fig. 3. The filed plots obtained with the standard rectangular and SR "ROTOR" FD-BPM numerical simulations, [6].

By FD-discretising transverse waveguide field $\mathbf{F_t}$ (Eqs. (3)) in the appropriate waveguide regions, an FD-BPM algorithm can be easily developed. BPM field plots obtained by standard rectangular and SR "ROTOR" numerical

simulation algorithm (both emloying a standard Crank-Nicolson (CN) method, an imaginary distance (ID) mode solver and the TBC boundary conditions at the edges of the computational window, [6]), after 500 BPM steps ($\triangle z = 1\mu$m), with $\triangle x = \triangle y = 0.1\mu$m, are given in Fig. 3.

Field plots are calculated for a symmetrical sloped walled rib waveguide ($\theta = 25^0$, $n_f = 1$, $n_c = 3.44$, $n_s = 3.34$), with $n_{reff} = n_0 = 3.40483$ obtained by ID mode solver. The deterioration of the field obtained in the standard FD-BPM simulation near the oblique boundaries is clearly indicated. SR-FD-BPM allows simulations with coarser meshes to a prescribed accuracy in comparison to the standard rectangular BPM, offering savings in computational time and memory.

B. SR-FD-BPM in the longitudinal direction

SR-FD-BPM approach is successfully applied to the photonic waveguide structures employing curved waveguide sections in the propagation direction. The 3D plot of the "S"-curved directional waveguide coupler is given in Fig. 4. 3D geometry of the curved directional "Y"-branch is presented in Fig. 5. Staircase approximation forced onto a rectangular grid in the longitudinal direction is shown in Fig. 6.

The approach of the longitudinally dependent SR-FD-BPM is illustrated by formulating the one-way scalar paraxial SR wave equation in the general 3D non-orthogonal co-ordinate system (u, v, w). If we choose the co-ordinates (u, w) as $x = f(u, w)$, $z = w$ and $y = v$, under the one-way paraxial approximation, assuming that propagation occurs in the w direction, $+w$, (i.e. in the $+z$ with rectangular co-ordinates),

$$\mathbf{\Phi_t}(u, v, w) = \mathbf{F_t}(u, y, w)e^{-j\beta w}, \tag{6}$$

we obtain the scalar SR wave equation from Eq. (1), [5], [9],

$$L\frac{\partial}{\partial w}\mathbf{F_t} = M\mathbf{F_t}\,, \tag{7}$$

where the operators L and M are shown to be, [5],

$$L = 2j\beta C + B\frac{\partial}{\partial u}, \tag{8}$$

$$M = A\frac{\partial^2}{\partial u^2} + (Bj\beta - D)\frac{\partial}{\partial u} + C(k^2 - \beta^2) + \frac{\partial^2}{\partial y^2}. \tag{9}$$

In Eqs. (6) and (7) the field envelope $\mathbf{F_t} = \mathbf{F_t}(u, y, w)$ can be either $\mathbf{E_t} = \mathbf{E_t}(u, y, w)$ or $\mathbf{H_t} = \mathbf{H_t}(u, y, w)$. $A = A(u, w)$, $B = B(u, w)$, $C = C(u, w)$ and $D = D(u, w)$ are functions of the partial derivatives of $f(u, w)$, [5], [9],

$$A = 1 + \left(\frac{\partial f}{\partial w}\right)^2,\quad B = 2\left(\frac{\partial f}{\partial u}\right)\left(\frac{\partial f}{\partial w}\right),\quad C = \left(\frac{\partial f}{\partial u}\right)^2,$$

$$D = \frac{1}{\frac{\partial f}{\partial u}}\left[A\frac{\partial^2 f}{\partial u^2} - B\frac{\partial^2 f}{\partial w\,\partial u} + C\frac{\partial^2 f}{\partial w^2}\right]. \tag{10}$$

The SR approach allows to designer the flexibility in analysis to use any curvature function $x = f(u, w)$ providing optimal integrated optic requirements. A standard CN method can be easily introduced in the 3D SR-FD-BPM algorithm, and for the well-confined waveguide fields TBCs are typically applied in a standard way.

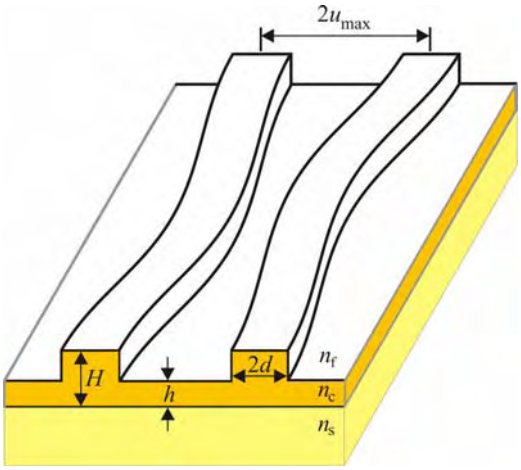


Fig. 4. Schematic diagram of the "S"-curved 3D coupler.

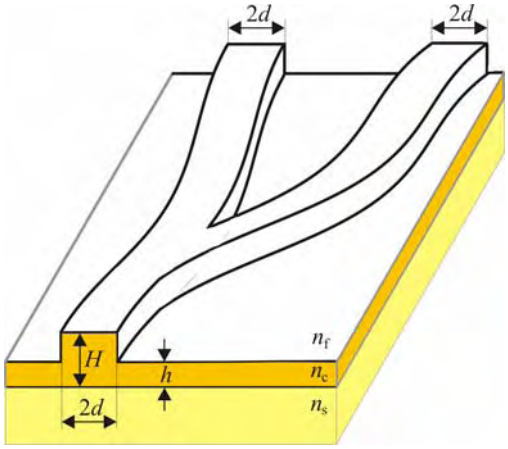


Fig. 5. Schematic diagram of the curved 3D "Y"-branch.

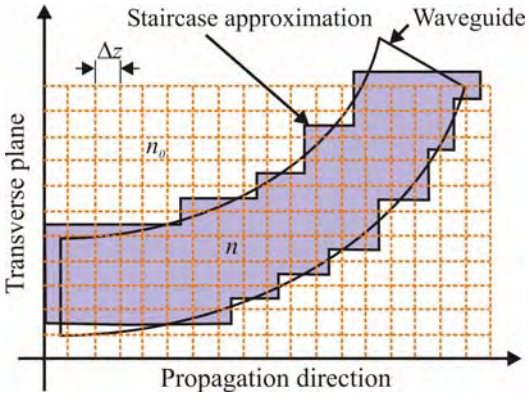


Fig. 6. Staircase approximation forced onto a rectangular grid in the longitudinal direction, causes non-physical scattering of the field.

Examples of 3D curved directional couplers design are presented to highlight the effectiveness of the SR FD-BPM. A symmetrical directional coupler made from two identical and

adjacent but spatially separated curved input (I) and output (II) waveguides is shown in Fig. 4.

The curvature function $x = f(u, w)$ of an "S" curved coupler can be given parametrically or explicitly. A cosine type SR geometry was considered in [9], where functions A, B, C and D, Eq. (10), are easily obtained analytically, however, they can be even computed numerically. If the curvature function changes slowly with w, only the function C varies, thus we can introduce $A \simeq 1, B \simeq 0, D \simeq 0$, which is usually the case except for the sharp guide bends. The discretisation mesh in the (u, w) plane is shown in Fig. 7.

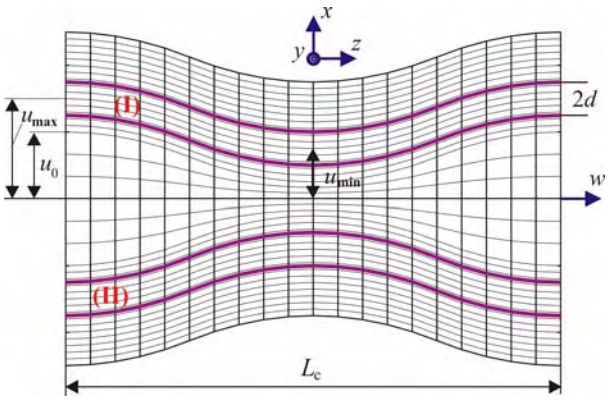


Fig. 7. Geometry of 3D coupler in SR co-ordinate system.

The total coupling length L_c is calculated, with both rectangular and SR based FD-BPM algorithms, for a 3D symmetrical rib waveguide coupler of the cosine type SR geometry and with (u, w) plane defined as in Fig. 7. Rib waveguides parameters are $n_f = 1$, $n_c = 3.44$, $n_s = 3.4$, $H = 1\mu$m, $h = 0.5\mu$m, $2d = 3\mu$m, $\lambda = 1.15\mu$m. The total coupling of the fundamental TE mode ($n_{ref_{TE}} = 3.41313$) is obtained to be $L_{c_{TE}} = 4826\mu$m, with $u_{max} = 3.0\mu$m and $u_{min} = 1.8\mu$m. For TM-mode ($n_{ref_{TM}} = 3.41161$) the total coupling length $L_{c_{TM}} = 4769\mu$m is obtained.

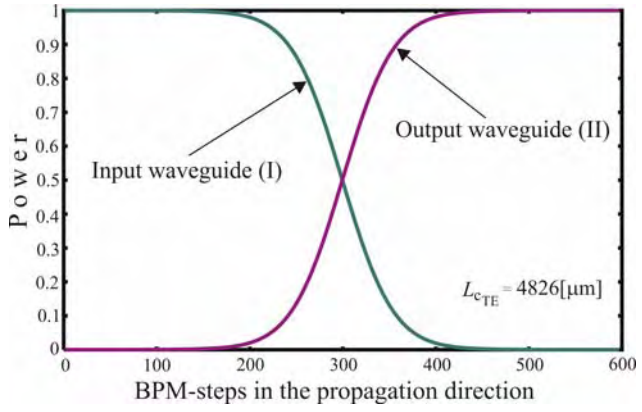


Fig. 8. Evolution of the total power transfer of the fundamental TE-mode field during the propagation in 3D directional coupler, [9].

The total power transfer evolution of the fundamental TE-field, using the *H*-field formulation of the algorithm, is shown in Fig. 8. The almost complete power transfer occurs in the coupler region where rib waveguides are closest to each other.

The SR scheme enables more accurate 3D simulations of the total coupling. In the SR scheme the simulation time is considerably shorter for the same order of accuracy in comparison to the rectangular FD-schemes. This leads to the conclusion that the accuracy of the BPM simulations depends strongly and mostly on the method used for the FD-discretisation of the dielectric waveguide boundaries.

III. Conclusion

Features of the co-ordinate transformation based structure related beam propagation method have been addressed and reviewed. The main advantage of the SR-based method is the exact modelling of the (transverse or/and longitudinal) geometry of the structure under analysis. Presented results obtained in SR-FD-BPM waveguide and directional coupler fundamental mode field simulations demonstrate the advantages and generality of SR over standard rectangular approach. The SR-FD-BPM method enable design flexibility, offering savings in both computational time and memory, thus it is an ideally suited method for optoelectronic CAD design.

References

[1] R. Scarmozzino, A. Gopinath, R. Pregla and S. Helfert, "Numerical techniques for modeling guided-wave photonic devices", IEEE J. of Sel. Top. In Quant. Electr., vol. 6, no. 1, pp. 150-162, 2000.

[2] C.L. Xu and W.P. Huang, "Finite-difference beam propagation method for guide-wave optics", PIER 11, pp. 1-49, 1995.

[3] T.M. Benson, P. Sewell, A. Vukovic, D.Z. Djurdjevic, "Advances in the finite difference beam propagation method", In: Proc. IEEE on Transp. Opt. Netw., vol. 2, pp. 36-41, 2001.

[4] T.M. Benson, D.Z. Djurdjevic, A. Vukovic and P. Sewell, "Towards numerical vector Helmholtz solutions in integrated photonics", In: Proc. IEEE on Transp. Opt. Netw., vol. 2, pp. 1-4, 2003.

[5] T. M. Benson, P. Sewell, S. Sujecki, and P. C. Kendall, "Structure related beam propagation", Optical and Quantum Electronics, vol. 31, pp. 689-703, 1999.

[6] D. Z. Djurdjevic, P. Sewell, T.M. Benson, and A. Vukovic, "Highly efficient finite-difference schemes for structures of nonrectangular cross-section", Microwave and Optical Technology Letters, vol. 33, no. 6, pp. 401-407, 2002.

[7] D.Z. Djurdjevic, P. Sewell, T.M. Benson, A. Vukovic, "Design of photonics structures with non-orthogonal cross-sections using structure-related finite difference methods", In: Proc. SIOE'02, Cardiff, Wales, 2002.

[8] D.Z. Djurdjevic, T.M. Benson, P. Sewell and A. Vukovic, 3D analysis of waveguide couplers using a structure related beam propagation algorithm, In: Proc. OSA/IEEE Integrated Photonics Research, Washington D.C, USA, pp. 137-139, 2003.

[9] D.Z. Djurdjevic, T.M. Benson, P. Sewell, and A. Vukovic, "Fast and accurate numerical analysis of 3D curved waveguide couplers", IEEE Journal of Lightwave Technology, vol. 22, no. 10, pp. 2333-2340, 2004.

[10] P. Sewell, T.M. Benson, A. Vukovic, D.Z. Djurdjevic, J.G. Wykes, Computational issues in the simulation of reflective interactions in integrated photonic components, In: Proc PIER Symposium, Pisa, Italy, 2004.

A Novel Approach for Finite Difference Treatment of Dielectric Interfaces in Photonics

Dušan Ž. Djurdjević[1]

***Abstract* – A general methodology that allows the derivation of high accurate finite difference (FD) formulas is presented. The methodology is made under a power series expansion of the transverse field components of the electromagnetic field. The use of derived FD formulas enables highly accurate modelling of dielectric interfaces and they have been validated in the electric field computations in electrostatics and full-vectorial waveguide simulation in photonics. Some illustrative results from full-vectorial waveguide eigenmode analysis in the frequency domain are presented.**

***Keywords* – Finite difference method, Helmholtz's equation, Numerical simulation, Photonics, Optoelectronics.**

I. INTRODUCTION

Amongst numerous techniques developed for solving partial differential equations, the finite difference method (FDM) is one of the most often-used numerical and simulation tool, [1]. FDM is ideally suited and often a favorite approach for CAD simulation software in microwaves, photonics (lightwave electromagnetics) and optoelectronics, because of its simplicity and flexibility.

In general, FDM technique appears in electromagnetics within two conceptually different approaches: the frequency domain FDM and the time domain FDM. Standard FDM usually refers to time and frequency independent problems described by Laplace's and Poisson's equations. FDM in the frequency domain is often-used approach in waveguide propagation numerical analysis in photonics, namely, as the finite difference beam propagation method (FD-BPM). Generally, the FD-BPM is a particular FDM technique in the frequency domain for the numerical finite difference solution of an exact vector Helmholtz's wave equation, [2], [3].

Standard FDM and the frequency domain FDM are usually implemented in a rectangular co-ordinate system. Certain restrictions arise when the FDM is applied to the structures with oblique or curved dielectric interfaces, because of the inevitable staircase approximation of the boundaries that occurs during the finite difference discretisation procedure. Many approaches have been proposed to improve the efficiency and accuracy of the FDM discretisation of the field near the dielectric interfaces. In area of optics engineering and research, those approaches are well known as the improved FD schemes, [4], [5].

In this paper, a new approach, which can be applied in standard FDM (electrostatics) and in frequency domain FD-BPM, in a uniform dielectric region, at a step-like dielectric interface of an arbitrary shape and near dielectric corner points, is presented. New FD-formulas are derived and used in numerical field calculations and BPM simulations. The results demonstrate the advantages and generality of the presented approach over standard and improved FD-approaches.

[1]Dušan Ž. Djurdjević is with the Faculty of Technical Sciences, Knjaza Miloša 7, 38220 Kosovska Mitrovica, Serbia, E-mail: drdusan.djurdjevic@gmail.com

II. OUTLINE OF THE METHOD

A. Theoretical Foundation – Helmholtz's wave equation

In linear and isotropic media, under the assumption that propagation is in perfectly insulating and z invariant media of refractive index n, in terms of the transverse electric field vector $\vec{\mathbf{E}}_\mathbf{t} = \vec{\mathbf{E}}_\mathbf{t}(x, y, z)$, Helmholtz's equation has a well known form, [2], [3],

$$\nabla^2 \vec{\mathbf{E}}_\mathbf{t} + k^2 n^2 \vec{\mathbf{E}}_\mathbf{t} = \nabla_t \left[\nabla_t \cdot \vec{\mathbf{E}}_\mathbf{t} - \frac{1}{n^2} \nabla_t (n^2 \vec{\mathbf{E}}_\mathbf{t}) \right], \quad (1)$$

where $k = \omega\sqrt{\varepsilon_0 \mu_0}$, $n = \sqrt{\varepsilon_r}$ and a differential operator ∇ is replaced as a sum of transverse and longitudinal part, $\nabla = \nabla_t + \nabla_z$. The similar equation can be derived for the transverse magnetic field vector $\vec{\mathbf{H}}_\mathbf{t}$.

The BPM is developed under paraxial approximation of Eq. (1), [2], [3]. Assuming that the forward traveling field of typical photonics waveguide-based structures has rapid phase variations along the guiding axis z and a slowly-varying vector envelope $\mathbf{E}_t$, one can rewrite Eq. (1) as,

$$j2kn_0 \frac{\partial \mathbf{E}_t}{\partial z} = \nabla_t \left[\frac{1}{n^2} \nabla_t (n^2 \mathbf{E}_t) \right] + k^2 (n^2 - n_0^2)\, \mathbf{E}_t, \quad (2)$$

where $j = \sqrt{-1}$, $\vec{\mathbf{E}}_\mathbf{t} = \mathbf{E}_t e^{-j\beta z} = \mathbf{E}_t e^{-jkn_0 z}$ and $\beta = kn_0$ is the background propagation constant with n_0 denotes a reference (modal) index. Eq. (2) is known as the paraxial full-vectorial approximation of Eq. (1).

In the 2D electrostatic approximation $(k \to 0,\ \partial/\partial z = 0)$, for the transverse vector $\vec{\mathbf{E}}_\mathbf{t} = \vec{\mathbf{E}}_\mathbf{t}(x, y)$, Eq. (1) simplifies to

$$\nabla_t \left[\frac{1}{n^2} \nabla_t (n^2 \vec{\mathbf{E}}_\mathbf{t}) \right] = 0, \quad (3)$$

where $n = \sqrt{\varepsilon_r}$. Eqs. (1), (2) and (3) contain a term,

$$\nabla_t \left[\frac{1}{n^2} \nabla_t (n^2 \mathbf{\Phi}_\mathbf{t}) \right], \quad (4)$$

where $\mathbf{\Phi_t}$ can be either vector $\vec{\mathbf{E}}_\mathbf{t}$ or slowly varying envelope vector $\mathbf{E_t}$. If n is a continuous function in the transverse plain (graded index cases), a differential term in Eq. (4) can be calculated analytically. However, if n is a discontinuous function (step-index cases), analytical solution is not possible, so there is a serious problem how to obtain consistent difference equations. The improved FD formulas approaches, [4], [5], take into account boundary conditions for the field and its derivatives near the dielectric interfaces. Approach presented in this paper improves FD-discretisation near step-like dielectric interfaces, but in a different, specific way, [6].

B. Derivation of FD-formulas, 2nd order of accuracy

Power series expansions of the electric field components as the functions of two Cartesian co-ordinates x and y are

$$E_x(x,y) = a_0 + a_1 x + a_2 y + a_3 x^2 + a_4 xy + a_5 y^2 + \cdots, \quad (5)$$

$$E_y(x,y) = b_0 + b_1 x + b_2 y + b_3 x^2 + b_4 xy + b_5 y^2 + \cdots. \quad (6)$$

In linear and isotropic source-free media, in the cases of 2D (xy) static field (e.g. electrostatics) and TEM field (e.g. transmission line field), from Maxwell's equations,

$$\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} = 0, \qquad \frac{\partial E_x}{\partial y} - \frac{\partial E_y}{\partial x} = 0, \quad (7)$$

we can evaluate power series expansion coefficients in Eqs. (5) and (6),

$$\mathbf{E}_x = a_0 + a_1 x + a_2 y + a_3 x^2 + a_4 xy - a_3 y^2 + \cdots, \quad (8)$$

$$\mathbf{E}_y = b_0 + a_2 x - a_1 y + \frac{a_4}{2} x^2 - 2a_3 xy - \frac{a_4}{2} y^2 + \cdots. \quad (9)$$

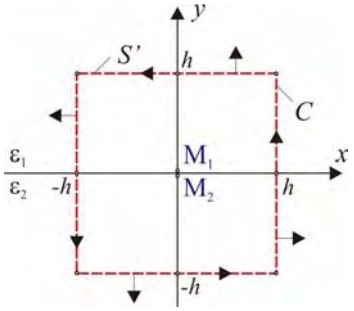


Fig. 1. Square 2D cell placed symmetrically at the interface of two dielectric media. Points M_1 and M_2 are placed close to interface. S' and C orientations are indicated with arrows.

If we assume that the dielectric interface has no charge, by applying the integral form of the Gauss' law on the Gaussian surface S (a contour S' in Fig. 1 is a cross-section of S in the transverse plane), and by using Eqs. (8) and (9), one can derive, [6],

$$\varepsilon_{r1} E_y\big|_{M_1} - \varepsilon_{r2} E_y\big|_{M_2} = \frac{h^2}{6}\left[\varepsilon_{r2}\frac{\partial^2 E_x}{\partial x\,\partial y}\bigg|_{M_2} - \varepsilon_{r1}\frac{\partial^2 E_x}{\partial x\,\partial y}\bigg|_{M_1}\right] + O(h^2). \quad (10)$$

By integrating electric field over contour C in Fig. 1, the similar formula to Eq. (10) can be obtained, [6],

$$E_x\big|_{M_1} - E_x\big|_{M_2} = \frac{h^2}{6}\left[\frac{\partial^2 E_y}{\partial x\,\partial y}\bigg|_{M_1} - \frac{\partial^2 E_y}{\partial x\,\partial y}\bigg|_{M_2}\right] + O(h^2). \quad (11)$$

When dielectric interface is arbitrary placed between two subsequent grid lines, integrations over S and C, Fig. 2, yield

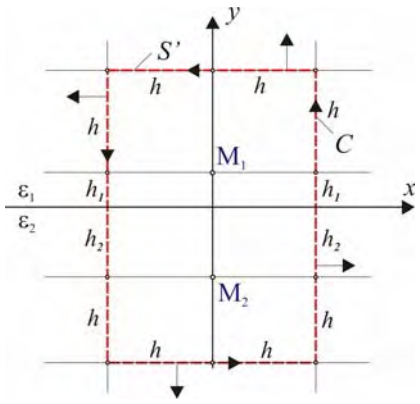


Fig. 2. Non-square 2D rectangular cell arbitrary placed at the interface of two dielectric media. Points M_1 and M_2 are placed at two subsequent grid lines.

$$\varepsilon_{r1} E_y\big|_{M_1} - \varepsilon_{r2} E_y\big|_{M_2} = \frac{h^2}{6}\left[(p_2 a_1 + q_2 a_4)_{M_2} - (p_1 a_1 + q_1 a_4)_{M_1}\right] + O(h^2), \quad (12)$$

and

$$E_x\big|_{M_1} - E_x\big|_{M_2} = \frac{h^2}{3}\left[(r_1 a_2 + s_1 a_3)_{M_1} - (r_2 a_2 + s_2 a_3)_{M_2}\right] + O(h^2). \quad (13)$$

In Eqs. (12) and (13) coefficients a_i, $i = 1, \cdots 4$, denote first and second derivatives of the field components in points M_1 and M_2, and coefficients p_i, q_i, r_i and s_i, $i = 1, 2$, are

$$p_i = -\varepsilon_{ri}\frac{6h_i}{h^2}, \quad q_i = \varepsilon_{ri}\left(1 - \frac{3h_i^2}{h^2}\right), \quad (14)$$

$$r_i = -\frac{6h_i}{h^2}, \quad s_i = \left(1 - \frac{3h_i^2}{h^2}\right). \quad (15)$$

In Eqs. (14) and (15), $h = \triangle = h_1 + h_2$ is a distance between two subsequent grid lines in the rectangular FD-mesh adopted, uniformly spaced both in x and y directions, Fig. 2.

C. Derivation of FD-formulas, 5th order of accuracy

Formulas given in Eqs. (10), (11), (12) and (13) are second order accurate. Formulas with higher order of accuracy can be derived in the similar way, by extending the order of power

series in Eqs. (5) and (6). The uniform rectangular grid scheme, in (x, y) co-ordinate system, used in derivation of the finite difference formulas near a dielectric interface, is shown in Fig. 3. The field components can be expanded in Eqs. (5) and (6) up to the 5th order, over the stencil diagramed in Fig. 3, resulting in 10 linear algebraic equations. Their solution yields FD-formulas for E_{x_0}, E_{y_0} and their derivatives with 5th order of accuracy. In Eqs. (16) to (21) 1st and 2nd derivatives are given only.

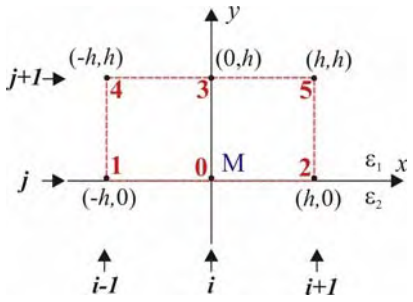


Fig. 3. Uniform rectangular scheme used in the finite difference formulas derivation. The point M is above the dielectric interface.

$$E_{x_0} = \frac{1}{10}[3E_{y_1} - 3E_{y_2} + 3E_{y_4} - 3E_{y_5} - (E_{x_1} + E_{x_2} - 10E_{x_3} - E_{x_4} - E_{x_5})] + O(h^5), \quad (16)$$

$$E_{y_0} = -\frac{1}{10}[3E_{x_1} - 3E_{x_2} + 3E_{x_4} - 3E_{x_5} + (E_{y_1} + E_{y_2} - 10E_{y_3} - E_{y_4} - E_{y_5})] + O(h^5), \quad (17)$$

$$a_1 = \frac{1}{10h}[(20E_{y_0} - 3E_{y_1} - 3E_{y_2} - 10E_{y_3} - 2E_{y_4} - 2E_{y_5}) + (E_{x_1} - E_{x_2} + E_{x_4} - E_{x_5})] + O(h^5), \quad (18)$$

$$a_2 = \frac{1}{10h}[(20E_{x_0} - 3E_{x_1} - 3E_{x_2} - 10E_{x_3} - 2E_{x_4} - 2E_{x_5}) - (E_{y_1} - E_{y_2} + E_{y_4} - E_{y_5})] + O(h^5), \quad (19)$$

$$a_3 = \frac{1}{4h^2}[(-10E_{x_0} + 2E_{x_1} + 2E_{x_2} + 4E_{x_3} + E_{x_4} + E_{x_5}) + (2E_{y_1} - 2E_{y_2} + E_{y_4} - E_{y_5})] + O(h^5), \quad (20)$$

$$a_4 = \frac{1}{2h^2}[-(10E_{y_0} - 2E_{y_1} - 2E_{y_2} - 4E_{y_3} - E_{y_4} - E_{y_5}) + (-2E_{x_1} + 2E_{x_2} - E_{x_4} + E_{x_5})] + O(h^5). \quad (21)$$

Similar FD-formulas can be derived for the stencils where the point M is set to be below, left and right to the dielectric interface. Combining those formulas with Eqs. (10) to (13) one can obtain FD formulas with high accuracy.

III. Numerical Results and Discussion

Performed test-computations in electrostatics have confirmed that derived finite FD formulas have truncation error proportional to the grid size, with the expected 5th order of accuracy. Some results obtained in simple FD-BPM test-simulations of buried waveguides are presented, [6].

The cross section of a simple buried dielectric waveguide is shown in Fig. 4. The most realistic cases - the full-vectorial real-axis and imaginary-axis beam propagation have been considered. The full-vectorial FD-BPM is adopted because of the presence of both transverse field components. Much simpler scalar and polarized formulations of Eq. (1) can not lead to the appropriate conclusions. Only TM propagation (E-field formulation) has been used in simulations.

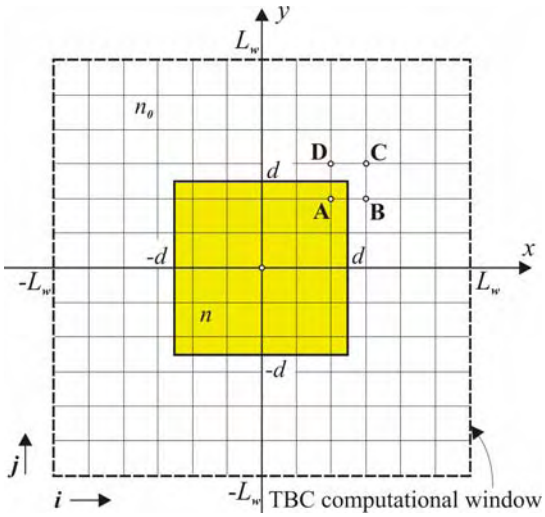


Fig. 4. Cross section of the buried waveguide. Dielectric interfaces are placed at the middle between two grid lines.

The numerical BPM simulation and the reference (modal) index evaluation are highly dependent on the electric field discretisation accuracy near the step-index interfaces. The approach presented in Section II is valid for the static and TEM fields. In TM propagation is $H_z = 0$, $\partial E_z/\partial z \cong 0$ (steady-state field regime), therefore the presented approach and derived FD-formulas can be used in TM case as well.

To avoid the influence of corners on the overall result, [7], FD-meshing in the transverse plane have been done by placing interfaces at the middle between two grid lines. For the field in points A, B, C and D, Fig. 4, the FD-formulas, with truncation error of order 3, have been derived and used:

$$E_{x_A} = E_{x\,i,j} = E_{x\,i-1,j-1} - E_{y\,i-1,j} + E_{y\,i,j-1} + O(h^3), \quad (22)$$

$$E_{y_A} = E_{y\,i,j} = E_{y\,i-1,j-1} + E_{x\,i-1,j} - E_{x\,i,j-1} + O(h^3), \quad (23)$$

$$E_{x_B} = E_{x\,i,j} = E_{x\,i+1,j-1} + E_{y\,i+1,j} - E_{y\,i,j-1} + O(h^3), \quad (24)$$

$$E_{y_B} = E_{y\,i,j} = E_{y\,i+1,j-1} - E_{x\,i+1,j} + E_{x\,i,j-1} + O(h^3), \quad (25)$$

$$E_{x_C} = E_{x\,i,j} = E_{x\,i+1,j+1} - E_{y\,i+1,j} + E_{y\,i,j+1} + O(h^3), \quad (26)$$

$$E_{y_C} = E_{y\,i,j} = E_{y\,i+1,j+1} + E_{x\,i+1,j} - E_{x\,i,j+1} + O(h^3), \quad (27)$$

$$E_{x_D} = E_{x\,i,j} = E_{x\,i-1,j+1} + E_{y\,i-1,j} - E_{y\,i,j+1} + O(h^3), \quad (28)$$

$$E_{y_D} = E_{y\,i,j} = E_{y\,i-1,j+1} - E_{x\,i-1,j} + E_{x\,i,j+1} + O(h^3). \quad (29)$$

In the uniform regions, the standard five-point FD formula has been used. A standard Crank-Nicolson algorithm has been used to simulate propagation in the z direction, and due to the well-confined waveguide field, transparent boundary conditions (TBC), [8], have been introduced at the edges of the computational window. For eigenmode solving, the imaginary distance BPM algorithm has been applied, [9].

TABLE I
CALCULATED MODAL INDEX AND CPU RUN-TIME USED IN IMAGINARY DISTANCE PROPAGATION

mesh size	$\triangle[\mu m]$	n_{ref}	$error_{rel.}$	CPU time
42×42	0.12500	1.222253	0.000727	00:01
82×82	0.06250	1.275934	0.000231	00:07
122×122	0.04166	1.274600	0.000125	00:22
162×162	0.03125	1.274077	0.000084	00:51
202×202	0.02500	1.271807	0.000063	01:36
242×242	0.02083	1.271645	0.000050	02:39
282×282	0.01785	1.271538	0.000041	05:04
322×322	0.01562	1.271462	0.000036	06:55
362×362	0.01389	1.271405	0.000031	09:50
402×402	0.01250	1.271361	0.000028	11:00

Numerical simulations have been performed for two different buried waveguide structures: low-index contrast one, with $n = 1.5$ $(\varepsilon_r = n^2 = 2.25)$, $2d = 1\,\mu$m, and the strong-index contrast waveguide, with $n = 3.4$ $(\varepsilon_r = 11.56)$, $2d = 0.5\,\mu$m. The wave-length has been kept at $\lambda = 1.5\,\mu$m in both cases. BPM step has been chosen to be $\triangle z = 0.1\,\mu$m, dimensions of the square computational domain have been truncated at $L_w = 2.5\,\mu$m in low, and at $L_w = 1.25\,\mu$m in strong-index contrast simulations.

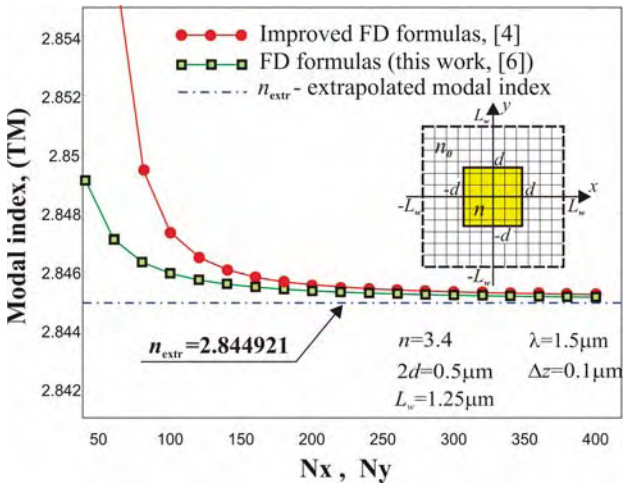


Fig. 5. Calculated modal index versus mesh size for waveguide shown in inset of Figure, strong-index contrast case.

Results for n_{ref} obtained in the imaginary distance BPM simulations in low-index contrast case are tabulated in Table I, together with data showing the total CPU use (in [min:s]) during the C++ code execution on PC (32-bit OS, 2.0 GHz) and the relative error computed against the extrapolated value for n_{ref} ($n_{ref_{extr.}} = 1.271301$, low-index contrast case, $n_{ref_{extr.}} = 2.844921$, strong-index contrast case). The CPU data represent the necessity of the accurate FD discretisation, since the mesh refinement implies a drastic increase in computer run-time.

The computed data for TM modal index n_{ref} versus number of grid lines in the x (N_x), and the y (N_y) directions, strong-index contrast case, are shown in Fig 5. Comparison with results obtained by using method published in [4], Fig. 5, shows that the present discretisation model demonstrates the faster convergence as the grid size is reduced. This result has been expected, as a consequence of the intrinsic property of the presented methodology enabling the true two-dimensional FD-discretising of the field near the dielectric interfaces and corners.

IV. CONCLUSION

The approach for finite difference treatment of dielectric interfaces has been presented. FD-formulas have been derived and applied to the electromagnetic field analysis and BPM simulations in photonics. Numerical simulations have been carried out to solve eigenmodes of TM mode field via the imaginary distance BPM for buried waveguide in the rectangular co-ordinate system, and results have been compared with those obtained from simulations based on previously reported improved FD algorithms. The flexibility of the approach, efficiency and accuracy of the derived FD-formulas recommends them for use in photonic CAD design.

REFERENCES

[1] C. Pozrikidis, *Numerical computation in Science and Engineering*, New York, Oxford University Press, 1998.

[2] C.L. Xu and W.P. Huang, "Finite-difference beam propagation method for guide-wave optics", PIER 11, pp. 1-49, 1995.

[3] T.M. Benson, D.Z. Djurdjevic, A. Vukovic and P. Sewell, "Towards numerical vector Helmholtz solutions in integrated photonics", In: Proc. IEEE on Transparent Optical Networks, vol. 2, pp. 1-4, 2003.

[4] Y.P.Chiou, Y.C. Chiang and H.C. Chang, "Improved three-point formulas considering the interface conditions in the finite-difference analysis of step-index optical devices", IEEE Journal of Lightwave Technology, vol. 18, no. 2, pp. 243-251, 2002.

[5] G.R. Hadley, "High-accuracy finite-difference equations for dielectric waveguide analysis I: Uniform regions and dielectric interfaces", IEEE Journal of Lightwave Technology, vol. 20, no.7, pp. 1210-1218, 2002.

[6] D.Z. Djurdjevic, "New finite-difference formulas for dielectric interfaces", Facta Universitatis (Niš), Ser.: Elec. Energ, vol. 23, no. 1, pp. 17-35, 2010.

[7] G.R. Hadley, "High-accuracy finite-difference equations for dielectric waveguide analysis II: Dielectric corners", IEEE J. of Lightwave Technology, vol. 20, no.7, pp. 1219-1231, 2002.

[8] G. R. Hadley, "Transparent boundary condition for beam propagation", Opt.Lett., vol.16, pp. 624-626, 1991.

[9] D.Yevick and B.Hermansson, "New formulations of the matrix beam propagation method: Application to rib waveguides", IEEE Journal of Quant. Elec., vol.25, pp. 221-229, 1989.

Theoretical Model of the MUD Algorithm Based on the Diagram of States and Markov's Chains

Iliya Georgiev Iliev[1] Marin Nedelchev[2] Boyan Ruzhkov Kehayov[3]

***Abstract* – The paper presents a development of a Markov's chains of the work of quasioptimal algorithm for multiuser detection. Through it identifies: functional accuracy of the algorithm-respectively the lower limit of the error probability in a channel with Rayleigh fading and its computational complexity. Program models are realized in MATLAB, in order to research the algorithm performance and their parameters. Graphic presentations from the measurements are given to prove the exactness of the analytical formulas.**

Keywords – MUD, CDMA, Markov chain, diagram of states.

I. INTRODUCTION

CDMA is an efficient method for sharing a mobile radio channel. The noise resistance of the system decreases due to internal system interference – MUI (multi user interference). The correlation receiver is optimal when no MUI exists. The multi user detection (MUD) is used to minimize the influence of the MUI [1].

The MUD receiver, based on the maximum likelihood criteria (ML) [1], gives an optimal solution and checks all possible combinations of transmitted symbols. The number of the calculations grows exponentially with the number of the active users, which is a disadvantage of ML MUD. There are many suggested methods and algorithms for suboptimal receiving, decreasing the needed number of detection calculations. They are compromise between calculation complexity and quality parameters of the receiver. [8,9].

The research is focused on the parameters and the possibilities of the quasioptimal algorithm of MUD [8,9], based on the serial search. The fast-response of the algorithm is due to strong criteria of search discontinuation and selection of start point of the optimization after the single correlation receiving.

A diagram of states and Markov's chains is used for modeling the process of the algorithm. The result of the analysis is the error probability of the algorithm, the dependency of its accuracy from the number of iterations and users.

[1] Ilia Iliev is with the Faculty of Telecommunications, TU-Sofia, Kliment Ohridski 8, Sofia 1000, Bulgaria, E-mail: igiliev@tu-sofia.bg

[2] Marin Nedelchev is with the Faculty of Telecommunications, TU-Sofia, Kliment Ohridski 8, Sofia 1000, Bulgaria, E-mail: mnedelchev@tu-sofia.bg

[3] Boyan Kehayov is with the Faculty of Telecommunications, TU-Sofia, Kliment Ohridski 8, Sofia 1000, Bulgaria, E-mail: bkehayov@tu-sofia.bg

II. MUD MODEL IN SYNCRONIOUS CDMA SYSTEM

A block diagram of the system is shown on figure 1. The signal processing is made in baseband.

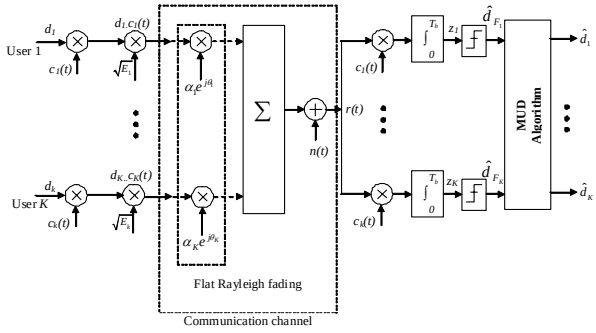


Fig. 1. Block diagram of MUD

The model is described in details in [8]. Here *K* users are using synchronous transition with direct spread spectrum – DSS and 2PSK modulation.

After applying criteria for maximum of aposterior probability (MAP), optimal MUI is achieved. A logarithmic function of likelihood is presented in matrix form [1]:

$$\Psi(\boldsymbol{d}) = \boldsymbol{d}^T \boldsymbol{EARA}^* \boldsymbol{Ed} - 2\Re(\boldsymbol{d}^T \boldsymbol{EA}^* \boldsymbol{z}) \qquad (1).$$

Base components in (1) are:

$\boldsymbol{d} = [d_1, d_2, ..., d_K]^T$ - a matrix, containing data, transmitted from user *k*;

$\boldsymbol{A} = diag[\alpha_1 e^{j\theta_1}, \alpha_2 e^{j\theta_2}, ..., \alpha_K e^{j\theta_K}]$ - a diagonal matrix with complex coefficients of transmission channel for the corresponding user. The amplitudes are with Rayleigh distribution. The phase shifting is with normal distribution in [0, 2π]. The channels of all users are statistically independent;

$\boldsymbol{E} = diag\left[\sqrt{E_1}, \sqrt{E_2}, ..., \sqrt{E_K}\right]$ - a diagonal matrix. $\sqrt{E_k}$ is the symbol energy of user *k*;

$\boldsymbol{c} = [c_1(t), c_2(t), ..., c_K(t)]$ - a matrix, where every row is a pseudorandom binary sequence (PBS) for the corresponding user;

$\boldsymbol{R}$ - a cross-correlation *KxK* matrix, which coefficients are the values of the normalized cross-correlation functions of PBS.

$\boldsymbol{n} = [n_1, n_2, ..., n_K]^T$ - Gaussian noise after the correlator, with the covariance matrix $\boldsymbol{R}_n = 0.5 N_o \boldsymbol{R}$.

III. THEORETICAL MODEL OF WORK OF THE MUD ALGORITHM BASED ON THE DIAGRAM OF STATES AND MARKOV'S CHAINS

For simplicity, we will consider the following conditions:

1. The communication channels are without additive white Gaussian noise. This assumption is made in order to calculate the error probability rate of the algorithm caused by its quasioptimal solution;
2. The power of the received signals from all users is equal;
3. Bipolar code sequences with uniform distribution of the values ±1 are used for DSS;
4. The process gain is equal to the length of the random sequence N, when it is sufficiently long.

The work of the algorithm could be described with a cell diagram, where the value of the target function $\Psi(\boldsymbol{d})$ is included.

Figure 2 shows a possible cell diagram for K=3. Each junction represents a data vector $\boldsymbol{d}$ located away from the optimal point $\boldsymbol{d_0}$ with certain code distance. The problem is to find the minimum of $\Psi(\boldsymbol{d})$.

For a given initial vector (point) $\boldsymbol{d_{l-1}}$ the algorithm goes through the vector space with Hamming distance H_d=1 with consecutive changes of each element of the vector [8]. For each vector from that space, $\Psi(\boldsymbol{d})$ is calculated. This represents a step of the algorithm to find the next point which is closer to transmitted vector $\boldsymbol{d_0}$.

The solution of the algorithm for the $\boldsymbol{l}$-th iteration is calculated with:

$$\boldsymbol{d_l} = \arg\left\{\min_{\boldsymbol{d}\in\boldsymbol{M_d}}\left[\Psi(\boldsymbol{d})\right]\right\} \text{ и } \quad M_d=\left\{d: H_d\left(\boldsymbol{d_{l-1}},\boldsymbol{d}\right)=1\right\} \qquad (2).$$

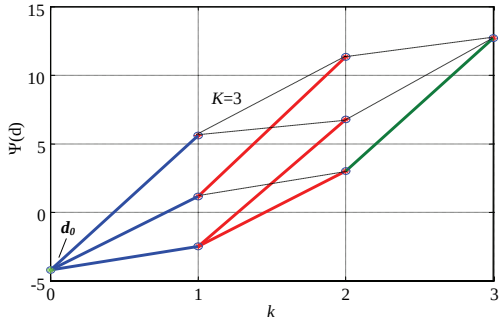


Fig. 2. Cell diagram for K=3

One step of iteration is inverting the value of one element of vector $\boldsymbol{d_{l-1}}$. The new vector $\boldsymbol{d_l}$ is defined. The attempt is successful when:

$$\max_{\boldsymbol{d}\in\boldsymbol{M_d}}\left\{[\Psi(\boldsymbol{d_{l-1}})-\Psi(\boldsymbol{d_l})]>0\right\} \text{ и } H_d(d_{l-1},d_0)>H_d(d_l,d_0) \qquad (3).$$

If the algorithm is in point $\boldsymbol{d_{l-1}}$, during the iteration, the following hypotheses are possible:

1. H_{+1} - **step forward** – A vector $\boldsymbol{d_l}$ is found, which code distance to the optimal point $\boldsymbol{d_0}$ is less than the one to $\boldsymbol{d_{l-1}}$. The hypothesis is true when the below conditions are satisfied:

   $\max_{\boldsymbol{d}\in\boldsymbol{M_d}}\left\{[\Psi(\boldsymbol{d_{l-1}})-\Psi(\boldsymbol{d_l})]>0\right\}$ и $H_d(d_{l-1},d_0)>H_d(d_l,d_0)$;
2. H_{-1} - **step backward** – The vector $\boldsymbol{d_l}$ has longer code distance than $\boldsymbol{d_{l-1}}$. The necessary conditions are:

   $\max_{\boldsymbol{d}\in\boldsymbol{M_d}}\left\{[\Psi(\boldsymbol{d_{l-1}})-\Psi(\boldsymbol{d_l})]>0\right\}$ и $H_d(d_{l-1},d_0)<H_d(d_l,d_0)$;
3. H_0 - **an extremum is reached** – $\boldsymbol{d_l}=\boldsymbol{d_{l-1}}$, $[\Psi(\boldsymbol{d_{l-1}})-\Psi(\boldsymbol{d_l})]\le 0$ for $\boldsymbol{d}\in\boldsymbol{M_d}$.

The extremum can be local, when the cost function is not unimodal. The global extremum corresponds to the transmitted vector. The local extremum may not be the same as the global extremum $\boldsymbol{d_0}$, because the algorithm is quasioptimal. If the impact of AWGN is ignored, the algorithm fully compensates the cross-channel interference if the point $\boldsymbol{d_{l-1}}$ has a code distance 1 to the global extremum $\boldsymbol{d_0}$, and then with the next iteration the global extremum would be found.

The algorithm performance, error decision probability and the minimal limit error probability with defined number of operations can be found with the help of the probability theory.

The probability to get to the global extremum can be specified by describing the algorithm by Markov chain. During iteration the probability for transition from one point to another can be described by directed graph shown on Figure 3.

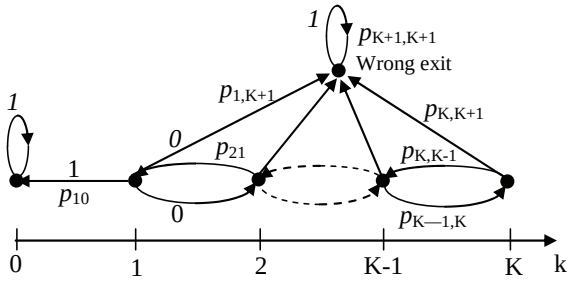


Fig. 3. Directed graph

Every point corresponds to a junction and is located on fixed code distance from the global extremum. The number of the junction is equal to the number of the users K+1. Junction 0 represents the transmitted vector, which is the global extremum $\boldsymbol{d_0}$. The probability p_{10} of passing through junction 1 to 0 is one. If there is no receiving error the algorithm will begin the search with 0 code distance and will remain in this junction – this is the probability p_{00}=1. Due to this condition the Markov chain is reducible. There is one more reducible state when the algorithm reaches a local extremum that is different from the global extremum. On figure 3 this state is marked with index K+1 – „Wrong exit". The algorithm will cancel the search and the received vector is with errors.

$$[P_M]=\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & L & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & L & 0 & 0 & 0 & 0 \\ 0 & p_{21} & 0 & p_{23} & 0 & L & 0 & 0 & 0 & p_{2,K+1} \\ 0 & 0 & p_{32} & 0 & p_{34} & L & 0 & 0 & 0 & p_{3,K+1} \\ M & M & M & M & M & L & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & L & p_{K-1,K-2} & 0 & p_{K-1,K} & p_{K-1,K+1} \\ 0 & 0 & 0 & 0 & 0 & L & 0 & 0 & 0 & p_{K,K+1} \\ 0 & 0 & 0 & 0 & 0 & L & 0 & p_{K,K-1} & 0 & 1 \end{bmatrix} \qquad (4).$$

The Markov transition matrix for iteration of the algorithm is described with (4).

The transition probabilities can be defined, if the distribution of the values of the target function is known. Due to the great number of random variables in the matrix it is suggested to operate with equivalent random variable *y*, which probability density is $f_y(y)$. It is the difference between the values of the target function $\boldsymbol{d_{l-1}(k)}$ and $\boldsymbol{d_l(k-1)}$, where the argument is the code distance to $\boldsymbol{d_0}$ and:

$$y_l(k) = \Psi(\boldsymbol{d_{l-1}}(k)) - \Psi(\boldsymbol{d_l}(k-1)) \qquad (5).$$

Once (5) is converted and the code distance between the starting point for an iteration and transmitted vector is $H_d(d_{l-1},d_0)=k$, (6) is obtained:

$$y_l(k) = 2(\sum_{i=1}^{k-1} z_i - 2\chi^2) \qquad (6),$$

where χ is a random variable with χ^2 distribution and variance $\sigma^2 = 1$ comes from the fact that the channel has Rayleigh distribution of transmission coefficient. Random variable z_i is: $z_i = \alpha_m \alpha_n \overline{r}_{mn}$, $\alpha_{m,n}$ takes values ±2 and discrete uniform distribution – depends on the transmitted symbols. Other random variable is:

$$\overline{r}_{mn} = R_{mn}[\mathrm{Re}(C_m)\mathrm{Re}(C_n) + \mathrm{Im}(C_m)\mathrm{Im}(C_n)],$$

where distribution of values of cross correlation function R_{mn} can be approximated with normal distribution with variance $\sigma^2=1/N$. The real part $\mathrm{Re}(C_m)$ is with normal distribution $\mu=0$ and mathematical expectation $\sigma^2=0.5$.

Analytical determination of the density distribution of $y_l(k)$ $f_y(y(k))$ is not an easy task, so computer modeling with method Monte Carlo is proposed. f_y distributions are independent and identical for each point, which is on the same code distance from $\boldsymbol{d_0}$.

Correct decision for the *l*-th iteration will occur (Hypothesis H_+), if both of the following conditions are observed:

1. At least one of the *k* points with code distance $H_d(d_l,d_0)<H_d(d_{l-1},d_0)$ satisfies $[\Psi(\boldsymbol{d_{l-1}}) - \Psi(\boldsymbol{d_l})] > 0$;
2. At least one of total *K-k* points with $H_d(d_l, d_0) > H_d(d_{l-1}, d_0)$ satisfies the condition $[\Psi(\boldsymbol{d_{l-1}}) - \Psi(\boldsymbol{d_l})] < 0$.

The probability of execution of the first and second condition and their independence for iteration (probability of correct decision (Hypothesis H_+)) is:

$$Pd_k = \left\{1 - \left[\int_{-\infty}^{0} f_y(y(k))dy\right]^k\right\}\left\{1 - \left[\int_{-\infty}^{0} f_y(y(k+1))dy\right]^{K-k}\right\} \qquad (7).$$

The probability for not finding a better point during next iteration and the end of the search algorithm (Hypothesis H_0 (an extremum is reached)) depends on the distribution of the random variable $[\Psi(\boldsymbol{d_{l-1}}) - \Psi(\boldsymbol{d_l})] \le 0$ where $\boldsymbol{d} \in \boldsymbol{M_d}$. In iteration, *k* attempts reduce the code distance and *K-k* attempts increase the code distance.

The probability algorithm remains at the same point and the search end is determined by:

$$Po_k = \left[\int_{-\infty}^{0} f_y(y(k))dy\right]^k \left[\int_{-\infty}^{0} f_y(y(k+1))dy\right]^{K-k} \text{ for } k = 1...K \qquad (8).$$

The algorithm's probability to make a mistake - to differ from the optimal point, hypothesis H_{-1}, is equal to:

$$Perr_k = 1 - Po_k - Pd_k \qquad (9).$$

Once the transition probabilities in the diagram of figure 3 are known, the elements of the matrix [Pm] are determined by:

$p_{00} = p_{10} = 1$, $p_{k,k-1} = Pd_k$, $p_{k,K+1} = Po_k$, $p_{k,k+1} = Perr_k$ for $k = 2..K-1$, $p_{K,K-1} = Pd_K$, $p_{K,K+1} = Po_K$, за $k = K$, $p_{K+1,K+1} = 1$.

IV. Functional precision of the algorithm

The probability, after detection with hard decision, for the algorithm to begin a search from an initial vector with code distance *k* is given by [8]:

$$P_K(k) = \frac{K!}{(K-k)!\,k!}\left(0.5 - 0.5/\sqrt{1+(K-1)/N}\right)^k \cdot \left(0.5 + 0.5/\sqrt{1+(K-1)/N}\right)^{K-k} \qquad (10).$$

The vector $[P_K]$, arranged by the $P_K(k)$ for $k=0...K$, is the initial vector of the initial states of the Markov Chain. The vector $[P_L]$, with probabilities of occurrence of *k* fold error ($k=0...K$) after *L* number of iterations, is determined by the matrix equation:

$$[P_L] = [P_K][P_M]^L \qquad (11),$$

The probability of starting the algorithm from the *k*-th code distance is defined by (10). The error probability of the algorithm after iteration is determined by two hypotheses. The first hypothesis is that the algorithm finds a vector with smaller code distance. The second one is to end the search – Wrong exit.

The matrix with error probability after the *L* number of iterations is:

$$[P_{err}]_L = [P_L] + [Po] \qquad [Po] = [Pu(:, K+2)']$$
$$[Pu] = [diag(P_K)][P_M]^L \qquad (12).$$

The vector $[Po]$ is composed by elements of the last column of the matrix $[Pu]$. Due to the independence of the communication channels, the average probability of error per bit after the number *L* of iterations of the algorithm can be calculated with:

$$P_{e_L} = \sum_{k=1}^{L} \frac{k}{K} p_{err}(k) \qquad (13).$$

V. RESULTS OF SIMULATION

For the purposes of the research a model program in MATLAB, which finds the distribution of random variable *y* using the Monte Carlo method, is created.

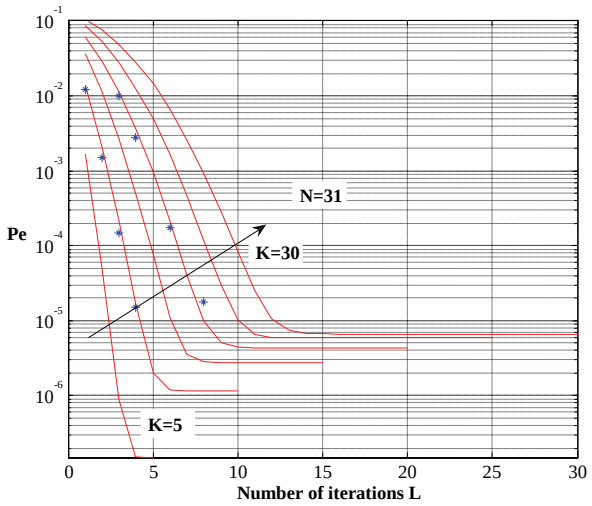


Fig. 4. P_e depending on *L* and *K*

Equation from formula (6) is used and with (4), (8), (9) and (12) the dependency of the probability of error per bit, depending on the number of iterations *L* with the number of users *K* as a parameter, is defined. The results are obtained for DS-CDMA system with transmission and reception of *K* number of users working simultaneously. The system operates in an environment with multipart slow Rayleigh fading. The results are shown on Figure 4. The red curves are theoretical estimates - from formula (6) and the blue dots (star *) are the results of the simulation.

The probability of error per bit, depending on the functional accuracy of the quasioptimal algorithm is shown in figure 5. The dependence is the number of active users *K*.

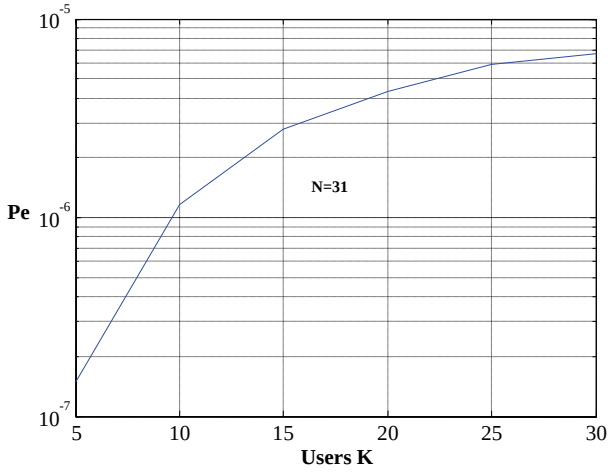


Fig. 5. P_e depending on *K*

VI. CONCLUSION

The paper presents a development of a Markov's chains of the work of quasioptimal algorithm for multiuser detection. Through it identifies: functional accuracy of the algorithm- respectively the lower limit of the error probability in a channel with Rayleigh fading and its computational complexity. Program models are realized in MATLAB, in order to research the algorithm performance and their parameters. Graphic presentations from the measurements are given to prove the exactness of the analytical formulas.

With this approach of modeling the algorithm of MUD with a Markov chains can be determined the average number of iterations needed to reach the extremum.

REFERENCES

[1] S. Verdú, *Multiuser Detection*. New York: Cambridge Univ. Press,1998.

[2] R. Lupas and S. Verd´u, "Linear multiuser detectors for synchronous code-division multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 35, pp. 123-136, Jan. 1989.

[3] Z. Xie, R. Short, and C. Rushforth, "A family of suboptimum detectors for coherent multiuser communications," *IEEE Journal on Selected Areas in Communications*, vol. 8, pp. 683-690, May 1990

[4] C. Ergun and K. Hacioglu, "Multiuser detection using a genetic algorithm in CDMA communications systems," *IEEE Trans. Commun.*, vol. 48, pp. 1374–1383, Aug. 2000.

[5] M. J. Juntti, T. Schl¨osser, and J. O. Lilleberg, "Genetic algorithms for multiuser detection in synchronous CDMA," in *IEEE International Symposium on Information Theory – ISIT'97*, (Ulm, Germany), p. 492, 1997.

[6] K. Yen and L. Hanzo, "Hybrid genetic algorithm based multi-user detection schemes for synchronous CDMA systems," in *Proceedings of the IEEE Vehicular Technology Conference (VTC)*, (Tokyo, Japan), May 15-18,2000.

[7] Peng Hui Tan, Lars K. Rasmussen, Multiuser Detection in CDMA—A Comparison of Relaxations, Exact, and Heuristic Search Methods, *IEEE TR. WIRELESS COMM.*, VOL. 3, N 5, SEPT. 2004

[8] Iliev I. G., Nedelchev M. V., Performance Analysis of a suboptimal multiuser detection algorithm, ICEST 2007, Macedonia, pp.531 - 534

[9] Iliev I. G., Nedelchev M. V., Antenna Diversity Multhi User Detection Algotithm for Synchronous CDMA System, ICEST 2009, V. Tarnovo, Proceedings of Papers, pp.67-71

Couplings of Microstrip Triangular Open-Loop Resonators

Marin V. Nedelchev[1], Ilia G. Iliev[2]

Abstract: This paper presents a study of the coupling topologies of microstrip triangular open-loop resonators and their application in filter design. Due to their shape, triangular resonators have many coupling topologies in order to achieve couplings of different nature-electric, magnetic, mixed. Each coupling topology is analyzed in full wave electromagnetic (EM) simulator in order to estimate the resonance peaks in the frequency response. Based on the simulation results, coupling coefficient graphs are presented. These graphs can be used in microstrip filter design.

Keywords: coupling coefficient, triangular open loop resonator, cross coupled filters.

I. INTRODUCTION

In the modern communication systems, high selectivity and low passband loss are the main requirements for the microstrip filters. Low passband loss increases the system sensitivity and the high selectivity decrease the guard interval between two channels in a communication system. Better spectrum efficiency is achieved. High filter selectivity requires high filter order and more resonators. Because of the low unloaded Q factor of the microstrip resonators, the passband loss increases. Both requirements become contradictory for cascaded microstrip filters. Filters satisfying the increased requirements are the cross-coupled filters. They have non-adjacent resonator coupling.

The synthesis of cross coupled filters is based on the early works from Atia and Williams [1], Cameron and Rhodes [2,3] considering waveguide cavity filter design. It is based on the deriving the coupling matrix from the transfer function and its reduction to the corresponding topology form. This technique is found to be useful in the design of microstrip cross-coupled filters. Generally, they are compact in size and weight, easy for manufacturing and adjustment. Therefore, there is a growing interest in the microstrip filter design.

The practical application of the design approach adopted from the waveguide filters, require knowledge about the mutual couplings between the microstrip resonators.

Triangular open loop resonators are proposed for filter design in [4]. Their geometrical form allows coupling in different coupling schemes, which exhibits various coupling types. The coupling topologies shown on Fig.1 may be used in the cross coupled filter design as well as classic filters. It is possible to form the resonator as an isosceles right-angled, isosceles or equilateral triangle, according to the desired topology.

[1]Marin Veselinov Nedelchev – Assistant, PhD in Dept. of Radiotechnic in Faculty of Communications and Communication Technologies in TU –Sofia E-mail mnedelchev@tu-sofia.bg

[2]Ilia Georgiev Iliev – Assoc. Professor, PhD in Dept. of Radiotechnic in Faculty of Communications and Communication Technologies in TU –Sofia E-mail: igiliev@tu-sofia.bg

The optimum form depends on the filter topology, circuit size and the necessary value of the coupling coefficient. This paper describes the four basic coupling topologies, shown on Fig.1. The form of the resonator used in simulations is isosceles right-angled triangle.

Full wave EM simulator is used for obtaining the resonance frequencies when introducing electrical and magnetic walls. The coupling mechanism is described in order to achieve better comprehension of the nature and the applicability of the coupling structures. Numerical results for the values of the coupling coefficients are presented.

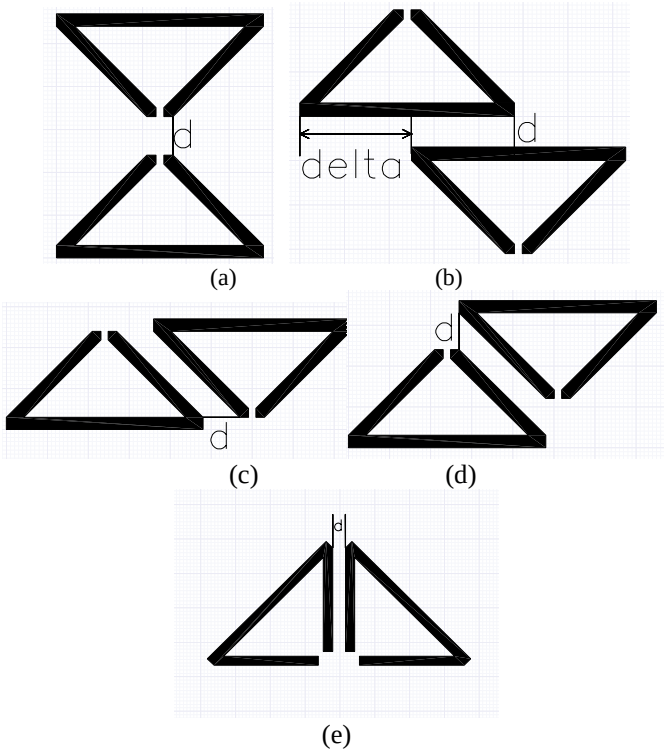


Fig.1 Topologies of triangular open loop resonators (a) electrical coupling, (b) magnetic coupling, (c), (d) and (e) mixed coupling

The obtained results are graphically shown and they are applicable for practical design.

II. TRIANGULAR OPEN LOOP RESONATOR COUPLING STRUCTURES

The triangular open loop resonator is half wavelength long for the central frequency of the filter. When constructing a filter, the exact form of the triangle should be chosen: right-angled, isosceles, equilateral or scalene. It is convenient for symmetrical filter topologies to use isosceles or equilateral

triangular resonators. More common case is to use right-angled isosceles resonators.

The coupling structures are formed by different orientation in the plane of identical resonators. They should be closely situated and separated by distance ***d.*** Both resonators in the coupling structure may have or not an offset in the direction of the coupled lines- denotes as ***delta*** (Fig.1b). This offset is made for two major reasons- to achieve loose coupling or to make possible connection of another resonator.

The coupling mechanism is based on the fringe fields of closely situated resonators. The nature of coupling depends on the resonator configuration. It is clear that half wavelength triangular resonator is symmetrical along the center of the hypotenuse. Consequently for the first resonance (odd mode), this point is an ideal short circuit (electric wall). The magnetic field will be predominant over the electric around this point. On the opposite side, the electric field will be dominant around the open ends of the resonators.

The coupling coefficient for synchronously tuned resonators can be calculated easily by the resonance frequencies of even and odd mode [5]:

$$k = \frac{f_e^2 - f_o^2}{f_e^2 + f_o^2} \tag{1}$$

The necessary condition for observing these resonance peaks is to set the resonator structure in overcoupled mode. In this case the coupling coefficient is larger than the critical coupling value of $1/Q$, where Q is the quality factor of the resonators [6]. A full wave EM simulator based on the Method of the Moments (MoM) is used to identify the resonance frequencies. Most of the coupling structures are simulated using the symmetry in their topology for electrical and magnetic wall introducing in-between. When symmetry does not present, the whole structure is simulated. This does not constrain in any way the obtained results. The simulation are carried out for standard FR-4 substrate with $\varepsilon_r = 4.4$, $h = 1.5mm$, $tg\delta = 0.02$.

A. Electrical Coupling

Following the above mentioned considerations, an electrical coupling presents in the structure shown on Fig.1a. Both resonators are close each other with their open ends. Because of the triangular form of the resonators, the coupling is realized by the apexes of resonators. Despite of the strong electrical field at this point, the small proximity areas leads to very small values of the coupling coefficient.

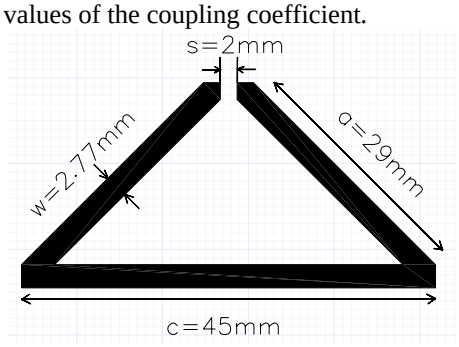


Fig.2 Triangular open loop resonator used in the simulations tuned on 811MHz

More over the sign of the coupling coefficient is negative, because the resonance frequency of odd mode (short circuit) is greater than the resonance frequency of even mode (open end) [5].

Figure 2 shows the topology of the microstrip triangular open loop resonator with the following dimensions: strip width $w = 2.77mm$, corresponding to 50Ω line on FR-4, hypotenuse length $c = 45mm$, cathetus length $a = 29mm$, gap between the cathetuses $s = 2mm$.

The first resonance frequency of the resonator is 811MHz. A number of full wave EM simulations are carried out in order to obtain the resonance frequencies in the frequency response. The coupling coefficient is calculated according to Eq.(1). The numerical results for the coupling coefficient are presented on Fig.3. It is clearly seen that the value of the coupling coefficient exponentially decays with the increase of the gap between the resonators. The results show that the coupling is very weak: the value of the coupling coefficient is about ten times less than in other coupling structures.

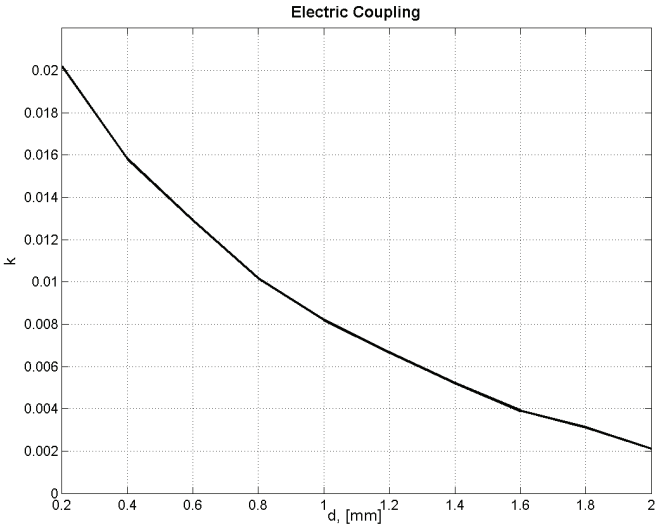


Fig.3 Coupling coefficient for electrically coupled resonators

The sign of the coupling coefficient is negative. This fact is very important for the cross coupled filters. The coupling between the non-adjacent resonators should be out-of-phase the other couplings. In the case of triangular coupled resonators, the coupling is not strong enough. This leads to transmission zeros in the frequency response away from the passband.

B. Magnetic Coupling

Magnetic nature of coupling presents in the coupling structure shown on Fig.1b, when the offset distance ***delta*** is zero. Both resonators are coupled through their hypotenuses, where short circuit is present for the resonance frequency. The magnetic field at this point is predominant over the electric. The electric field can be neglected in this case. Using the same resonator, shown on Fig.2, full wave EM simulations are carried out in order to obtain the resonance peaks in the frequency response of the coupling structure. The results are shown on Fig.4.

When $delta = 0$, there is no offset between the resonators. The coupling nature is pure magnetic. Since the hypotenuse is the longest side of the triangular resonator, combined with the strong magnetic field, the value of coupling coefficient is high. This topology is suitable for wideband filters, while maintaining reasonable distance d between the resonators. The inconvenience of using this structures rises from the fact that the cathetuses are the only available sides for couplings. This constrains the number of filter topologies involving the coupling structure.

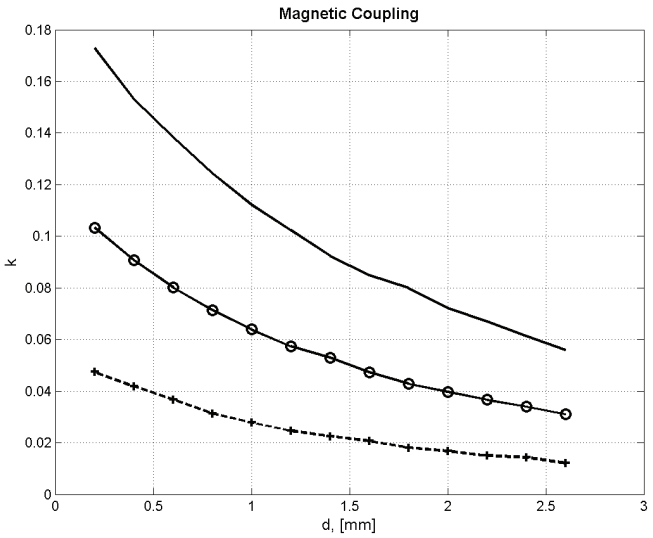


Fig.4. Coupling coefficient for magnetic coupling. Solid line- $delta = 0$ (no offset), circles- $delta = 0.5c = 22.5mm$, dashed- $delta = 0.25c = 11.25mm$.

More flexible form of predominant magnetic coupling is when an offset is presented in the structure (Fig.1b). The coupling is not purely magnetic in nature because of the offset. When $delta$ is one half of the hypotenuses, the one more side (the other side of the hypotenuses) is present for additional coupling. In this way the coupling coefficient lowers its values, but remains the monotonically decreasing behavior (Fig.4 circles).

When the offset $delta$ is a quarter of the hypotenuses, the coupling becomes weaker (Fig.4 dashed). The character of the coupling is not longer magnetic, but mixed. This topology combines the possibility of coupling to another resonators and asymmetric filter topologies.

C. Mixed Coupling

Mixed coupling presents, when both triangular resonators are close with their cathetuses (Fig.1c,d). The coupling resonators are opposite situated in the plane. The currents in the coupled lines are equal in amplitude and in-phase. It cannot be estimated which component of the field- electric or magnetic is dominant in this coupling structure. However, the electrical filed decays more rapidly with the distance from the open ends, than the magnetic field. The value of the coupling coefficient is with positive sign.

Since the coupling lines are tilted by 45^o, the distance between them is hard to be controlled. There is no closed form formulas for synthesis of microstrip coupled lines tilted by arbitrary angle. The only way for characterization them is to use a full wave EM simulator.

The approach chosen in this paper for obtaining the coupling coefficient is to fix one of the resonators and to move the other one either in horizontal (Fig.1c), or in vertical (Fig.1d) direction. This approach has an important advantage when adjusting the filter. Then it is easy to measure exactly the gap d between the resonators and to verify the coupling coefficient according to the graphical results presented here.

When horizontal spacing is introduced, the graphical results are presented on Fig.5. It is clearly seen that the value of the coupling coefficient decreases with increasing the spacing between the resonators.

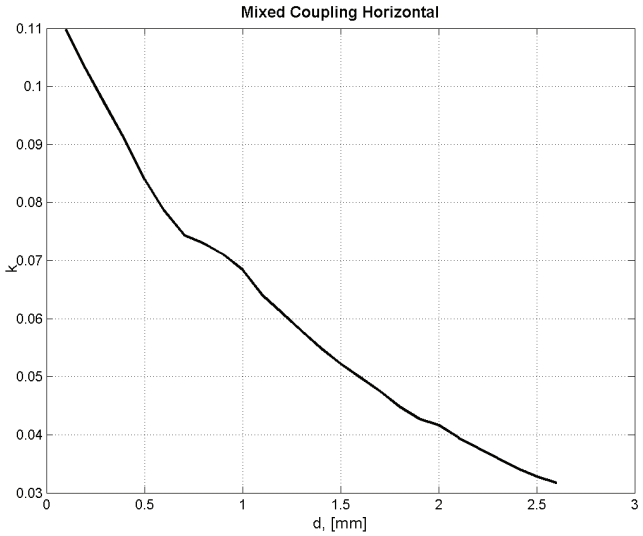


Fig.5 Mixed coupling with spacing in horizontal direction

The graphical results for vertical spacing between the resonators (Fig.1d) are presented on Fig.6.

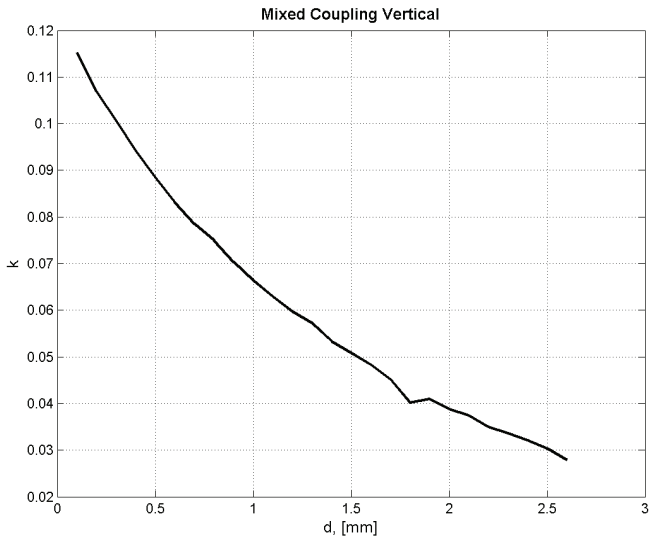


Fig.6 Mixed coupling with spacing in vertical direction

The coupling coefficient has the same behavior as the coupling coefficient for horizontal spacing.

The mixed coupling presents in the coupling structure shown on Fig.1e. As it is seen both resonators are coupled with their cathetuses and they are parallel in the plane. The

currents in the coupled lines are equal in amplitudes, but out-of-phase. This fact supposes low value of the coupling coefficient. The coupling structure is simulated in full wave EM simulator according to the above described method and the resonance peaks' frequencies are obtained. The graphical results are presented on Fig.7.

As it is clearly seen, the magnetic and electric field acts in opposite directions. This leads to coupling coefficient values about two times less than in the topologies shown on Fig.1 (c and d). It is reported in [5,7], that the coupling coefficient for this type of mixed coupling has not monotonic behavior. This is the case when the electric and magnetic filed annihilate each other. Then the coupling coefficient could be zero [7].

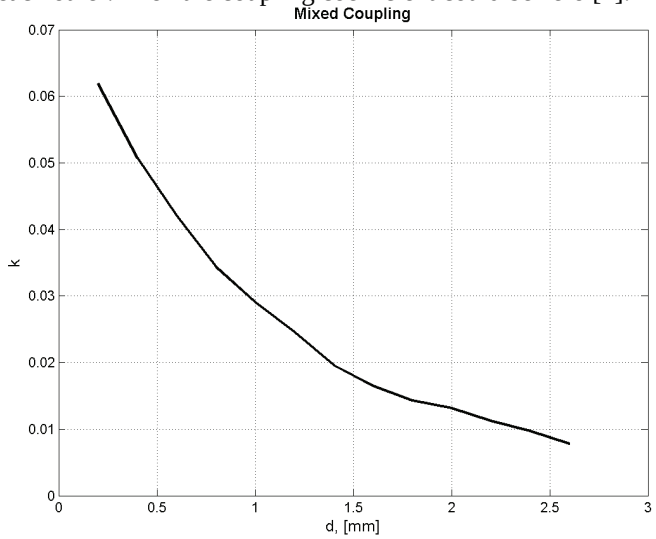


Fig.7 Mixed coupling for the structure shown on Fig.1e.

The coupling coefficient for the structure on Fig.1e has two times faster decreases than the other coupling coefficients. For the investigated region of spacing $d = (0.2 \div 2.6)mm$, the coupling coefficient decreases more than 6 times. The corresponding coupling coefficient for mixed coupling structure shown on Fig.1c and d decreases 3.5 times. The magnetic coupling with *delta=0, delta=0.5c* and *delta=0.25c* decreases around 3 times . The electric coupling the "dynamic range" of the coupling coefficient is about 10 times.

This observation may be taken into account, when consider the sensitivity of the coupling structures with the manufacturing tolerances and the impact of the adjustment of the couplings on the overall response of the filter. It is seen that the weak couplings lead to higher sensitivity to the spacing tolerances.

The presented graphs for different coupling structures of microstrip triangular open loop resonators can be used in the filter synthesis, when FR-4 substrate is utilized.

V. Conclusion

This paper presents a survey of the coupling coefficient for microstrip triangular open loop coupled resonators. The form of the resonator used is isosceles right-angled triangle. There are described the four basic coupling topologies with electric, magnetic and mixed coupling. The coupling mechanism is described in order to achieve better comprehension of the nature and the applicability of the coupling structures. In order to characterize the coupling coefficient, a full wave EM simulator is used for obtaining the resonance frequencies when introducing electrical and magnetic walls. Numerical results for the values of the coupling coefficients are presented. The numerical results can be used in cross coupled filter synthesis.

References

[1] A.E. Atia and A.E. Williams. "Narrow-Bandpass Waveguide Filters." 1972 Trans. on Microwave Theory and Techniques 20.4 (Apr. 1972 [T-MTT]): 258-265

[2] Cameron, Rhodes, J.D., Asymmetric Realization for Dual-Mode Bandpass Filters, IEEE Trans on MTT 1981 Jan., pp.51-58

[3] Cameron, R., Advanced Coupling Matrix Synthesis Techniques for Microwave Filters, IEEE Trans on MTT-50, Jan.2003, pp.1-10

[4] A. GÖRÜR, and C. KARPUZ, "Cross-coupled bandpass filter using microstrip triangular open-loop resonators", Proceedings of 31th The European Microwave Conference, CD-ROM, Sep 24-28, 2001, Excel, London, ENGLAND

[5] Hong J.-S. , M.J. Lancaster. "Couplings of Microstrip Square Open-Loop Resonators for Cross-Coupled Planar Microwave Filters." 1996 Transactions on Microwave Theory and Techniques 44.11 (Nov. 1996 [T-MTT]): 2099-2109.

[6] Matthaei, G., Young, L., Jones, E.M.T., Microwave Filters, Impedance-Matching Networks, and Coupling Structures, Artech House, Norwood, MA, 1980

[7] Nedelchev, M., Coupling Coefficient analysis in synchronously tuned square open-loop microstrip filters. Electronics and Electrotechnics Journal, 2005 (in Bulgarian)

SESSION SP I

Signal Processing I

Algorithm for Fast Complex Hadamard Transform

Rumen P. Mironov[1], Roumen K. Kountchev[2]

***Abstract* – An algorithm for fast complex Hadamard transform is presented. The complex Hadamard matrices are factorized with set of sparse matrices on the base of classical Cooley-Tukey algorithm and obtained signal flow graphs of order 4, 8 and 16 are shown. The developed algorithm is simulated on Matlab 6.5 environment and the computational complexity is calculated.**

***Keywords* – digital signal processing, Complex Hadamard Transform, fast orthogonal transforms.**

I. INTRODUCTION

Discrete orthogonal (unitary) transforms [1], [2] have found applications in many areas of N-dimensional signal processing, spectral analysis, pattern recognition, digital coding, computational mathematic and etc. Dimensionality reduction in computation is a major signal processing application. Stated simply, these transform coefficients that are small may be excluded from processing operations, such as filtering, without much loss in processing accuracy. The discrete integer Walsh Hadamard Transform (WHT) is a fairly simple orthogonal transform and is an example of a generalized class of Fourier transforms [3]. The idea of using complex, rather than integer transforms matrices for spectral processing, analysis and watermarking has been shown in [4], [5], [6] and [7]. From the Complex Hadamard Transform (CHT), several complex decisions diagrams are derived and analysis of more general CHT properties for 1D and 2D signals are investigated [8].

In this paper an algorithm for Fast Complex Hadamard Transform (FCHT) is developed, using factorization of basis CHT matrices by the sparse matrices and the signal flow graphs illustrating the computation of orders 4, 8 and 16 are shown. The obtained results are similar to factorization and flow graphs of real Fast Hadamard Transform (FHT), which leads to considerable decreasing of mathematical computations. The difference between FCHT and FHT is entirely into the last iteration, which include the all complex operations. Comparison with the presented in [9] fast conjugate symmetric sequence-ordered Complex Hadamard Transform is made and the results are given.

The developed FCHT algorithm is simulated on Matlab 6.5 environment and it requires $N.log_2 N$ additions or subtractions and $N/2$ complex operations in the last iteration.

[1]Rumen P. Mironov is with the Faculty of Telecommunications, Technical University of Sofia, Boul. Kl. Ohridsky 8, Sofia 1000, Bulgaria, E-mail: rmironov@tu-sofia.bg

[2]Roumen K. Kountchev is with the Faculty of Telecommunications, Technical University of Sofia, Boul. Kl. Ohridsky 8, Sofia 1000, Bulgaria. E-mail: rkountch@tu-sofia.bg

II. MATHEMATICAL DESCRIPTION

The coefficients of Complex Hadamard Transform matrix $[CH_N]$ with dimension N by N can be represented by the following equations [7], [8]:

$$\begin{vmatrix} c(u,v) = j^{uv} s(u,v) \\ c^*(u,v) = (-j)^{uv} s(u,v) \end{vmatrix}, \quad (1)$$

where: $N = 2^n$, $j = \sqrt{-1}$, u, v = 0,1,…2^n-1 and

$$s(u,v) = \begin{cases} 1 & \text{for } n = 2 \\ (-1)^{\sum_{r=3}^{n} \lfloor u/2^{r-1} \rfloor \lfloor v/2^{r-1} \rfloor} & \text{for } n = 3,4,5..... \end{cases}, \quad (2)$$

is the sign function. Here $\lfloor . \rfloor$ is an operator, which represents the integer part of the result, obtained after the division.

From the equations (1) and (2) the CHT basis matrix of order 2^n and complex conjugated matrix calculated for n=2 and u, v = 1,2,3,4 are presented as follows:

$$[CH_4] = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & j & -1 & -j \\ 1 & -1 & 1 & -1 \\ 1 & -j & -1 & j \end{bmatrix} \quad [CH_4]^* = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -j & -1 & j \\ 1 & -1 & 1 & -1 \\ 1 & j & -1 & -j \end{bmatrix}. \quad (3)$$

The basis complex Hadamard matrices of order 2^n (n>2) can be received as the Kroneker product of a number of identical "core" matrices of order 2^{n-1} in the following way:

$$[CH_{2^n}] = \begin{bmatrix} [CH_{2^{n-1}}] & [CH_{2^{n-1}}] \\ [CH_{2^{n-1}}] & -[CH_{2^{n-1}}] \end{bmatrix}. \quad (4)$$

Using the basic forward one-dimensional complex Hadamard transform for n=2 from the input signal vector $\overset{\nu}{A} = [a_1, a_2, a_3, a_4]$, the output spectral vector - $\overset{\nu}{B} = [b_1, b_2, b_3, b_4]$ is received by the equations [7]:

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & j & -1 & -j \\ 1 & -1 & 1 & -1 \\ 1 & -j & -1 & j \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix}, \quad \begin{vmatrix} b_1 = a_1 + a_2 + a_3 + a_4 \\ b_2 = a_1 + ja_2 - a_3 - ja_4 \\ b_3 = a_1 - a_2 + a_3 - a_4 \\ b_4 = a_1 - ja_2 - a_3 + ja_4 \end{vmatrix}. \quad (5)$$

The complex Hadamard matrix $[CH_4]$ can be decomposed of series of two sparse matrixes of order 4:

$$[CHI_4]=\begin{bmatrix}1&0&1&0\\0&1&0&1\\1&0&-1&0\\0&1&0&-1\end{bmatrix}=\begin{bmatrix}[I_2]&[I_2]\\[I_2]&-[I_2]\end{bmatrix}, \quad (6)$$

$$[CHJ_4]=\begin{bmatrix}1&1&0&0\\0&0&1&j\\1&-1&0&0\\0&0&1&-j\end{bmatrix}, \quad (7)$$

where $[I_{2^n}]$ is the identity matrix and decomposition can be written by the following:

$$[CH_4]=[CHJ_4][CHI_4]. \quad (8)$$

The first matrix $[CHI_4]$ is presented with a first sub-graph and the second matrix $[CHJ_4]$ is presented with a second one in the generalized CHT signal flow graph, shown on Fig.1:

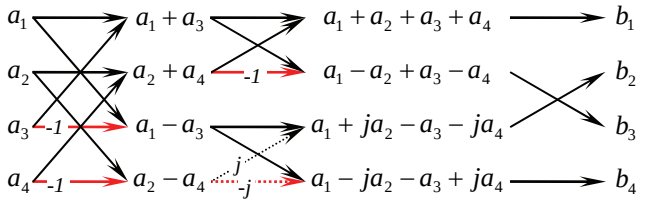


Fig. 1. Fast CHT signal flow graph of order 4.

The third sub-graph presents reordering of spectrum elements from equation (5). Multipliers are *+1* and *-1* as indicated by the black and red solid lines in real parts of sub-graphs and $+j$ and $-j$ as indicated by the black and red dashed lines in complex parts, respectively.

Using this approach the fast CHT algorithm and the corresponding signal flow graph of order 8 can be constructed in the following way:

- record the basic foreword complex Hadamard transform as linear system of 8 unknown values:

$$\begin{aligned}
b_1 &= a_1 + a_2 + a_3 + a_4 + a_5 + a_6 + a_7 + a_8\\
b_2 &= a_1 + ja_2 - a_3 - ja_4 + a_5 + ja_6 - a_7 - ja_8\\
b_3 &= a_1 - a_2 + a_3 - a_4 + a_5 - a_6 + a_7 - a_8\\
b_4 &= a_1 - ja_2 - a_3 + ja_4 + a_5 - ja_6 - a_7 + ja_8\\
b_5 &= a_1 + a_2 + a_3 + a_4 - a_5 - a_6 - a_7 - a_8\\
b_6 &= a_1 + ja_2 - a_3 - ja_4 - a_5 - ja_6 + a_7 + ja_8\\
b_7 &= a_1 - a_2 + a_3 - a_4 - a_5 + a_6 - a_7 + a_8\\
b_8 &= a_1 - ja_2 - a_3 + ja_4 - a_5 + ja_6 + a_7 - ja_8
\end{aligned} \quad (9)$$

- decomposition of complex Hadamard matrix $[CH_8]$ by the series of three sparse matrixes of order 8:

$$[CH_8]=[CHJ_8][CHI_8]'.[CHI_8]'', \quad (10)$$

$$[CHI_8]''=\begin{bmatrix}1&0&0&0&1&0&0&0\\0&1&0&0&0&1&0&0\\0&0&1&0&0&0&1&0\\0&0&0&1&0&0&0&1\\1&0&0&0&-1&0&0&0\\0&1&0&0&0&-1&0&0\\0&0&1&0&0&0&-1&0\\0&0&0&1&0&0&0&-1\end{bmatrix}=\begin{bmatrix}[I_4]&[I_4]\\[I_4]&-[I_4]\end{bmatrix} \quad (11)$$

$$[CHI_8]'=\begin{bmatrix}1&0&1&0&0&0&0&0\\0&1&0&1&0&0&0&0\\1&0&-1&0&0&0&0&0\\0&1&0&-1&0&0&0&0\\0&0&0&0&1&0&1&0\\0&0&0&0&0&1&0&1\\0&0&0&0&1&0&-1&0\\0&0&0&0&0&1&0&-1\end{bmatrix}=\left[\begin{array}{c:c}[CHI_4]&0\\\hdashline 0&[CHI_4]\end{array}\right] \quad (12)$$

$$[CHJ_8]=\begin{bmatrix}1&1&0&0&0&0&0&0\\0&0&1&j&0&0&0&0\\1&-1&0&0&0&0&0&0\\0&0&1&-j&0&0&0&0\\0&0&0&0&1&1&0&0\\0&0&0&0&0&0&1&j\\0&0&0&0&1&-1&0&0\\0&0&0&0&0&0&1&-j\end{bmatrix}=\left[\begin{array}{c:c}[CHJ_4]&0\\\hdashline 0&[CHJ_4]\end{array}\right] \quad (13)$$

where: $[CHI_8]''$, $[CHI_8]'$ and $[CHJ_8]$ are real and complex factorization sparse matrices;

- construction of signal flow graph for complex Hadamard transform of order 8, which is shown on Fig.2;

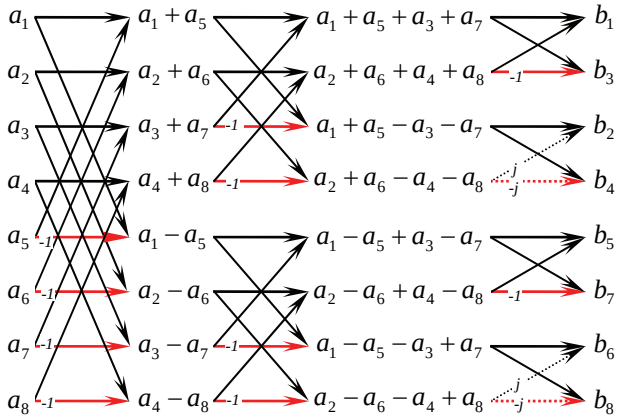


Fig. 2. Fast CHT signal flow graph of order 8.

Decompositions, shown in (8) and (10) can be generalized using Cooley-Tukey algorithm ([2], [3]) for the real [*CHI*] matrices factorization and multiplying with the complex

matrix [*CHJ*]. The complex Hadamard matrix $[CH_N]$ of order $N=2^n$ can be presented by the equation:

$$[CH_N]=[CHJ_N]\prod_{r=1}^{n-1}[G_r(N)], \qquad (14)$$

where: $n=\log_2 N$, $[G_r(N)]$ are the sparse matrices with two non-zero elements in each row, which have the following block-diagonal structure:

$$[G_r(N)]=\begin{bmatrix}[A(r)] & 0 & ..... & 0\\ 0 & [A(r)] & ..... & 0\\ ..... & ..... & ..... & .....\\ 0 & 0 & ..... & [A(r)]\end{bmatrix}. \qquad (15)$$

The sub-matrices $[A(r)]$ are defined as Kroneker product of the matrices:

$$[A(r)]=[H_2]\otimes[I_{2^r}], \qquad (16)$$

where: $[I_{2^r}]$ is identity matrix of size $2^r x 2^r$, and $[H_2]$ is real basic Hadamard matrix of order 2.

Using definitions in (7) and equations (14)-(16) the fast CHT for *N*-components vector $\overset{\nu}{A}=[a_1,a_2,.....,a_{N-1},a_N]$ can be presented in the following way:

$$\overset{\nu}{B}=[CHJ_N][G_1(N)][G_2(N)].....[G_{n-1}(N)]\overset{\nu}{A}, \qquad (17)$$

or as a sequence of elementary transformations:

$$\begin{aligned}&\overset{\nu}{C}_1=[G_{n-1}(N)]\overset{\nu}{A}\\ &\overset{\rho}{C}_2=[G_{n-2}(N)]\overset{\rho}{C}_1\\ &.....\\ &\overset{\rho}{B}=[CHJ_N]\overset{\rho}{C}_{n-1}\end{aligned}, \qquad (18)$$

where: $\overset{\nu}{C}_1,\overset{\nu}{C}_2,....\overset{\nu}{C}_{n-1}$ is the sequence of "intermediated" vector-iterations, which are received by the transformations with sparse matrices.

As a sample, from equation (14) for $[CH_8]$ factorization of order *N*=8 are obtained:

$$[A(1)]=[H_2]\otimes[I_2]; [A(2)]=[H_2]\otimes[I_4]$$

$$[CH_8]=[CHJ_8]\prod_{r=1}^{2}[G_r(8)]=[CHJ_8][G_1(8)][G_2(8)]$$

$$[G_1(8)]=\left[\begin{array}{cc:cc} [I_2] & [I_2] & & 0\\ [I_2] & -[I_2] & & \\ \hdashline & & [I_2] & [I_2]\\ 0 & & [I_2] & -[I_2]\end{array}\right]=\left[\begin{array}{c:c}[CHI_4] & 0\\ \hdashline 0 & [CHI_4]\end{array}\right]. \qquad (19)$$

$$[G_2(8)]=\left[\begin{array}{c:c}[I_4] & [I_4]\\ \hdashline [I_4] & -[I_4]\end{array}\right]$$

The received from (19) equations are completely identical with the equations (10)-(13). The developed fast complex Hadamard transform (FCHT) algorithm can be illustrated by the signal flow graph, shown on Fig.2.

The described algorithm can be used for the reverse complex Hadamard transformation. The flow graphs are identical and all output components must be divided on the *N*.

III. Experimental Results

The developed FCHT algorithm is simulated on Matlab 6.5 environment. The basis complex Hadamard matrices are calculated using equation (4) and factorization sparse matrices using equation (14) for orders *N*=4, 8, 16, 32, 64, 128, 256 and 512 are constructed. The received results are identical, which indicates that proposed factorization is correct. The FCHT algorithm is similar with the real FHT algorithm, which leads to considerable decreasing of mathematical computations. The difference between FCHT and FHT is entirely into the last iteration, which includes all complex operations.

The developed FCHT algorithm requires $N.log_2 N$ additions or subtractions and *N/2* complex operations in the last iteration. The presented in [9] fast conjugate symmetric sequence-ordered Complex Hadamard Transform algorithm requires $N.log_2 N$ additions or subtractions, but the complex operations are 3 times more than FCHT.

As a sample the obtained signal flow graph of order 16 is given on the Fig.3.

Using the matrix descriptions for the 2D Complex Hadamard Transform in [8], the 2D FCHT algorithm can be realized by applying of 1D FCHT on the rows of the input image matrix and after then applying the 1D FCHT on the columns of the obtained matrix. The calculation complexity of 2D FCHT can be evaluated from the complexity of 1D FCHT and require: $2.N^2.log_2 N$ additions or subtractions and N^2 complex operations.

IV. Conclusion

A class of complex Hadamard matrix is presented. An algorithm for fast CHT matrices calculation is developed on the base of classical Cooley-Tukey algorithm. Additional factorization matrices for the complex calculations are constructed.

The main advantages of the developed FCHT algorithm are:

- minimizing of complex calculations;
- the input components of each iteration can be substituted with the resulting ones, which reduces the used memory only for calculated *N* components;
- the 2D FCHT algorithm is based on the described 1D FCHT algorithm and is calculated by the same way as a real 2D FHT.

The presented fast ordered Complex Hadamard Transform can be used in digital signal processing for spectral analysis, pattern recognition, digital watermarking, transformation,

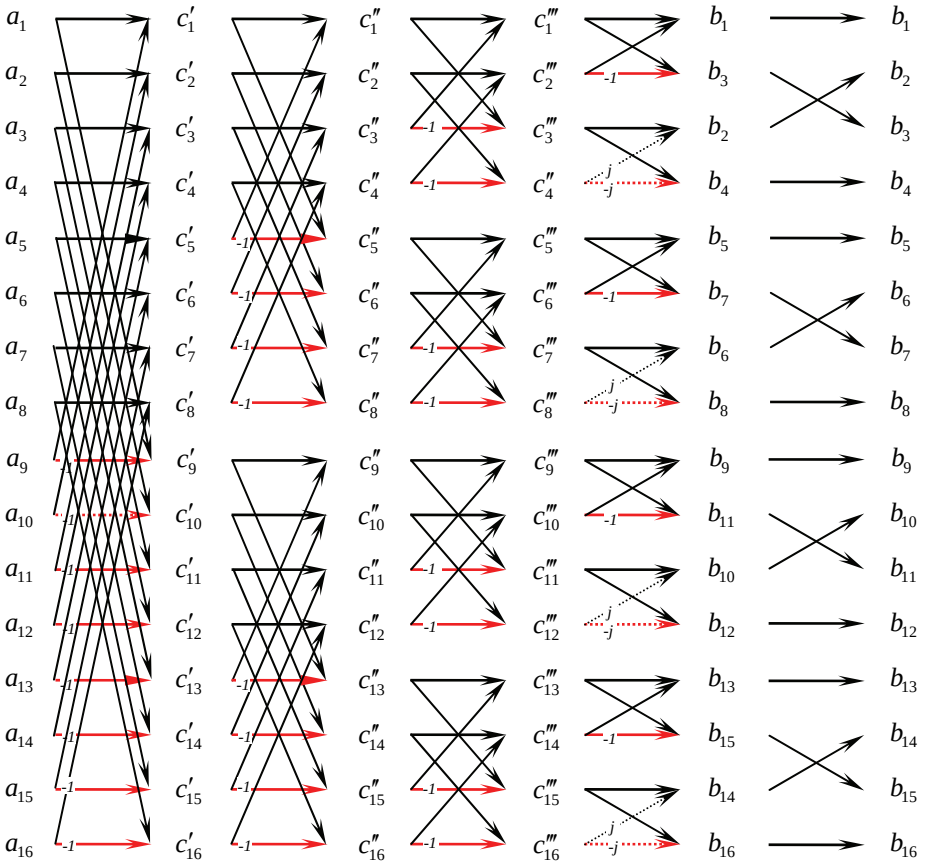


Fig. 3. Fast CHT signal flow graph of order 16.

coding and transmission of one-dimensional and two-dimensional signals.

V. Acknowledgement

The authors thank the National Fund for Scientific Research of the Bulgarian Ministry of Education and Science for the financial support by the contract VU-I-305/2007.

VI. References

[1] N. Ahmed, K. R. Rao. *Orthogonal Transforms for Digital Signal Processing*, Springer-Verlag Berlin, Heidelberg, 1975.

[2] A. D. Poularikas. *The Transforms and Applications Handbook*, Second Ed., CRC Press, 2000.

[3] W. K. Pratt. *Digital Image Processing*, John Wiley&Sons, N.Y., 2001.

[4] B. Falkowski, S. Rahardja. "Complex Spectral Decision Diagrams", *Proc. of the 26th Int. Symposium on Multiple Valued Logic*, Vol. ISMVL'96, 1996.

[5] S. Rahardja, B. Falkowski. "Complex Composite Spectra of Unified Complex Hadamard Transform for Logic Functions", *IEEE Trans. on Circuits and Systems-II: Analog and Digital Signal Processing*, Vol. 47, No. 11, November 2000.

[6] B. Falkowski, S. Rahardja. "Complex Hadamard Transforms: Properties, Relations and Architecture", *IEICE Trans. Fundamentals*. Vol. E87-A, No.8, August 2004.

[7] R. Kountchev, R. Mironov. "Audio Watermarking in the Phase-Frequency Domain", *XL Intern. Scientific Conference on Information, Communication and Energy Systems and Technologies*, ICEST'2005, Nis, Serbia and Montenegro, 2005.

[8] R. Mironov, R. Kountchev. "Analysis of Complex Hadamard Transform Properties", *XLI International Scientific Conference on Information, Communication and Energy Systems and Technologies*, ICEST 2006, 26 June – 1st July, Sofia, Bulgaria, 2006, pp.173-176.

[9] Aye Aung, Boon Poh Ng, S. Rahardja. "Conjugate Symmetric Sequency-Ordered Complex Hadamard Transform". *IEEE Transactions on Signal Processing*, July 2009, Vol. 57, No. 7, pp. 2582-2593.

Wavelet Thresholding by Using Multiscale Correlation

Cvetko Mitrovski and Mitko Kostov[1]

***Abstract* – In this paper we present a technique for thresholding wavelet coefficients based on multiscale product analysis. We apply experiments both on noise-free and noisy signals; noise is signal dependent. Experimental results show that the noise is effectively removed.**

***Keywords* – Denoising, threshold, wavelet domain filtering, multiscale products, signal-dependent noise.**

I. INTRODUCTION

Lately, there are many developed methods for image noise filtration in a transformation domain. In the last decade the stress on researches in this field is put on the signal processing in the wavelet domain. Wavelet transforms are multiresolution representations of signals and images. They decompose signals and images into multiscale details.

The reason of using the wavelet transform for denoising purposes is that the basis functions used in wavelet transforms are locally supported; they are nonzero only over part of the domain represented. Hence, adequately chosen wavelet basis groups the coefficients in two groups – one with a few coefficients with high SNR, and other with a lot of coefficients with low SNR. In addition, the signal power at large scales corresponds to that at low frequencies in the Fourier transform; the power at small scales corresponds to that at high frequencies in the Fourier transform.

In case of white Gaussian noise, the noise level is same through whole signal and for all the wavelet coefficients, independently on the signal. Hence, a well-estimated global threshold shrinks all the coefficients for an equal portion and removes the noise. But, in some signals, like nuclear medicine (NM) images, the noise level is proportional to the local signal intensity. Obviously, denoising them by using a global threshold is not the best solution.

In this paper we propose removing signal noise by exploiting wavelet multilevel correlation: the small scale data is passed at positions where the correlation is large and suppressed if the correlation is small. The paper is organized as follows. Section II briefly outlines wavelet theory. Section III discusses how to analyze the signal through the scale. Section IV verifies the validity of our approach on deterministic signals contaminated with signal dependent noise. At the end, Section V concludes the paper.

[1]Mitko Kostov and Cvetko Mitrovski are with the Faculty of Technical Sciences, I.L.Ribar bb, 7000 Bitola, Macedonia, E-mails: mitko.kostov@uklo.edu.mk, cvetko.mitrovski@uklo.edu.mk

II. DISCRETE WAVELET SHRINKAGE METHOD

The Discrete Wavelet Transform (DWT) decomposes a signal into a set of orthogonal components describing the signal variation across the scale [1]. The orthogonal components are generated by dilations and translations of a prototype function ψ called mother wavelet.

In analogy with other function expansions, a function f may be written for each discrete coordinate t as a sum of a wavelet expansion up to certain scale J plus a residual term, that is:

$$f(t) = \sum_{j=1}^{J} \sum_{k=1}^{2^{-j}M} d_{jk} \psi_{jk}(t) + \sum_{k=1}^{2^{-J}M} c_{Jk} \phi_{Jk}(t) \qquad (1)$$

where ψ_{jk} is component obtained by dilation and translation of the mother wavelet. The approximation coefficients c_{Jk} contain the signal identity while the detail coefficients d_{jk} likely contain noise and need to be processed in order to remove the noise.

The most popular form of conventional wavelet-based signal filtering [1], can be expressed by:

$$\begin{aligned} &\{\mathbf{A}^{(k)}, \mathbf{D}^{(1)}, \mathbf{D}^{(2)}, \Lambda\ , \mathbf{D}^{(k)}\} = \mathrm{DWT}(\mathbf{s}+\mathbf{n}), \\ &\mathbf{s}^{*} = \mathrm{IDWT}\left(f\left(\mathbf{A}^{(k)}, \mathbf{h}^{(1)}.\times \mathbf{D}^{(1)}, \mathbf{h}^{(2)}.\times \mathbf{D}^{(2)}, \Lambda\ , \mathbf{h}^{(k)}.\times \mathbf{D}^{(k)}\right)\right) \end{aligned} \qquad (2)$$

where $\mathbf{s}$ is noise-free signal, $\mathbf{n}$ is noise, $\mathbf{s}^{*}$ is filtered signal, $\mathbf{A}^{(k)}$ and $\mathbf{D}^{(k)}$ are approximation and detail coefficients at level k, respectively, f is a function of the modified detail and approximation coefficients, .× is element-by-element multiplying and

$$\mathbf{h}^{(k)} = \left[h_1^{(k)}, h_2^{(k)}, \Lambda\ , h_j^{(k)}\right]^T \qquad (3)$$

are weighting coefficients of the corresponding detail coefficients at level k.

In case of conventional hard threshold filtering the weighting coefficients are

$$h_j^{(k)}(hard) = \begin{cases} 1, & \text{if } \left|D_j^{(k)}\right| > \tau^{(k)} \\ 0, & \text{otherwise} \end{cases}, \qquad (4)$$

while for the soft threshold filtering they are

$$h_j^{(k)}(soft) = \begin{cases} 1 - \dfrac{\tau^{(k)}\,\mathrm{sgn}\left(D_j^{(k)}\right)}{D_j^{(k)}}, & \text{if } \left|D_j^{(k)}\right| > \tau^{(k)} \\ 0, & \text{otherwise} \end{cases}, \qquad (5)$$

where $\tau^{(k)}$ is user specified threshold for the k-th level details.

III. MULTISCALE CORRELATION

A. Algorithm

A noisy signal **x** can be expressed as a sum of noise-free signal **s** and noise **n**:

$$\mathbf{x} = \mathbf{s} + \mathbf{n}. \tag{6}$$

An orthogonal wavelet transformation of the noisy input yields an equivalent model in each wavelet sub-band:

$$\boldsymbol{\omega} = \boldsymbol{\beta} + \boldsymbol{\nu}. \tag{7}$$

where $\boldsymbol{\beta}$ are noise-free signal wavelet coefficients and $\boldsymbol{\nu}$ are noise wavelet coefficients.

For a given noise-free signal coefficient β_p (at position p) in a reference to a certain threshold τ_p, two hypotheses can be defined [2]:

H_0: $|\omega_p| \le \tau_p$, the signal of interest is absent (in the given coefficient);

H_1: $|\omega_p| > \tau_p$, the signal of interest is present (in the given coefficient).

We want to estimate the level of probability that the wavelet coefficient ω_p represents signal of interest. For this aim we want to exploit the correlation between the coefficients at the neighboring levels. This can be done by multiplying small scale coefficients and the coefficients derived from the larger scale, after the larger scale coefficients are shifted first. Hence, we define a shrinkage rule for estimating the coefficient $\hat{\beta}_p$:

$$\hat{\beta}_p = \omega_p \cdot P(H_1 \mid \omega_p), \tag{8}$$

where

$$P = \begin{cases} 1 & \text{if } |\omega_p^{(k)}| \cdot |\lambda_p^{(k)}| \cdot \sqrt{\dfrac{\sum_j (\omega_j^{(k)})^2}{\sum_j (\omega_j^{(k)} \lambda_j^{(k)})^2}} > |\omega_p^{(k)}| \\ |\omega_p^{(k)}| \cdot |\lambda_p^{(k)}| \cdot \sqrt{\dfrac{\sum_j (\omega_j^{(k)})^2}{\sum_j (\omega_j^{(k)} \lambda_j^{(k)})^2}} & \text{otherwise} \end{cases} \tag{9}$$

$$\lambda_p^{(k)} = \max\{\Lambda\ \omega_{p-1}^{(k+1)}, \omega_p^{(k+1)}, \omega_{p+1}^{(k+1)} \Lambda\} \tag{10}$$

$\text{length}\{\boldsymbol{\lambda}\} = 4k$, k – decomposition level.

The coefficients $\boldsymbol{\lambda}^{(k)}$ in (10) are derived from the coarser level coefficients $\boldsymbol{\omega}^{(k+1)}$ and in this way we ensure that (shifted) maximums from the coarser level correspond to adequate spikes from the finer level. In addition, since the product coefficients $\boldsymbol{\omega}^{(k)}\boldsymbol{\lambda}^{(k)}$ in (9) have much bigger energy than the finer level coefficients $\boldsymbol{\omega}^{(k)}$, the coefficients $\boldsymbol{\omega}^{(k)}\boldsymbol{\lambda}^{(k)}$ are normalized. The criterion at (9) keeps a coefficient if it contains signal of interest or reduces it if not.

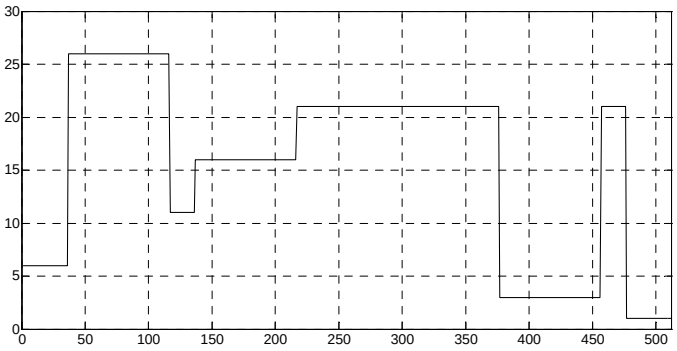


Fig. 1. Deterministic test signal.

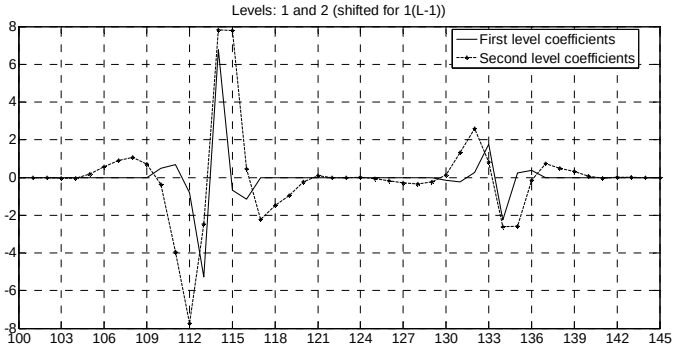


Fig. 2. Wavelet coefficients from 1st and 2nd level obtained by decomposing the signal with sym4 wavelet.

B. Analysis

If we consider signal decomposition in wavelet domain, starting from the finest level and moving through the scale, the signal energy increases and its shape slightly changes; the signal portions at the coarser scales contain more the signal identity than the portions at the finer scales. We want to use direct spatial correlations of the wavelet transform at different scales to identify the edges: the small scale data is passed at positions where the correlation is large and suppressed if the correlation is small. But, a small scale spike that appears at certain position stretches (dilates) in the next scale (coarser level), i.e. moves slightly around the considered positions in the next larger scale. This is shown by a wavelet decomposition of the deterministic noise-free signal shown in Fig. 1. The signal decompositions at different levels are carried out by using different wavelets and the results are shown in Figs. 2–5. Fig. 2 shows the signal decomposition carried out by wavelet sym4. The spikes at positions 77 and 78 at the first level are stretched in the second level, i.e. they are shifted to the positions 76 and 79. Similarly, the spikes at 497 and 498 (517 and 518) are shifted to positions 496 and 499 (516 and 519). Same results are obtained when other wavelets are used.

In addition, when wavelet coefficients from neighboring (k and k+1) levels are compared, the large level (k+1) coefficients are shifted for $k\times(L-1)$ samples, where L is the filter length.

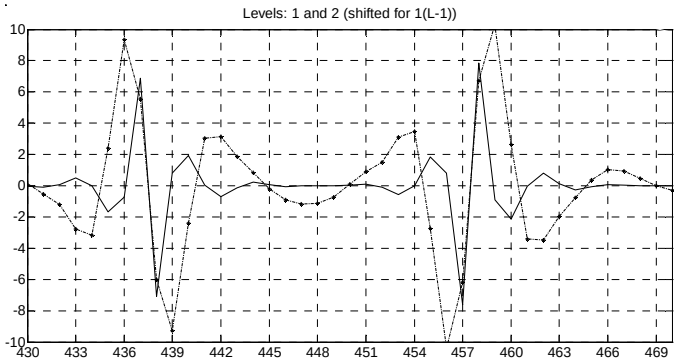
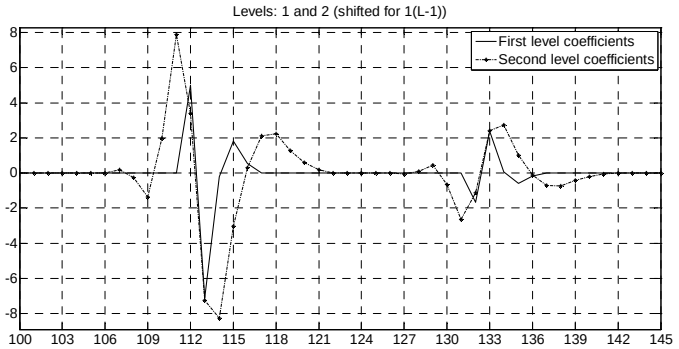


Fig. 3. Wavelet coefficients from 1st and 2nd level obtained by decomposing the signal with coif5 wavelet.



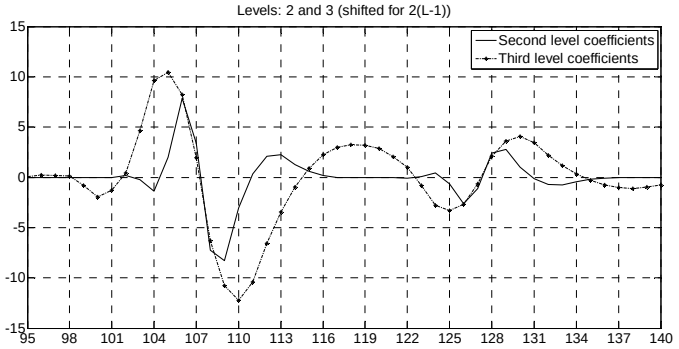


Fig. 4. Wavelet coefficients from 1st, 2nd and 3th level obtained by decomposing the signal with sym3 wavelet.

This correlation can be used to eliminate noise in the noisy signal. But, this filtering cannot be carried out by simple multiplying of the coefficients from the neighboring levels. We propose all the small scale coefficients to multiply by maximums in adequate large scale regions as it is shown in next section.

IV. EXPERIMENTS

In this Section, our experimental results are explained. The experiments are made with the noisy signal given in Fig. 6. The signal contains Poisson noise. It can be noticed in Fig 6 that there are signal information around the signal levels changes that should be preserved while filtering.

The results of filtering by using the proposed algorithm are shown in Figs. 7–9. It can be visually seen that the noise energy in the filtered signal is much lower than in the noisy

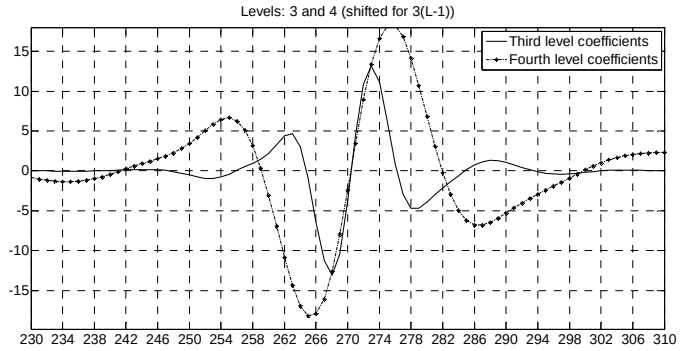


Fig. 5. Wavelet coefficients from 3th and 4th level obtained by decomposing the signal with coif5 wavelet.

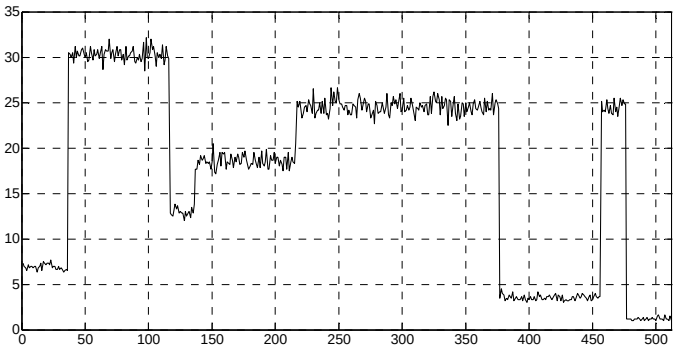


Fig. 6. Signal with noise.

MSE (Fig. 6)						
Wavelet	coif5		db4		sym3	
Level	1	2	1	2	1	2
Proposed	0.1446	0.1154	0.1419	0.1135	0.1382	0.1078
Multiscale	0.2703	0.3579	0.2041	0.2985	0.2411	0.4004
Universal	0.9077	1.8941	1.5997	2.0458	1.1990	4.1028

Table 1. Comparison of the proposed algorithm with other methods.

signal. The filtered coefficients preserve the signal information (big spikes) and the small spikes are eliminated.

In order to quantitatively compare the proposed method to some known wavelet based methods, we use the mean squared error of the remained noise in the filtrated signal $\mathbf{s}_1$ (normalized to the energy of the noise free signal **s**) as a measure:

$$MSE = \frac{1}{N}\sum_i \left(s(i) - s_1(i)\right)^2 \qquad (11)$$

where N is the signal length.

The results of filtering of signal are shown in Table 1. The signal is reconstructed from the first and second level filtrated detail coefficients and second level approximation coefficients. The table shows that when the proposed approach is applied, the noise energy is weaker compared to filtering by Donoho's universal threshold [3, 4] and multiscale product algorithm [5].

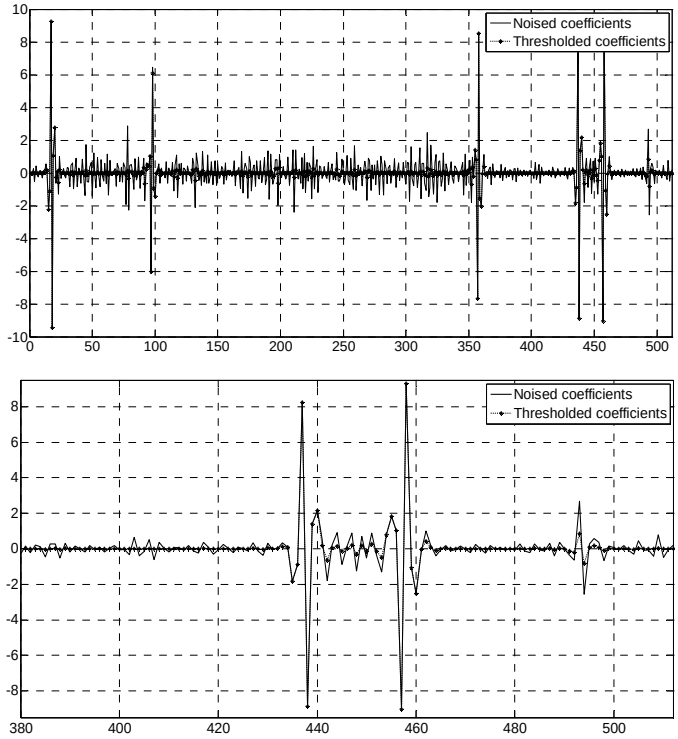


Fig. 7. Wavelet noised coefficients and thresholded coefficients at first level obtained with coif5.

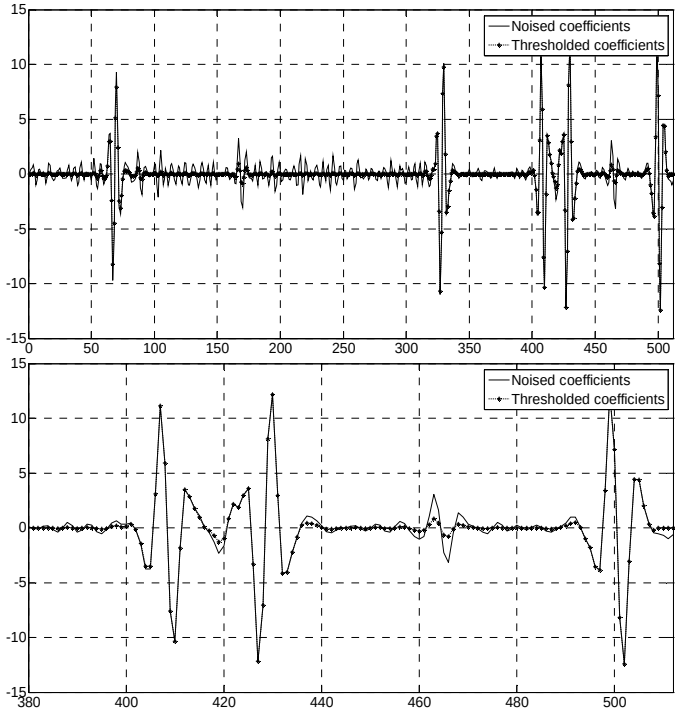


Fig. 8. Wavelet noised coefficients and thresholded coefficients at second level obtained with coif5.

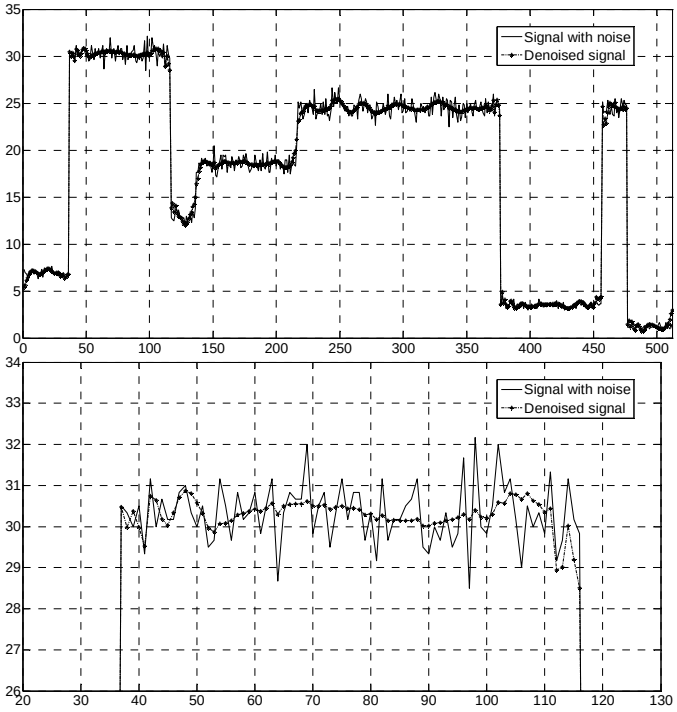


Fig. 9. Noise-free signal and denoised signal by using wavelet decomposition in two levels with coif5.

V. CONCLUSION

In this paper we propose removing signal noise by exploiting wavelet multilevel correlation: the small scale data is passed at positions where the correlation is large and suppressed if the correlation is small. Experimental results show that the noise is effectively removed.

REFERENCES

[1] G. Strang, T. Nguyen, Wavelets and Filter Banks, Wellesley-Cambridge Press, 1996.

[2] A. Pižurica, W. Philips, “Estimating the probability of the presence of a signal of interest in multiresolution single- and multiband image denoising”, *IEEE Trans. Image Process.*, vol. 15, no. 3, pp. 645–665, Mar. 2006.

[3] D.L. Donoho, "Wavelet Thresholding and W.V.D.: A 10-minute Tour", *Int. Conf. on Wavelets and Applications*, Toulouse, France, Jun. 1992.

[4] D.L. Donoho, I.M. Johnstone, "Ideal spatial adaptation via wavelet thresholding", *Biometrika*, vol. 81, pp. 425-455, 1994.

[5] Y. Xu, J.B. Weaver, D.M. Healy, Jr., J. Lu, “Wavelet Transform Domain Filters: A Spatially Selective Noise Filtration Technique”, *IEEE Trans. on Image Processing*, vol. 3, no. 6, pp. 747-758, Nov. 1994.

Construction of a Hybrid Quantizer with Huffman Coding for the Laplacian Source

Zoran H. Perić[1], Jelena R. Lukić[2], Dragan B. Denić[3], Milan R. Dinčić[4]

***Abstract* – This paper presents a hybrid model of a speech signal quantizer consisted of the uniform scalar quantizer and the optimal Lloyd-Max's quantizer. The proposed hybrid quantizer is developed for a small number of quantization levels *N*. The paper also gives an overview of the numerical results that indicate the advantages of the developed hybrid quantizer over the conventional uniform and nonuniform quantizer, where in all three cases was used Huffman coding method.**

***Keywords* – Hybrid quantizer, Huffman coding, Uniform and nonuniform quantizer**

I. INTRODUCTION

Speech signal compression involves quantization process, i.e. rounding off the previously sampled values of the signal from an infinite set to a values from the finite set, and the coding process, i.e. representing quantized signal values with bit sequences (code words).

There are a number of signal compression techniques, i.e. preparation of signals for transmission and storage. Compression techniques can be harshest divided into lossy (with losses) and lossless (with no losses) [5], [7]. The usage of lossy compression techniques leads to a loss of information about the signal which is compressed through a process of quantization, i.e. rounding off on the previously determined values (levels). Error added by the quantization process is measured by the distortion, while the quality of the compressed signal is measured by the value of signal-to-quantization noise ratio.

In order to raise the quality of voice signal it is important to choose an adequate quantizer, and a specific coding technique with it, in order to realize the lossless type of compression. For the realization of lossless types of compression the entropic codes are often used, and they are used for coding of a finite set of quantized signal values. The implementation of these codes brings additional complexity in the process of compression.

Although there are many realizations of quantizers, there is still a plenty of room for development of new models, especially when we need to raise the quality of the signal. In addition, with appropriate coding technique it should be easy to produce a digital signal which has a smaller impact on a channel range over which the it will be transferred (i.e. to reduce the average bit rate).

By following these requirements, in this paper we suggested a model of hybrid quantizer which output is coded with entropic Huffman code. Hybrid quantizer is consisted of the uniform and the optimal Lloyd-Max's scalar quantizer. It is known that the uniform quantizer with Huffman coding achieves lower average bit rate, while the nonuniform quantizer with Huffman coding achieves better signal-to-quantization noise ratio [4], [5], [7]. The proposed model of the hybrid quantizer incorporate both of the advantages mentioned above, which makes the hybrid quantizer better than the uniform and nonuniform quantizer. The speech signal at the entrance of the hybrid quantizer is modeled by the Laplacain probability density function. With this model of the hybrid quantizer and with Huffman coding we wanted to raise the quality of the signal, i.e. increase the value of signal-to-quantization noise ratio for the specific bitrate and simultaneously tended to reduce the average bit rate for a given value of signal-to-quantization noise ratio. On the other hand, the introduction of the coding techniques with variable length code words resulted in increased complexity of the proposed process of compression [4], [5], [7].

In this paper, beside the basic characteristics of the hybrid quantizer, will be presented the advantages achieved by applying Huffman coding algorithm. The advantages of the proposed hybrid quantizer over the standard uniform and nonuniformni scalar quantizer will be pointed out with a serie of numerical results. In order to compare the quantizers properly, the Huffman coding algorithm was applied in all three cases.

II. MODEL OF THE HYBRID QUANTIZER

In this paper is proposed a model of the hybrid scalar quantizer which is consisted of the simple uniform scalar quantizer and the optimal Lloyd-Max's scalar quantizer [1], [2], [3]. The model was developed for a small number of quantization levels N = 8, 16 and 32. Hybrid quantizer with N segments, was carried out in a way which means that the uniform quantizer part occupies the central part of the range of hybrid quantizer and include N-2L segments of equal width, which boundaries are defined by the decision levels x_i

[1]Zoran H. Perić is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: zoran.peric@elfak.ni.ac.rs

[2]Jelena R. Lukić is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: jelenalukicpk@gmail.com

[3]Dragan B. Denić is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: dragan.denic@elfak.ni.ac.rs

[4]Milan R. Dinčić is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: mdincha@hotmail.com

(i=L,...,N-L), and Lloyd- Max's quantizer part which contains 2L segments and occupies the outer part of the complete range. We only discussed the values of 1, 2 and 4 for the number of segments L. In order to achieve the lossless effect of the compression, we applied the Huffman coding algorithm with variable length code words. The advantage of the hybrid quantizer, achieved by applying the Huffman coding algorithm, is reflected on the level of quality of the compressed signal, with reduction of the average bit rate. The complexity of the proposed compression method mainly comes from the variable length coding algorithm. The proposed hybrid quantizer is designed for the small number of quantization levels, and for that reason with the uniform quantizer is used optimal Lloyd-Max's scalar quantizer. It is known that for the realization of the hybrid quantizer for a grater number of quantization levels, the compander is preferred instead of the Lloyd-Max's quantizer, as nonuniform quantizer [1], [2], [3].

How much the process of quantization affects the quality of compressed voice signal with unit variance modeled by the Laplacian probability density function,

$$p(x)=\frac{1}{\sqrt{2}}e^{-|x|\sqrt{2}} \tag{1}$$

can be estimated on the basis of the value of distortion D, i.e. the error introduced by the quantization procedure

$$D=\frac{2}{3}\frac{x_{N-L}^2}{(N-2L)^2}\int_0^{x_{N-L}}p(x)dx+$$

$$+2\sum_{i=N-L}^{N-1}\int_{x_i}^{x_{i+1}}(x-y_{i+1})^2p(x)dx, \tag{2}$$

for N=1,...,32 and L=1,2,4. The $x_{\text{N-L}}$ value, from the previous expression, is the decision threshold level that separates the range of the uniform quantizer from the range of the Lloyd-Max's quantizer within a range of the hybrid quantizer. Decision threshold level is determined in accordance with minimal distortion condition, i.e. equating the first derivate of distortion by the requested value with zero. When the value of the decision threshold level is known, the width of the segments in the uniform part of the range of the hybrid quantizer can easily be determined by dividing the double value of $x_{\text{N-L}}$ with N-2L. The parameters, $y_{\text{i+1}}$, are known as the representation levels in the part of the range with Lloyd-Max's quantization [4]. The quality of the signal after compression can be estimated on the basis of the value of signal-to-quantization noise ratio

$$SQNR[dB]=10\log_{10}\frac{1}{D} \tag{3}$$

Analysing the dependence of the signal-to-quantization noise ratio $SQNR$ on the average bit rate $\overline{R}$ (calculated in accordance with the Huffman coding algorithm), easily can be noticed the advantages of the hybrid quantizer over the uniform and nonuniform quantizer. All about this, as well as all about the other features of the proposed hybrid quantizer can be found in more words in the section with numerical results.

III. HUFFMAN CODING PROCEDURE

In this paper is analyzed the Huffman coding algorithm, which is implemented using already worked out MATLAB program code. Huffman coding algorithm is used for coding of the representation levels from the segments of the hybrid quantizer, or simply, the segments (because each segment contains a single representation level).

The Huffman algorithm will be described in the following. Let's analyse a source with the symbols s_i (i=0,1,...,q), which probabilities of appearance are P_i (i=0,1,...q) [6]. In the particular case of hybrid quantizer as symbols s_i (i=0,1,...q, q=N-1) are taken the segments from the range of the hybrid quantizer. Probabilities P_i (i=0,1,...q, q=N-1) are the probabilities of appearance of segments within a range of the hybrid quantizer and they are defined as follows:

$$P_i=\int_{x_i}^{x_{i+1}}p(x)dx, \quad i=0,...,N-1 \tag{4}$$

where parameters x_i are the decision levels ($x_0\rightarrow-\infty$, $x_n\rightarrow\infty$) [4], [5], [7]. The number of symbols is equal with number of segments (representation levels) N. The symbols should be regulated by the non-growing probabilities so that $P_0\geq P_1\geq...\geq P_q$. If the symbols have the same probability, its order in the editing does not matter. Now, it is needed to replace two the least probable symbols by a single, equivalent symbol, i.e. to form their union. Probability of appearance of this symbol is equal to the sum of the probabilities of the symbols that are replaced. In this way, the reduction of the source S with q symbols on the source S_1 which has q-1 symbols, is performed. Now, it is needed to re-edit the symbols of the reduced source by non-growing probabilities (because the new symbol does not need to be the least probable) and make a new (second) reduction to obtain a source S_2 with q-2 symbols. This process continues until q-1 reductions are made and the S_{q-1} source with only two symbols is got. At each process of editing symbols by non-growing probabilities the order within the group of symbols with equal probabilities is not important. Now we proceed from the reduced source with two symbols and begin to write code words in such a manner so the one symbol gets 0 as the code word, and the other gets bit 1 (arbitrarily is taken to place 0 in the code word of the symbol written above and to place bit 1 in the code word of the symbol written below). Since these symbols are the unions of the symbols from the previous reduction, the next process is to go step by step backward and break up the unions. At each breaking, bit 0 or bit 1 are added to the appropriate places in the code words, until the original source is reached [6]. For example, in the case of the hybrid quantizer with N=8 levels (the source with eight symbols) and L=1, the calculated probabilities of the segments in the entire range are given, from the first to the eighth: P_0=0.027, P_1=0.045, P_2=0.118, P_3=0.31, P_4=0.31, P_5=0.118, P_6=0.045

and P_7=0.027, respectively. Before these probabilities are put in the MATLAB program code, it is required to set them in non-growing order. After the algorithm is over the length of the code words, which are used to represent the previosly mentioned segments, are calculated and given here: l_0=4, l_1=4, l_2=3, l_3=2, l_4=2, l_5=3, l_6=4 and l_7=4, respectively. The average length of the various length code words, was calculated as follows:

$$\overline{R} = \sum_{i=0}^{N-1} P_i l_i \tag{5}$$

In this particular example, the average length of various length code words, i.e. average bit rate is 2.524 bits/sample.

IV. THE NUMERICAL RESULTS

In this section will be given all the numerical results of the analysis of the proposed hybrid quantizer with Huffman coding. In Table I are given the maximal values of signal-to-quantization noise ratio which are achieved by the proposed hybrid quantizer, while are in Table II exhibited average bit rate values at which the maximal values of signal-to-quantization noise ratio from Table I are achieved. It can be seen, from the displayed tables, that the better signal quality is achieved by the quantizer with a grater number of levels, but on the other hand this results in increased average bit rate. On the basis of the offered numerical results the optimum solution of the hybrid quantizer can be chosen in accordance with the requirements for better signal quality and/or lower average bit rate.

TABLE I
THE REWIEV OF MAXIMAL VALUES OF SIGNAL-TO-QUANTIZATION NOISE RATIO

SQNR (dB)	L=1	L=2	L=4
N=8	11.918	12.464	/
N=16	16.517	17.328	17.959
N=32	21.192	22.076	23.010

TABLE II
THE REWIEV OF AVERAGE BIT RATES AT WHICH THE MAXIMAL VALUES OF SIGNAL-TO-QUANTIZATION NOISE RATIO ARE ACHIEVED

$\overline{R}$ (bits/sample)	L=1	L=2	L=4
N=8	2.524	2.588	/
N=16	3.168	3.368	3.496
N=32	3.918	4.082	4.269

On Figs. 1, 2 and 3 are graphically illustrated functionalities of signal-to-quantization noise ratio from average bit rate for hybrid quantizer, when the Huffman coding algorithm is in use. By observing these graphics it can be seen that the higher quality of the signal can be achieved by using the quantizer with grater number of levels, which results in average bit rate increase (the same thing could be concluded on the basis of the numerical results from Table I and Table II). The aim of the analysis of these graphics is to see in which ranges of average bit rate the specific quantizer (referring to the values of N and L) gives the best quality of the signal at the output.

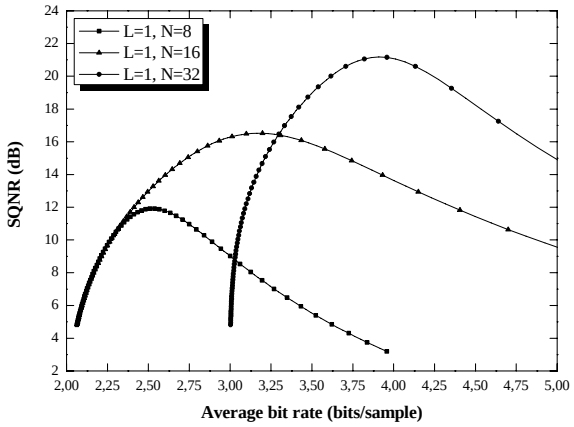


Fig. 1. The graphical interpretation of functionality of signal-to-quantization noise ratio from average bit rate for the hybrid quantizer, L=1 segment and N=8, 16 and 32 levels

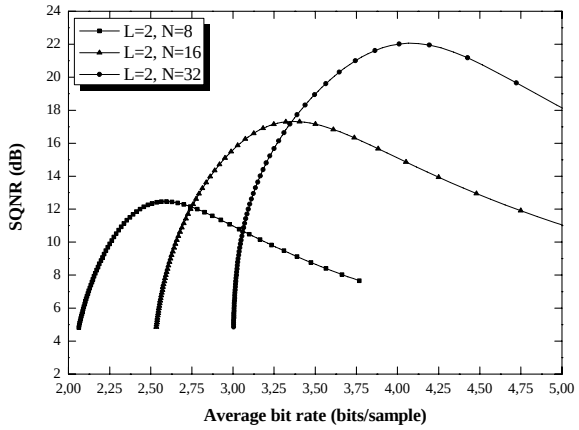


Fig. 2. The graphical interpretation of functionality of signal-to-quantization noise ratio from average bit rate for the hybrid quantizer, L=2 segments and N=8, 16 and 32 levels

Thus, from the Fig. 1 it can be determined that it is better to use the quantizer with N=8 levels for bit rates up to 2.315 bits/sample (for the mentioned range of the bit rate the quantizer with 8 levels gives the same quality as the quantizer with N=16 levels, but with less complexity), for the bit rates ranging from 2.315 to 3.298 bits/sample better quality is achieved with quantizer which has N=16 levels, while for the bit rates higher than 3.298 bits/sample it is better to use the quantizer with N=32 levels. By analyzing the graphics on Fig. 2 it can be determined that it is better to use the quantizer with

N=8 levels for bit rates up to 2.747 bits/sample, for the bit rates ranging from 2.747 to 3.36 bits/sample it is better to use the quantizer with *N*=16 levels, while for the bit rates higher than 3.36 bits/sample it is the best to use the quantizer with *N*=32 levels. In the same manner from the Fig. 3 can be concluded that it is better to use the quantizer with 16 levels for the bit rates up to 3.518 bits/sample, while for the higher bit rates better quality is achieved by using the quantizer with *N*=32 levels.

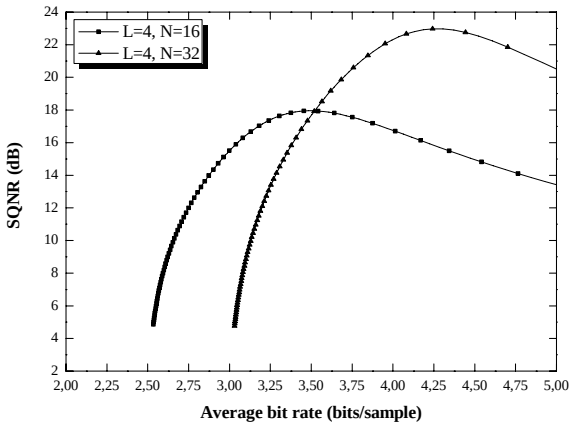


Fig. 3. The graphical interpretation of functionality of signal-to-quantization noise ratio from average bit rate for the hybrid quantizer, *L*=4 segments and *N*=8, 16 and 32 levels

It is known that the signal-to-quantization noise ratio of the standard uniform quantizer increases 6.02 dB with each increase of the bit rate for one bit, when the coding technique with constant length of the code words is used [4]. The thing which interested us, is the value of the average bit rate, of the hybrid quantizer, to which it can be increased by one bit, so that the gain in quality still be higher than that achieved with the standard uniform quantizer, or just to be greater than 6.02 dB. Specific values are shown in Table III.

TABLE III
VALUES UP TO WITCH IT MAKES SENSE TO INCREASE THE AVERAGE BIT RATE

$\overline{R}_{\text{lim.}}$ (bits/sample)	L=1	L=2	L=4
N=8	2.492	2.466	/
N=16	2.821	3.126	3.201
N=32	3.676	3.744	3.953

The analysis was performed for all combinations of the number of levels *N* and the number of segments *L*. From the mathematical point of view, it means that the slope of the curve *SQNR* ($\overline{R}$), at any point up to the value of the limit bit rate $\overline{R}_{\text{lim.}}$, is greater than 6.02.

To indicate the advantages of the hybrid model, we compared its performances regarding to the signal quality and average bit rate, with the standard uniform and nonuniform quantizer.

In order to make the comparation properly, we applied the Huffman coding algorithm in all three cases. The advantages of the hybrid quantizer versus the uniform and nonuniform quantizer can be reflected in the fact of how much is lower the average bit rate achieved by the hybrid quantizer when the value of signal-to-quantization noise ratio is fixed, or what is the gain in quality achieved by the hybrid quantizer, when the value of the average bit rate is fixed. The values of the signal-to-quantization noise ratio and average bit rate, which are fixed, refer to the uniform and nonuniform quantizer, and these fixed values are maximal signal quality and average bit rate at which it is achieved. Compared to the uniform quantizer, the hybrid quantizer achieves the quality of 16 dB at a bit rate lower for 0.15 bits/sample and achieves the rise in quality of 0.82 dB at a bit rate of 3.82 bits/sample. Compared to the nonuniform quantizer, the hybrid quantizer achieves the quality of 17.55 dB at a bit rate lower for 0.152 bits/sample and achieves the rise in quality of 1.36 dB at a bit rate of 2.606 bits/sample. The above results demonstrate that the adequate selection of the number of quantization levels *N*, the proper distribution of the segments in the entire range of the hybrid quantizer, and proper coding algorithm can result in a better signal quality and better compression, at the same time.

V. CONCLUSION

In this paper is proposed the hybrid quantizer which output is coded using Huffman coding algorithm with variable length code words. Based on the theoretical results obtained by the proper analysis of the proposed hybrid quantizer we showed that its advantages clearly can be seen in terms of signal quality as well as in the value of the average bit rate. Therefore, we believe that the proposed hybrid quantizer with Huffman coding technique, represent an efficient solution when the desired signal quality is needed to be achieved with a lower average bit rate.

REFERENCES

[1] Z. Perić, M. Dinčić, M. Petković, “Design of a Hybrid Quantizer with Variable Length Code”, IOS Press, Fundamenta Informaticae, vol. 98, no. 2-3, pp. 233-256, 2010.

[2] Z. Perić, J. Nikolić, D. Pokrajac, “Hybrid scalar quantizer for the Laplacian source”, WSEAS Transactions on Communications, vol. 6, no. 1, pp. 60-65, 2007.

[3] Z. Perić, L. Velimirović, J. Lukić, A. Jocić and D. Denić, "Konstrukcija kvantizera promenljive dužine i male bitske brzine za Laplasov izvor", TELFOR'09, Conference Paper, pp.576-579, Belgrade, Serbia, 2009.

[4] N. Jayant, *Digital Coding of Waveforms, Principles and Applications to Speech and Video*, New Jersey, Prentice Hall, 1984.

[5] D. Salomon, *A Concise Introduction to Data Compression*, London, Springer, 2008.

[6] D. Drajić, *Uvod u teoriju informacija i kodovanje*, Beograd, Akademska misao, 2004.

[7] K. Sayood, *Introduction to Data Compression*, Amsterdam, Elsevier, 2006.

Optimization of Polynomial Approximation in Series Based Software Direct Digital Synthesis of Signals

Borko G. Boyanov[1] and Mariana Iv. Shotova[2]

***Abstract* – In this paper an optimization of the software method for Direct Digital Synthesis of signals, based on series approximation of the sine wave is discussed. Three known sine-wave approximations are compared considering the spurious free dynamic range of the spectrum. An optimization of the polynomial approximation is proposed and discussed here. The signal synthesis requires reduced number of mathematical operations by taking advantage of the sine wave symmetry. Additional increase of the dynamic range is achieved by genetic algorithm optimization of the polynomial coefficients.**

***Keywords* – DDS, series, polynomial, optimization, GAO.**

I. INTRODUCTION

The Direct Digital Synthesis (DDS) is a technique for generating a high quality sine wave through a digitally defined frequency. The software implementation of DDS (SDDS) based on digital signal processor has two main versions – using ROM table of the sine wave, and series approximation of the sine wave.

The first one is most common and faster, however, due to restricted ROM table size, the spurious free dynamic range (SFDR) D of the spectrum of the synthesized signal, which is measure of quality, is also limited [1].

The advantage of the SDDS exploiting series approximation is elimination of the ROM table. A drawback however is the bigger number of the required mathematical operations, which results in lower sampling frequency.

There are several basic sine wave polynomial approximations, which can be used in DDS [2].

Analog Devices Inc. [3] suggested a 5-th order polynomial approximation of the sine wave in the range [0, π/2]. The dynamic range D is around 105 dB. The number of operations is 5 multiplications and 4 addition/subtractions.

More recently, methods combining the use of smaller tables with the evaluation of a low-degree polynomials have been proposed [4], [5]. However some of them exhibit argument-dependent execution time, which is not acceptable in real word DDS.

The SDDS, proposed here, requires reduced number of mathematical operations by taking advantage of the sinus' symmetry – using approximation in the range [-π/2, π/2], thus eliminating the even-order components of the polynomial. This is implemented here using MATLAB and its ***polyfit*** function. The SDDS suggested here, is based on 7-th order polynomial and features dynamic range of 128dB.

The complicated nature of the effect of the polynomial coefficients on the spectral spur levels is factor, which suggests utilization of generic algorithm optimization (GAO) of the coefficients, aimed at minimization of the spectral spur levels. The application of GAO to SDDS results in sets of coefficients of the 7-th order polynomial, which increase the SFDR up to 133dB.

II. SINE WAVE SERIES AND POLINOMIALS

The series, which can be used for sine wave synthesis is [2]

$$\sin(\alpha) = \alpha - \frac{\alpha^3}{3!} + \frac{\alpha^5}{5!} - \frac{\alpha^7}{7!} + \ldots \qquad (1)$$

In this series, due to sinus' odd-symmetry, even-order components are excluded. To minimize the error due to discarded series' component, the argument range is limited to [0, π/2]. A plot of calculation error $e_{(\alpha)}$, over the range [0, 2π] when 4 components of the series are used, is shown in Fig.1 with solid line. The dashed line represents the discarded 9-th order component $x_{9(\alpha)}$, which dominates the error.

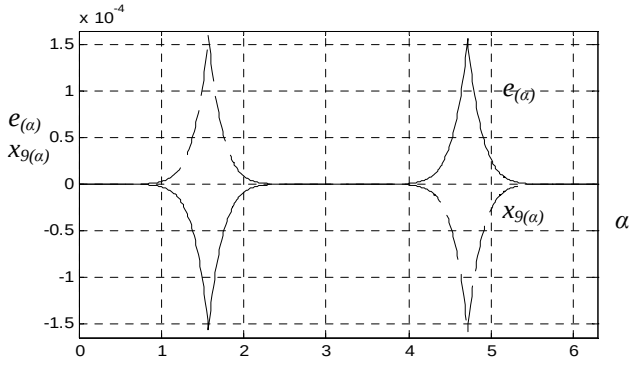


Fig.1. Calculation error $e_{(\alpha)}$ and 9-th order component $x_{9(\alpha)}$.

The elimination of the high-order components of the series results in large error at angles close to odd multiples of π/2, while the error at angles close to multiples of π is very small.

In the case of the DDS, the error spectrum is of interest. An error signal, which contains 16 periods of error "wave" $e_{(\alpha)}$ is composed, and FFT is applied. The main part of the spectrum is shown in Fig.2. The spectral line at k = 16 is the fundamental frequency, while the other spectral lines represent odd-order harmonics. Since the amplitude of the synthesized signal is A = 1, (0 dB), the SFDR is defined by the level of the third harmonic at k = 48 $D = -L_3$ = 91.4dB.

[1]Borko G. Boyanov is with the Faculty of Electronics, Technical University of Varna, 9010, Varna, Bulgaria, E-mail: bgboyanov@gmail.com

[2]Mariana Iv. Shotova is with the Faculty of Electronics, Technical University of Varna, 9010, Varna, Bulgaria, E-mail: marianashotova@gmail.com

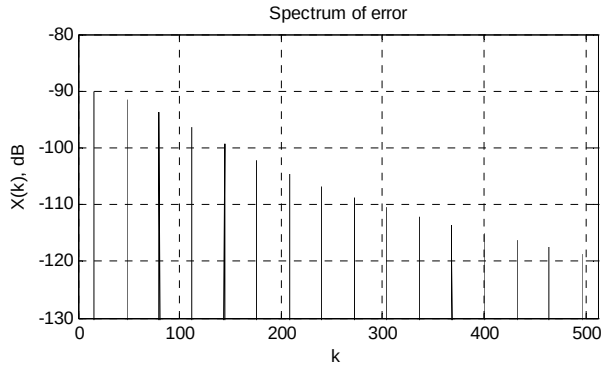


Fig.2. Main part of the FFT error spectrum.

For its 16-bit DSP Analog Devices Inc. suggested a 5th order polynomial approximation of the sine wave in the range [0, π/2]

$$\sin(\alpha) = 3.140625\beta + 0.02026367\beta^2 - 5.325196\beta^3 + 0.5446778\beta^4 + 1.800293\beta^5 , \quad (2)$$

where $\beta = \alpha/\pi$, $0 \le \beta \le 0.5$.

Rearranging the series in the form, which is closer to DSP,

$$\sin(\alpha) = \beta(a_1 + \beta(a_2 + \beta(a_3 + \beta(a_4 + \beta a_5)))), \quad (3)$$

the number of the required operations is 5 multiplications and 4 addition/subtractions.

A plot of calculation error $e_{(\alpha)}$, over the range [0, 2π] with this approximation is shown in Fig.3.

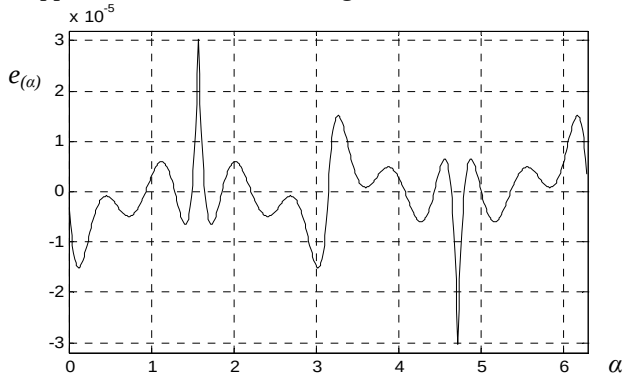


Fig.3. Sine wave error with ADI polynomial.

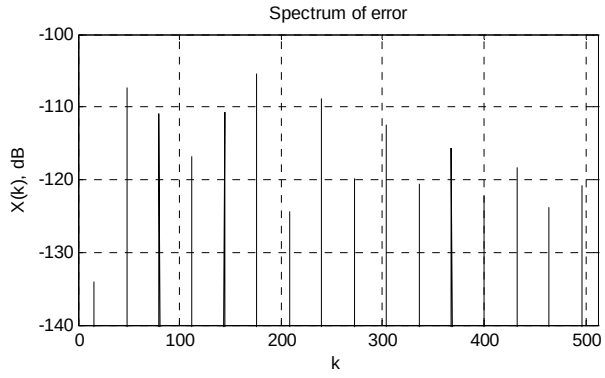


Fig.4. Main part of the error spectrum with ADI polynomial.

The main part of the spectrum is shown in Fig.4. The SFDR is defined by the level of the eleventh harmonic at k = 176 $D = -L_{11}$ = 105dB.

III. IMPROVEMENT TO THE POLINOMIAL APPROXIMATION

A drawback of the approximation (2) is that it does not take into account the sinus' odd symmetry. The approximation suggested here in the range [-π/2, π/2] eliminates the even-order components of the polynomial. This is implemented by MATLAB's *polyfit* function, which minimizes the root-mean-square error. The 7-th order polynomial obtained by this function is

$$\sin(\alpha) = a_1\alpha + a_3\alpha^3 + a_5\alpha^5 + a_7\alpha^7 , \quad (4)$$

where a_1 = 0.9999974, a_3 = -0.1666513, a_5 = 0.0083092, and a_7 = -0.00018437.

A plot of the error $e_{(\alpha)}$, over the range [-π/2, 3π/2] with this approximation is shown in Fig.5.

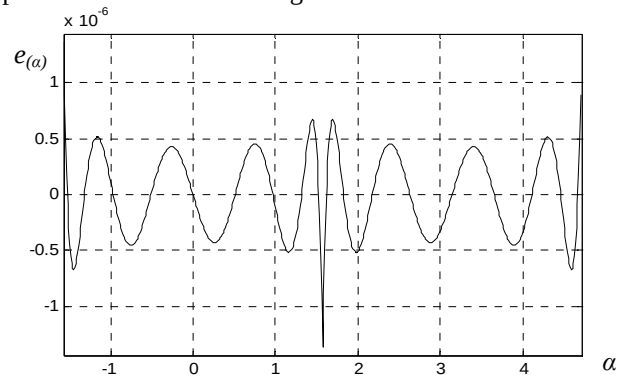


Fig.5. Sine wave error with 7th order polynomial.

The rearranged form of the polynomial is

$$\sin(\alpha) = \alpha(a_1 + \alpha^2(a_3 + \alpha^2(a_5 + a_7\alpha^2))), \quad (5)$$

where α^2 is calculated in advance. With this in mind the number of the required mathematic operations is 5 multiplications and 3 addition/subtractions.

The main part of the spectrum is shown in Fig.6.

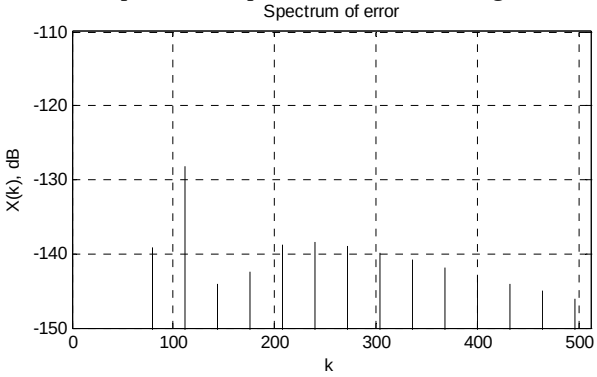


Fig.6. Error spectrum with 7th order polynomial.

The SFDR is defined by the level of the seventh harmonic at k = 112 , $D = -L_7$ = 128dB. Thus, exploiting the sinus' odd symmetry, the number of the polynomial components is reduced, while the dynamic range of the DDS is increased.

The seventh harmonic dominates the spectrum, and well the error in the time domain in Fig.5.

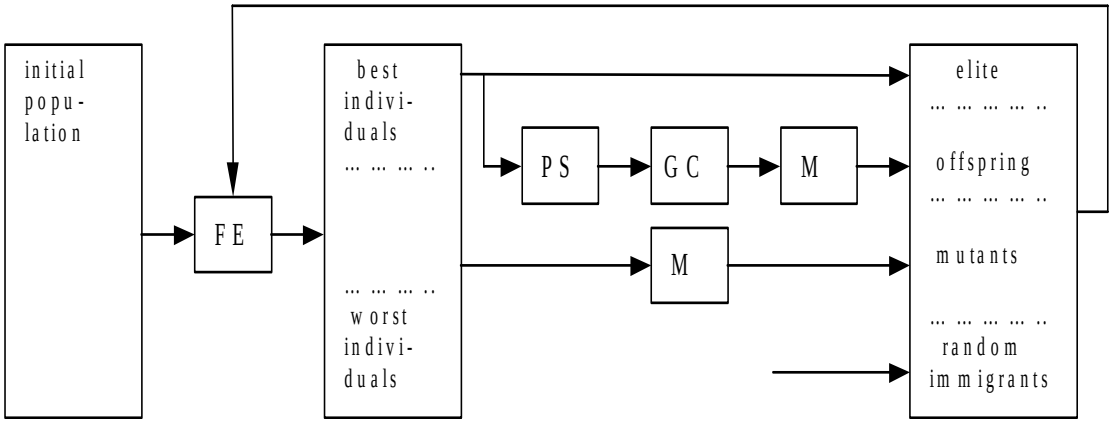


Fig.7. Generalized block diagram of a genetic algorithm.

IV. GENETIC OPTIMIZATION OF THE POLINOMIAL

A drawback of the *polyfit* function, from DDS point of view, is that the approximation is based on minimization of the root-mean-square error, while for DDS it is important to minimize the spectral spur levels.

The complicated nature of the effect of the polynomial coefficients on the levels of the harmonics is factor that suggests the utilization of generic algorithm. This is a relatively new optimization method, which provides stochastic search over the parameter space, guided by fitness evaluation towards specific goal. Although relatively slow, GAO can handle complex optimization problems, especially when the goal is to find near-best extreme in multimodal function domain. GAO also features global search from many parameter space points, capability to escape from local extreme [6], [7].

Generalized block diagram of a genetic algorithm is shown in Fig.7. Usually, GAO comprises three operations - Selection, Genetic operations, and Replacement, which are repeated in genetic cycle. As in nature, population consists of individuals, each represented by its chromosome. The chromosome encodes all parameters of an individual. The genetic cycle starts with evaluation of the fitness (FE) of each individual. Then the individuals are ranked, accordingly to their fitness values, and a particular group of best individuals (parents) is selected by operation, called parent selection (PS), to generate Offspring. Genetic operators as Genetic crossover (GC) and Mutation (M) are applied then to produce offspring. Exploiting replacement strategy individuals of the current population are replaced and thus a new population is composed. Such genetic cycle is repeated until certain termination criterion is reached (predefined number of genetic cycles, or fitness threshold).

Binary coding of the coefficients and chromosome

There are four coefficients a_i to be encoded. In order to facilitate the flexibility in exploring the parameter space, the following mapping is exploited

$$a_i = a_{i0} - \Delta_i / 2 + \delta_i \sum_{j=0}^{b-1} b_{ij} 2^j , \; i = 1,3,5,7. \qquad (6)$$

where a_{i0} are the initial values, calculated by *polyfit* function, Δ_i are the ranges of the variation, b_{ij} are the bits, b is the length of the binary string for each coefficient, and δ_i are the quantizing steps. These parameters are predefined for each coefficient Δ_1 = 0.00002, Δ_2 = 0.0002, Δ_3 = 0.0003, Δ_4 = 0.0020, $\delta_i = \Delta_i \, 2^{-b}$. In the developed GAO gene length b = 12 is chosen, which defines chromosome length L = 48 bits, and a search space of 2^{48} points.

Fitness evaluation

Since the objective of GAO is to reduce the spectral spur levels, the error signal and its spectrum are to be calculated. For each individual of the population the chromosome is split into genes and, accordingly to the mapping scheme (6), the coefficients are calculated. This set of coefficients is passed to the error and spectrum calculation function, which returns the levels of spectral spurs in decibels.

Parent selection and genetic operations

The number of the best individuals, selected for reproduction, is N_b = 30. These parents produce 60 child chromosomes by single-point crossover and double mutation with probability p_m = 0.995. Mixed replacement strategy is exploited: 30 best chromosomes are directly copied into the new population, which represents elitism, 210 mid-rank chromosomes undergo mutations and 120 random immigrants replace the worst chromosomes.

Results of the GA optimization

Tens of runs of the dedicated GAO program in MATLAB were performed, several of which resulted in sets of polynomial coefficients ensuring SFDR $D \approx$ 133dB. One such set is a_1=0.999997489622539, a_3=-0.166651585015804, a_5=0.008310170923798, and a_7 = -0.000184768956791.

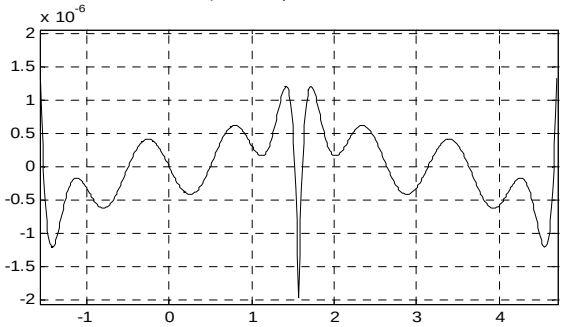


Fig.8. Sine wave error with GAO polynomial.

A plot of the error $e_{(\alpha)}$ over the range $[-\pi/2, 3\pi/2]$ with these coefficients is shown in Fig.8. The shape of the error, excluding the fundamental component, is more irregular, compared to Fig.5, spreading the energy over the spectrum.

The main part of the error spectrum is shown in Fig.9. The SFDR is defined by the seventh harmonic at $k = 112$, $D = -L_7 = 133$dB. The strongest component at $k = 16$ indicates deviation of the amplitude of the fundamental component from the nominal value **A**=1. Such an inaccuracy of the magnitude (relative value $\approx 5 \cdot 10^{-7}$) is negligible in practice.

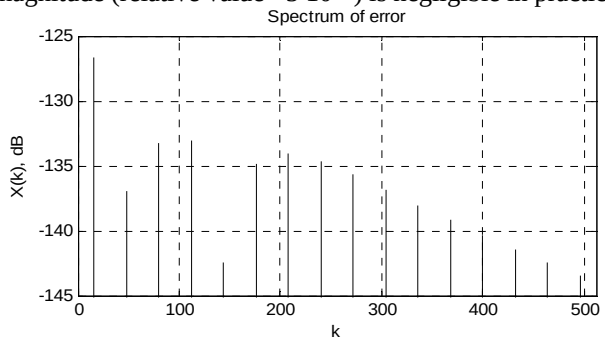


Fig.9. Error spectrum with GAO polynomial coefficients.

The development of the GA optimization over the time is illustrated in Fig.9, where the decrease of the maximal spur level L_s is shown.

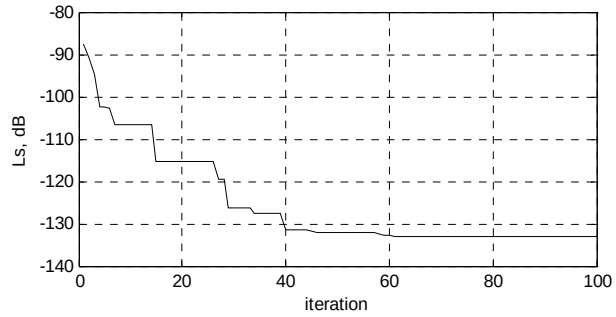


Fig.9. Development of the GAO over the time (iterations).

A distribution of the fitness of the individuals within the population at the end of the GAO is shown in Fig.10. About 30 of the individuals (the elite) have fitness close to the best one (-133dB).

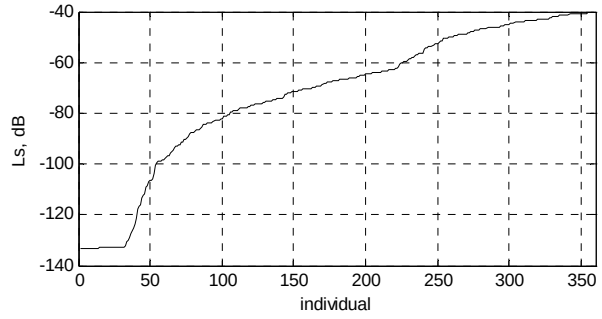


Fig.10. Fitness of the population at the end of GAO.

The GA optimization was performed on desktop machine with AMD Athlon 64X2 DualCore 5200+ 2.7GHz processor. Single run with specified GAO parameters and 1000 iterations takes approximately 80 s.

V. CONCLUSION

The optimization of the polynomial approximation of the sine wave aimed at Direct Digital Synthesis of sine wave signal was considered. Taking into account the symmetry of the sine wave, and applying GAO, minimizing the spectral spur levels, a polynomial of 7^{-th} order, which increases the spurious free dynamic range D up to 133 dB is suggested.

Further research, relevant to the problem, may focus on the following topics:

- polynomial approximation with polynomial of other order;
- quantizing effects in 16- and 32-bit implementations;
- fusing the GAO and the Gradient based search.

REFERENCES

[1] B. Boyanov, “Single table direct digital synthesis of signals, based on dsPIC digital signal controllers”, II International Congress on MEEMI 2005, Varna, Bulgaria, 2005.

[2] M. Abramowitz and I. A. Stegun. Handbook of mathematical functions with formulas, graphs and mathematical tables. Applied Math. Series 55. National Bureau of Standards, Washington, D.C., 1964.

[3] "ADSP-218x DSP Instruction Set Reference", Analog Devices, Inc, 2004.

[4] J.-A. Pineiro, S. Oberman, J.-M. Muller, and J. Bruguera, "High-Speed Function Approximation Using a Minimax Quadratic Interpolator", IEEE Transactions on Computers, vol. 54, No. 3, March 2005.

[5] N. Brisebarre, J.-M. Muller, and A. Tisserand. “Computing machine-efficient polynomial approximations” ACM Trans. Mathematical Software, 32(2): 236–256, June 2006.

[6] Foundations of Genetic Algorithms 2, Edited by L. Darrel Whittley, Morgan Kaufmann Publishers, San Mateo, CA, 1993.

[7] K. S. Tang, K. F. Man, S. Kwong and Q. He, “Genetic algorithms and their Applications”, IEEE Signal Processing Magazine, vol. 13, No. 6, pp. 22-37, Nov. 1996.

Text Line Segmentation by Gaussian Kernel Extended with Binary Morphology

Darko Brodić[1] and Branko Dokić[2]

***Abstract* – In this paper, new approach to the text segmentation by Gaussian kernel is presented. As a result of basic algorithm, defined area exploited for text segmentation and text parameter extraction. To improve text segmentation process, basic method is extended by binary morphological operations. Basic and extended algorithm is examined and evaluated under different text samples. Results are examined, analyzed and discussed.**

***Keywords* – OCR, Document image processing, Text Line Segmentation, Gaussian kernel, Morphological operation.**

I. INTRODUCTION

Printed and handwritten text is characterized by its attributes and features diversity. Hence, text parameters extraction procedure can be quite dissimilar one. But, such algorithm should be valid for printed as well as for handwritten text. Text line segmentation is the major step in document processing procedure. Although some text line detection techniques are successful for printed documents, processing of handwritten documents has remained a key problem in OCR [1,2]. Most text line segmentation methods are based on the assumptions that distance between neighboring text lines is significant as well as that text lines are reasonably straight. However, these assumptions are not always valid for handwritten documents. Hence, text line segmentation is a leading challenge in document processing.

Related work on text line segmentation can be categorized in few directions [3]: projection based methods, Hough transform methods, smearing methods, grouping methods, methods for processing overlapping and touching components, stochastic methods, others method.

Projection base methods have been primarily used for printed document segmentation, but it can be adapted for handwritten documents as well. It uses the vertical projection profile (VPP), which is obtained by summing pixel values along the horizontal axis for each y value. This is accomplished by finding its maximum and minimum value [4]. Because of method drawbacks, short lines will provide low peaks, and very narrow lines. Hence, method failed to be efficient for multi-skewed text lines.

The Hough transform [5] is a widespread technique for finding straight lines in the images. Consequently, image is transformed in the Hough domain. Potential alignments are hypothesized in Hough domain and validated in the image domain. The direction for the maximum variation is determined by a cost function. The "voting" function in Hough domain determine slope of the straight line [6].

In smearing methods the consecutive black pixels along the horizontal direction are smeared [7]. This way, enlarged area of black pixels is formed. It is so-called boundary growing area. Consequently, the white space between black pixels is filled with black pixels. It the valid only if their distance is within a predefined threshold.

Grouping methods is based on building alignments by aggregating them [8]. The units may be pixels or connected components, blocks or other features such as salient points. These units are joined together to form alignments. The joining scheme is based on both local and global criteria used for checking consistency. If the nearest neighbour belongs to another line, then the nearest-neighbour joining scheme will fail to group complex handwritten units.

Method for overlapping and touching components detects such components during the grouping process when a conflict occurs between two alignments [9]. Further, it applies a set of rules to label overlapping or touching components. The rules use as features the density of black pixels of the component in each alignment region, alignment proximity and positions of both alignments around the component. The frontier segment position is decided by analyzing the component VPP. If VPP includes two peaks, the cut will be done in the middle way from them. Otherwise, component will cut in two equal parts.

Stochastic method is based on probabilistic algorithm, which accomplished non-linear paths between overlapping text lines. These lines are extracted through hidden Markov modelling (HMM) [10]. This way, the image is divided into little cells. Each one them correspond to the state of the HMM. The best segmentation paths are searched from left to right. In the case of touching components, the path of highest probability will cross the touching component at points with as less black pixels as possible. However, the method may fail in the case that contact point contains a lot of black pixels.

In this paper, modification of the base method proposed in [12] are implemented, analyzed, examined and compared. It is simple and efficient method in terms of accuracy and computations. Its primary role is to perform text segmentation and to estimate the skew angle of document image. The proposed method is implemented and "measured" in different sample text examples and evaluated as well.

Organization of this paper is as follows. Section II includes brief description and information on proposed algorithm. In Section III text experiments are defined. Further, in Section IV given results are compared and discussed. In Section V conclusion is made and further investigation is pointed out.

[1]Darko Brodić is with the University of Belgrade, Technical Faculty Bor, Vojske Jugoslavije 12, 19210 Bor, Serbia, E-mail: dbrodic@tf.bor.ac.rs

[2]Branko Dokić is with the University of Banja Luka, Faculty of Electrical Engineering, Patre 5, 51000 Banja Luka, Republika Srpska, BIH, E-mail: bdokic@etfbl.net

II. PROPOSED ALGORITHM

The principal stages in document processing system are scanning, binarization, text segmentation, text parameter extraction, text recognition and conversion to ASCII. However, that procedure can be represented with three main stages as shown in Fig.1.

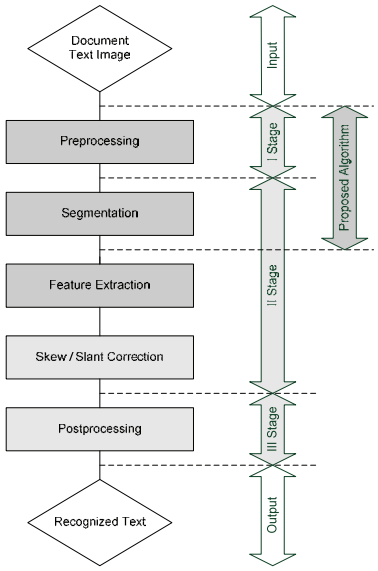


Fig.1. Document processing procedure

In preprocessing stage, algorithms for document text image binarization and normalization are applied. During the processing stage, algorithms for text segmentation as well as for reference text line estimation and skew rate identification are enforced. After that, skew angle is corrected. At the end, in postprocessing stage character recognition process is applied. Final sub stage is the conversion to ASCII characters.

A few assumptions should be made before algorithm description. In this paper, there is an element of preprocessing. After that, document text image is prepared for feature extraction. Main tasks are text segmentation as well as text parameter extraction, specifically reference text line identification and skew rate estimation.

Document text image is an input of text grayscale image described by following intensity function:

$$D\,(m,n) \in [0,\ldots,255] \quad , \tag{1}$$

where m ∈ [0, M-1] and n ∈ [0, N-1].

After applying intensity segmentation with binarization, intensity function is converted into binary intensity function given by:

$$D_{bin}(m,n) = \begin{cases} 1 & for \quad D(m,n) \geq D_{th} \\ 0 & for \quad D(m,n) < D_{th} \end{cases} \quad , \tag{2}$$

where D_{th} is given by Otsu algorithm [13].

Now, extracted text lines are represented as digitized document image by matrix **X** featuring M rows by N columns. Further, document text image is represented as black and white image. Hence, it consists of the only black and white pixels. Each character or word consists of the only black pixels. Every pixel $X(i,j)$ is represented by number of coordinate pairs such as:

$$X\,(i,\,j) \in [0,255] \quad , \tag{3}$$

where i = 1, …, M, j = 1, …, N of matrix **X** [14]. In addition, value 0 represents black pixels, while value 255 represents white pixels.

Prior to processing stage, document text image should be "prepared" for it. It is assumed text area is extracted by some appropriate method. Further, morphological preprocessing is performed to make document text image "noiseless". The morphological preprocessing was defined in [14-15] by following steps: document image erosion: $\mathbf{X} \ominus \mathbf{S}_1$, document image opening: $\mathbf{X} \circ \mathbf{S}_1$, dilatation of the opening the document image: $(\mathbf{X} \circ \mathbf{S}_1) \oplus \mathbf{S}_1$ and closing of the opening the document image: $(\mathbf{X} \circ \mathbf{S}_1) \bullet \mathbf{S}_1$. For these morphological operations, structuring element $\mathbf{S}_1$ dimension 3x3 is used [14-16].

For the processing stage Gaussian kernel algorithm is used. It is based on 2D Gaussian function given by [12]:

$$f(x,y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-b_x)^2+(y-b_y)^2}{2\sigma^2}} \quad , \tag{4}$$

where b_x is shift along x-axis, b_y is shift along y-axis, σ is curve spread parameter and A is the amplitude.

From (4) it is obvious that curve spread parameter σ is equal for x as well for y-axis. This way, Gaussian function is isotropic. Converting Gaussian function into point spread function, Gaussian kernel is obtained.

Algorithm using Gaussian kernel expands black pixel area by scattering every black pixel in its neighborhood. Around every black pixel new pixels are non-uniformly dispersed. Those pixels have lower intensity of black i.e. level of grayscale. Its intensity depends on their position i.e. distance from original center black pixel. Now, document image matrix is represented as grayscale image. Hence, intensity pertains in level region {0-255}. Black pixel of interest has coordinate $X(i,j)$ and intensity of 255, while neighbor pixels have around coordinates and intensity smaller than 255 i.e. grayscale level. So, after applying Gaussian kernel, equal to 2K+1 in x-direction as well as in y-direction, text is scattered forming enlarged area around it. Converting all non black pixels in the same area, as well as inverting image, forms the black pixel expanded areas. Those areas named boundary growing areas.

Boundary growing areas form control image with distinct objects that are prerequisite for document image text segmentation. These objects represent different text lines needful for text segmentation i.e. for disjoining text lines.

To further extend boundary growing area made some additional method is needed. Morphological dilatation is one of the possible solutions. It is given as [16]:

$$X \oplus S_2 \quad . \tag{5}$$

Dilatation structuring element $\mathbf{S}_2$ is used. It is a line defined by its height h=1, width w=2(R-K)+1 and parameter λ=R/K.

Main difference between original algorithm [12] and our approach are in text segmentation domain. In our approach, morphological dilatation extends boundary growing area leading to better text segmentation. Morphologically expanded boundary growing area is given in Fig.2.

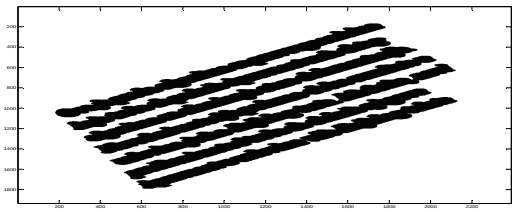


Fig.2. Morphologically expanded boundary growing areas (λ>1)

After segmenting text into separate text lines, primary task is text parameters extraction. Reference text line and skew

III. EXPERIMENTS

Algorithm quality examination consists of few text experiments representing test procedure. Basic and extended approach to algorithm is evaluated by combined text experiments framework. In this paper, only text line segmentation quality is examined. It is based on the following tests: multi-line text segmentation test, multi-line waved text segmentation test, and multi-line fractured text segmentation text. The schematic test procedure is shown in Fig.3.

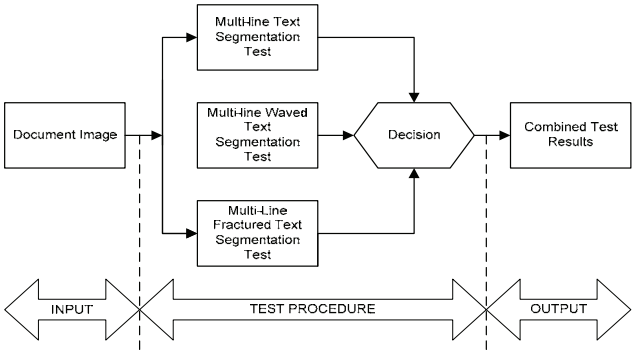


Fig.3. Schematic test procedure

The first experiment in combined text experiments framework is multi-line printed text sample [17]. This text with its skew angle parameter α is shown in Fig.4.

(a) (b)

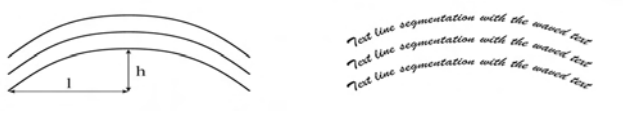
Fig.4. (a) Printed multi-text skew definition (α is parameter), (b) Printed multi-line text sample

Consequently, the number of existing text objects in multi-line text image relate to text segmentation quality success. Hence, the less objects the better segmentation process, except the number may not be less than text lines number. As a quality measure, the root mean square error $RMSE_{seg}$ has been used. It is calculated as [17,18]:

$$RMSE_{seg} = \sqrt{\frac{1}{P}\sum_{k=1}^{P}(O_{k,ref} - O_{k,est})^2} \quad , \qquad (6)$$

where k=1,…,P is the number of examined text samples, $O_{k,ref}$ is the number of referent objects in text i.e. number of text lines, and $O_{k,est}$ is the number of obtained objects in text by the applied algorithm.

The second text line segmentation experiment is based on multi-line waved text [17]. Sample text is formed as a group of text lines using the waved referent line as a basis. Referent line is defined by the parameter ε = h/l. Typically, ε is used from the set {1/8,1/6,1/4,1/3,…}. Multi-line waved text sample for this experiment is shown in Fig.5.

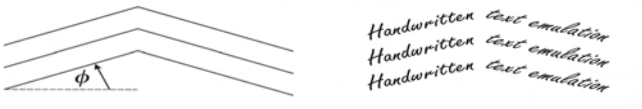


(a) (b)

Fig.5. (a) Waved text referent line shape definition (h and l are parameters), (b) Waved multi-line text sample

Similarly, the number of existing text objects after applied algorithm relate to the text segmentation quality success. Again, for the quality measure, *RMSE* has been used. Currently in (6) instead of RMS_{seg}, k, P, $O_{k,ref}$ and $O_{k,est}$ variables $RMSE_{seg,wav}$, l, R, $O_{l,ref}$ and $O_{l,est}$ are used, respectively. Alternatively, l=1,…,R is the number of examined text samples, $O_{l,ref}$ is the number of referent objects in text i.e. number of text lines, and $O_{l,est}$ is the number of obtained objects in text by the applied algorithm.

The last text line segmentation experiment is based on multi-line fractured text [17]. This text is formed by using the fractured referent line as a basis. Fractured text referent line is defined by the slope angle ϕ, as a parameter. Typically, ϕ is used from the set {5°,10°,15°,20°}. Multi-line fractured text for the last segmentation experiment is shown in Fig.6.



(a) (b)

Fig.6. (a) Fractured text reference line slope definition (ϕ is parameter), (b) Fractured multi-line text

Again, the number of existing text objects relate to the text segmentation quality success. *RMSE* has been used as a quality measure. Currently in (6) instead of RMS_{seg}, k, P, $O_{k,ref}$ and $O_{k,est}$ variables $RMSE_{seg,frac}$, m, Q, $O_{m,ref}$ and $O_{m,est}$ are used, respectively. Alternatively, m=1,…,Q is the number of examined text samples, $O_{m,ref}$ is the number of referent objects in text i.e. number of text lines, and $O_{m,est}$ is the number of obtained objects in text by the applied algorithm.

In above experiments, printed text could be interchanged by handwritten text written on the defined shape referent line.

IV. RESULTS AND DISCUSSION

In the first experiment, character height $H_{ch} \approx 100$ px is used. From [19] parameter K value may not exceed 1/5 of H_{ch}. In fact, bigger K could lead to text lines merging. Number of objects inspection is given in Fig.7.

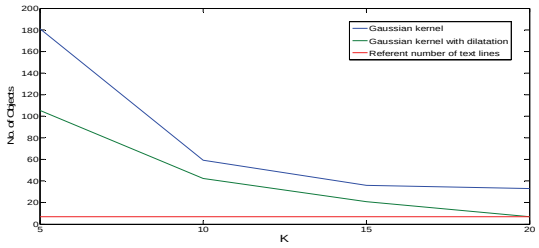


Fig.7. Number of objects from multi-text segmentation experiment

It is obvious that obtained results for extended algorithm are quite better than for original one. Further, in Fig.7 specific case is for K=20. Still, enlarging K above 20 is forbidden. Furthermore, K should be up to 20% of the H_{ch} to significantly close to boundary condition [18].

$RMS_{seg,wav}$ and $RMS_{seg,frac}$ is shown in Fig.8. and 9., respectively.

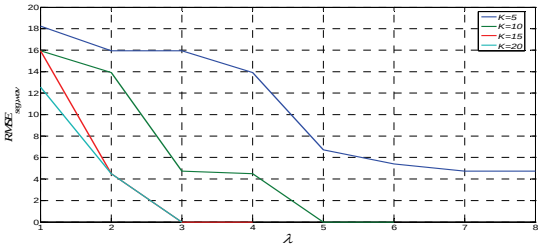


Fig.8. $RMS_{seg,wav}$ from waved text segmentation experiment

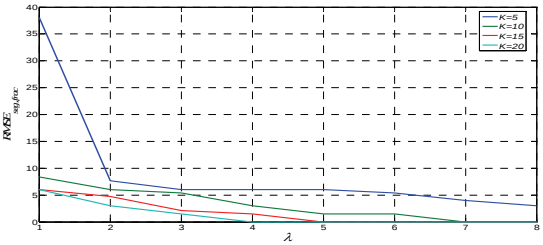


Fig.9. $RMS_{seg,frac}$ from fractured text segmentation experiment

Consequently, summary results for the original Gaussian kernel algorithm are not so convincing. Because of its faulty results from segmentation experiment, Gaussian kernel extended by morphological dilatation is promising.

V. CONCLUSION

In this paper, new approach to Gaussian kernel algorithm for text segmentation is presented. It assumes creation of boundary growing area around text based on Gaussian kernel algorithm extended by morphological dilatation. Those boundary growing areas form control image with distinct objects that are prerequisite for text segmentation. After text segmentation, reference text line and skew rate are calculated based on numerical method. Algorithm quality and robustness is examined by three experiments. Results are evaluated by RMSE method. All obtained results are compared with basic Gaussian kernel method.

Extended algorithm proved to be advanced in the domain of text segmentation which is of primary importance. Further investigation should be toward creating optimal adjusted and dilated Gaussian kernel rotated by the initial skew step.

REFERENCES

[1] A. Amin, S. Wu, "Robust Skew Detection in mixed Text/Graphics Documents", 8th International Conference on Document Analysis and Recognition (ICDAR 2005), Conference Proceedings, Seoul, Korea, 2005.

[2] S. Basu, C. Chaudhuri, M. Kundu, M. Nasipuri, D. K. Basu, "Text Line Extraction from Multi-Skewed Handwritten Documents", Pattern Recognition, vol. 40, pp. 1825-1839, 2006.

[3] Likforman-Sulem, L., Zahour, A., Taconet, B., "Text Line Segmentation of Historical Documents: A Survey", International Journal on Document Analysis and Recognition (IJDAR), vol.9(2-4), pp.123-138, 2007.

[4] Silva, L. F., Conci, A., Sanchez, A., "Automatic Discrimination between Printed and Handwritten Text in Documents", In Proceedings of XXII Brazilian Symposium on Computer Graphics and Image Processing, pp. 261-267, Rio de Janeiro, Brazil, 2009.

[5] Ballard, D. H., "Generalizing the Hough Transform to Detect Arbitrary Shapes", Pattern Recognition, vol.13(2), pp. 111-122, 1981.

[6] Amin, A., Fischer, "A Document Skew Detection Method Using the Hough Transform", Pattern Analysis & Applications, vol.3(3), pp. 243-253, 2000.

[7] Shi Z., Govindaraju V., "Line Separation for Complex Document Images Using Fuzzy Runlength", In Proceedings of the International Workshop on Document Image Analysis for Libraries, Palo Alto, U.S.A., 2004.

[8] Likforman-Sulem, L., Faure, C., "Extracting Lines on Handwritten Documents by Perceptual Grouping, in Advances in Handwiting and drawing: a multidisciplinary approach", Faure, C., Keuss, P., Lorette, G., Winter A. (Eds.), pp. 21-38, Europia, Paris, 1994.

[9] Zahour, A., Taconet, B., Mercy, P., Ramdane, S., "Arabic hand-written text-line extraction", In Proceedings of the 6th International Conference on Document Analysis and Recognition (ICDAR'01), pp. 281-285, Seattle, U.S.A., 2001.

[10] Koshinaka, T., Ken'ichi, I., Akitoshi, O., "An HMM-based Text Segmentation Method using Variational Bayes Approach", IEIC Technical Report, vol.104(87), pp. 19-24, 2004.

[11] Brodić, D., Milivojević, Z., "An Approach to Modification of Water Flow Algorithm for Segmentation and Text Parameters Extraction", In Emerging Trends in Technological Innovation. Camarinha-Matos, L.M., Pereira, P., Ribeiro, L., (Eds.), IFIP AICT, vol.314, pp. 324-331, Springer, Boston, U.S.A., 2010.

[12] Y. Li, Y. Zheng, D. Doermann, S. Jaeger, "A New Algorithm for Detecting Text Line in Handwritten Documents", 18th International Conference on Pattern Recognition (ICPR 2006), Conference Proceedings, vol.2, pp. 1030-1033, Hong Kong, China, 2006.

[13] Otsu, N., "A threshold selection method from gray-level histograms", IEEE Transactions on Systems, Man, and Cybernetics, vol. 9(1), pp. 62–66, 1979.

[14] R. C. Gonzalez, R. E. Woods, Digital Image Processing, 2nd ed., New Jersey, Prentice-Hall, 2002.

[15] M. Sonka, V. Hlavac, R. Boyle, Image Processing, Analysis and Machine Vision, Toronto, Thomson, 2008.

[16] Y. F. Shih, Image Processing and Mathematical Morphology – Fundamentals and Applications, Boca Raton, CRC Press, Taylor & Francis Group, 2009.

[17] Brodić D., Milivojević D. R., Milivojević Z., "Basic Test Framework for the Evaluation of Text Line Segmentation and Text Parameter Extraction", Sensors, vol.10(5), pp. 5263-5279, 2010.

[18] Bolstad, W. M., Introduction to Bayesian Statistics, John Wiley & Sons, NJ, U.S.A

[19] Brodić D., Dokić B., "Initial Skew Rate Detection Using Rectangular Hull Gravity Center", 14th International Conference on Electronics (E2010), Vilnius, Lithuania, 2010.

Performance of the Hybrid Method of Image Retrieval

Igor Stojanovic[1], Sofija Bogdanova[2], Momcilo Bogdanov[3] and Ivan Kraljevski[4]

***Abstract* – We study the performance of the hybrid method of image retrieval (content-based followed by pixel-based retrieval) from image databases. We carry out a series of image search experiments that cover the following scenarios: a) the given image is present in the database; b) copies of the given image are present but with different names; c) similar (but not identical) images are present; and d) the given image is not present. Experiments are performed with databases of up to 1000 images, using the Oracle database and the Matlab component Database Toolbox for operations with databases.**

***Keywords* – Database, image retrieval, normalized correlation, progressive wavelet correlation.**

I. INTRODUCTION

Images, drawings, and photographs have been part of everyday life for a long a time as a means of communication among people. The easy to use World Wide Web, the reduced price of storage devices as well as the increased calculating power allow an essential and efficient management of large quantities of digital information.

However, image digitalization systems do not allow for an easy management of collections of images. The need for efficient storage and retrieval of images was recognized by managers of large collections of images long time ago and was studied at a workshop sponsored by the American National Science Foundation in 1992. Data representation, feature extractions and indexing, image query matching, and user interfacing were identified as areas with a potential for improvement. One of the issues identified was the problem of locating a picture in a large and diverse collection of images.

The earliest and the most sophisticated descriptor-based image search engine is IBM QBIC [1]. Examples for such tools are WebSEEk and ImageRover. All of these techniques improve the accuracy and usefulness of image searching and may have commercial appeal in the future. The accuracy of searching has a tendency to improve by means of increasing the descriptors' resolution and pre-processing images for locating more appropriate descriptors.

[1]Igor Stojanovic is with the Faculty of Computer Science, University Goce Delcev, Toso Arsov 14, 2000 Stip, Macedonia, E-mail: igor.stojanovik@ugd.edu.mk

[2]Sofija Bogdanova is with Faculty of Electrical Engineering and Information Technologies, Karpos II b.b., P.O. Box 574, 1000 Skopje, Macedonia, E-mail: sofija@ feit.ukim.edu.mk

[3]Momcilo Bogdanov is with Faculty of Electrical Engineering and Information Technologies, Karpos II b.b., P.O. Box 574, 1000 Skopje, Macedonia, E-mail: bogdanov@feit.ukim.edu.mk

[4]Ivan Kraljevski is with Faculty for ICT, FON University, bul. Vojvodina bb, 1000 Skopje, Macedonia; E-mail: ivan.kraljevski@fon.edu.mk

The described image search engines are based on compact image representation in order to produce high processing speed. Generally there is no time for image analysis during the search phase.

The main descriptor-based search techniques are limited in two ways: Firstly, they do not contain the details of the image, and secondly, they do not cover the overall features within the frames of the image descriptors. Large complex images can contain many details and as a result a suitable compact presentation will be hard to determine.

On the other side, pixel-based search represents a somewhat different direction that has promise for applications that require high resolution, such as satellite images and medical images, particularly in geosciences. Pixel-based search techniques work by locating a particular pattern in a given image library. Popular criteria for matching are the normalized correlation coefficients [2], which measure the differences between images and patterns from the library. The particular strength of these criteria is that they are insensitive to uniform differences in brightness.

Some of the work done in the area of PWC (Progressive Wavelet Correlation) [2] is outlined in Section II. Our proposal about an application of the hybrid method for searching images stored in a database is presented in section III. Results of experiments that use the hybrid system are presented in Section IV.

II. PROGRESSIVE WAVELET CORRELATION

A. Overview

The PWC pixel-based method for retrieval of images involves correlation in the frequency domain for fast computation, discrete cosine transform (DCT) in factorizing form and wavelet representation of signals for efficient compression [2]. The primary idea is based on Vaidyanathan's theorem [3] for computing convolutions of wavelet-packet representations of signals. Progressive wavelet correlation summarizes Vaidyanathan's results replacing the operation of convolution in the wavelet domain with the equivalent operation in the Fourier-transform domain.

In this method the correlation between the two signals is formed in their original domain, without reverting from the transform domain. This method is progressive in the sense that each resolution level is calculated based only on the preceding level; lower resolution levels are irrelevant.

The algorithm can be described as follow:

Step 1: A candidate image is coarsely correlated with the pattern. Every eighth point of the correlation is generated.

Step 2: It is determined whether the pattern suitably matches the candidate image. If not, then another candidate image may be chosen or the search abandoned.

Step 3: If the match was suitable, then the candidate image is medium correlated with the pattern. We obtain the correlation at indices that are multiples of 4 mod 8 of the full correlation.

Step 4: Another similar match test is performed.

Step 5: A candidate image is finely correlated with the pattern. The fine correlation obtains the correlation at indices that are multiples of 2 mod 8 and 6 mod 8 of the full correlation.

Step 6: Another similar match test is performed.

Step 7: Full correlation: Obtain the correlation at odd indices.

Step 8: If a suitable match is found for the fully correlated image, then the image has been found.

B. Extension to two dimensions

Let the image size be N by N. In step 1, we have 64 subbands of length $N^2/64$. We perform one step of the inverse 2D JPEG transfer function, and one 2D step of the forward Fourier transform function. The next step involves adding the 64 subbands point by point to create a 2D array of size $N/8$ by $N/8$. Taking the inverse Fourier transform, we obtain the correlations at points that lie on a grid that is coarser than the original pixel grid by a factor of 8 in each dimension. In step 2, we obtain 16 subbands of size $N^2/16$ by adding the 16 subbands point by point, and taking the Fourier inverse. We obtain the correlation values on a grid that is coarser than the original grid by a factor of 4 in each dimension. In step 3, we obtain 4 subbands of size $N^2/4$. Finally, in step 4, the full resolution is obtained.

Formulas for calculating normalized correlation coefficients that measure differences between images and patterns are given in [2]. Normalized correlation coefficients can be computed from the correlations described above. The normalization is very important because it allows for a threshold to be set. Such a threshold is independent of the encoding of the images.

III. HYBRID SYSTEM FOR IMAGE RETRIEVING

A. Adaptation of PWC for searching in a database

The progressive wavelet correlation provides guidelines on how to locate an image in the image library. To make this method practical, we must first decide how to store the images. The initial choice is to store them in a disk file system. This can be seen as the quickest and simplest approach. A better alternative that should be considered is to store those images in a database. Databases offer several strengths over traditional file system storage, including manageability, security, backup/recovery, extensibility, and flexibility.

We use the Oracle Database for investigation purposes. There are two ways of storing an image into the Oracle Database. The first one is the use of Large Objects – LOB, and the second one is the use of Oracle *inter*Media.

To store images into the database we use the BLOB datatype. After creation of one BLOB column defined table we also create a PL/SQL package with a procedure for loading images (named load). This procedure is used to store images into the database.

The implementation of the progressive wavelet correlation in Matlab and the connection between the algorithm with the database are the next steps. The Database Toolbox is part of an extensive collection of toolboxes for use with Matlab.

Before the Database Toolbox is connected to a database, a data source must be set. A data source consists of data for the toolbox to access, and information about how to find the data, such as driver, directory, server, or network names. Instructions for setting up a data source depend on the type of database driver, ODBC or JDBC. For testing purposes JDBC drivers were usually used [4].

After setting up the data source for connecting to and importing data from a database we have used several standard functions of the Matlab Database Toolbox. We can retrieve BINARY or OTHER Java SQL data types. However, the data might require additional processing once retrieved. For example, data can be retrieved from a MAT-file or from an image file. Matlab cannot process these data types directly. One needs knowledge of the content and might need to massage the data in order to work with it in Matlab, such as stripping off leading entries added by the driver during data retrieval.

The last step in the adaptation is to create Matlab applications that use the capabilities of the World Wide Web to send data to Matlab for computation and to display the results in a Web browser. In the simplest configuration, a Web browser runs on a client workstation, while Matlab, the Matlab Web Server (matlabserver), and the Web server daemon (httpd) run on another machine. In a more complex network, the Web server daemon can run on a separate machine [4].

The practical implementation of progressive wavelet correlation includes two main subsystems: the server subsystem and the client subsystem. The server subsystem handles the processes of image storing in a database and similarity measure. The client subsystem handles the process of queries.

B. Hybrid system

Pixel-based search using progressive wavelet correlation is too expensive computationally to apply to large collections of images, especially when it is possible to discover in advance that no match is likely. It has to be combined with descriptor-based search or some other means of reducing the search space. After descriptor matches narrow the search, pixel-based search can find matches based on detailed content.

Results obtained from a number of performed experiments by retrieving images from a database via application of both CBIR [5] and progressive wavelet correlation led us to the idea of combining them in order to make the best use of their positive features. The initial idea was extended into a developed proposal for a new algorithm intended for

searching and retrieving images from a database. The modular scheme of this algorithm is given in Fig. 1.

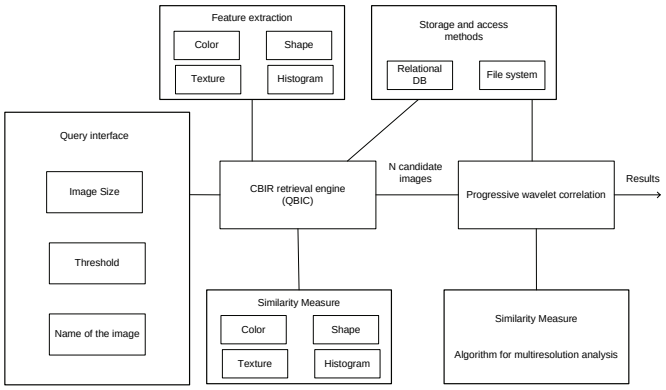


Fig. 1. A modular scheme for the proposed retrieval system

IV. Experimental results

This section represents experimental results obtained by means of image retrieval through an algorithm of progressive wavelet correlation.

Different experiments were set up as follows:

- The required image is included several times in the database with different names;
- The image is included only once in the database;
- Aside from the required image, the database also contains an image very similar to the required one (smudged in some parts or an image generally slightly different);
- The required image is not present in the database.

The experiments are carried out with databases including between 250 and 1000 store images. Oracle 10g version 10.1.0.2.0, served as our database, while we used Matlab version is 7.0.4.365 (R14) Service Pack 2 for image search.

Using QBIC, we established a database for the following characteristics of images: color, text, color histogram, and texture feature. We query the database for those images in the library with the most similar characteristics to the input image. On this set of candidates, we apply the normalized correlation coefficients to obtain the desired image. Using QBIC, we isolated ten candidate images based on the Color Histogram Feature. After that these images are subjected to detailed pixel-based search based on the normalized correlation coefficients.

A. The required image is present several times in the database

Two images, called flower01.jpg and flower10.jpg, served as search targets. Image flower01.jpg appears eight times under different names in the database, while the image flower10.jpg appears six times. The database contains images that are more visually similar to the image flower10.jpg.

Evaluation of the quality of the system concerning its precision p is estimated using the following the definition:

$$p = \frac{|A(q) \cap R(q)|}{|A(q)|} \qquad (1)$$

where q stands for query, $R(q)$ signifies a set of relevant images for the query in the database, while $A(q)$ stands for the set of images returned as a response to the set query q. The results obtained for precision are shown in tables.

TABLE I
FLOWER01.JPG

Threshold	Retrieved images	Precision
0.2	9	0.89
0.3	8	1
0.4	8	1
0.5 - 1	8	1

TABLE II
FLOWER10.JPG

Threshold	Retrieved images	Precision
0.2	10	0.6
0.3	10	0.6
0.4	10	0.6
0.5	10	0.6
0.6	10	0.6
0.9 - 1	6	1

The Table II shows that the image flower10.jpg, when in the database is more visually similar to her, is extracted with a threshold equal to or greater than 0.7.

B. The required image is present only once in the database

The results presented in this part refer to two different images 21.jpg and 40.jpg. Each of these images is included only once in the searched database.

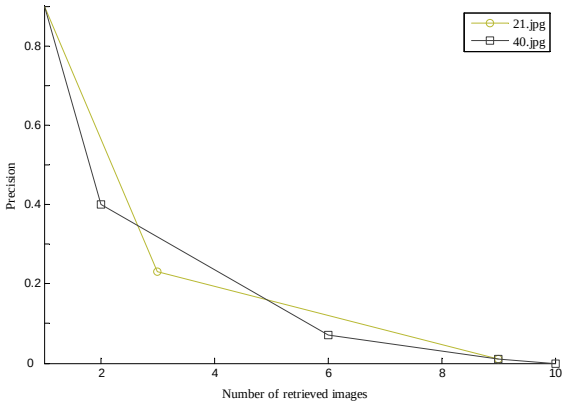


Fig. 2. Relationship between precision and the number of retrieved images for the images 21.jpg and 40.jpg

C. The database contains two very similar images

The next example refers to image 181.jpg. This example is specific because the database contains a similar image 183.jpg with its vertical sides slightly smudged. The similar images are shown in Fig. 3 and Fig.4.

Fig. 3. Image 181.jpg

Fig. 4. Image 183.jpg

In the previous examples, the number of candidate images that QBIC gives for further processing is 10, and in part *C* and *D* that number is 20. These images are then processed by PWC.

TABLE III
IMAGE 181.JPG

Threshold	Retrieved images	Precision
0.3	20	0.05
0.4	20	0.05
0.5	17	0.12
0.6	2	1
0.7	2	1
0.8	2	1

It is evident from Table III that when correlation threshold values are 0.6, 0.7 and 0.8, both the images 181.jpg and 183.jpg are retrieved. If the correlation threshold is equal to or greater than 0.9 only image 181.jpg is retrieved.

D. The required image is not present in the database

Table IV gives the number of retrieved images for different values of the correlation threshold for image 50.jpg, which is not present in the database. For correlation threshold values greater than or equal to 0.5 there are no images retrieved from the database.

TABLE IV
IMAGE 50.JPG

Threshold	Retrieved images
0.2	17
0.3	16
0.4	7
0.5 - 1	0

V. CONCLUSION

Based on our experience and experimental evidence we conclude that the hybrid system is a useful tool for image search from databases. The main feature of PWC is its high accuracy. With the choice of an adequate correlation threshold it is possible to detect if the given image is present in the database, whether there are images similar to the required one with different names, whether there are images slightly different from the required one, and whether the required image is present in the database.

We conclude that by changing the threshold value of correlation it is possible to identify all four cases examined by the hybrid system. In addition, the hybrid system locates and retrieves images faster than the PWC. This is achieved by using QBIC as a first step in the hybrid algorithm for obtaining a small number of candidate images.

The efficiency of joint content and pixel-based retrieval of images from a database can be improved in some applications such as satellite and medical imaging by appropriate selection of the threshold value. Content-based image retrieval is fast, but it normally gives more than one image due to the vast differences in perception capacity between humans and computers. On the other hand, pixel-based retrieval using progressive wavelet correlation is impractical, due to the numerous operations per image it entails. However, a positive outcome should be expected if we combine the good features of content-based and pixel-based searches: speed and accuracy.

REFERENCES

[1] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, P. Yanker, “Query by image and video content: The QBIC system,” IEEE Comp., vol. 28, pp. 23-32, Sept. 1995.

[2] H. S. Stone, “Progressive Wavelet Correlation Using Fourier Methods,” IEEE Trans Signal Processing, vol. 47, pp. 97-107, Jan. 1999.

[3] P. P. Vaidyanathan, “Orthonormal and biorthonormal filter banks as convolvers, and convolutional coding gain,” IEEE Trans. Signal Processing, vol. 41, pp. 2110-2130, June 1993.

[4] I. Stojanovic, S. Bogdanova, and M. Bogdanov, “Retrieving Images Using Content-Based Followed by Pixel-Based Search,” 15th Int. Conf. on Systems, Signals and Image Processing, pp. 271-274, Bratislava, Slovakia, June 25-28, 2008.

[5] Ritendra Datta, Dhiraj Joshi, Jia Li and James Z. Wang, ``Image Retrieval: Ideas, Influences, and Trends of the New Age," ACM Computing Surveys, vol. 40, no. 2, article 5, pp. 5:1-60, April 2008.

Applying of the Algorithm of Lagrange Multipliers in the Removal of Blur in Images

Igor Stojanovic[1], Ivan Kraljevski[2] and Slavcho Chungurski[3]

***Abstract* – This paper presents a non-iterative method that finds application in a broad scientific field such as digital image restoration. The method, based on algorithm of Lagrange multipliers, is use for the removal of blur in an X-ray caused by uniform linear motion. This method assumes that linear motion corresponds to an integral number of pixels. The resolution of the restored image remains at very high level. The main contribution of the method was found on the Improvement in Signal to Noise Ration (*ISNR*) that has been increase significantly compared to the classic techniques.**

***Keywords* – Image restoration, Lagrange multipliers, matrix equation, X-ray.**

I. INTRODUCTION

Images are produced to record or display useful information. Due to imperfections in the imaging and capturing process, however, the recorded image invariably represents a degraded version of the original scene. The undoing of these imperfections is crucial to many of the subsequent image processing tasks. There exists a wide range of different degradations that need to be taken into account, covering for instance noise, geometrical degradations (pin cushion distortion), illumination and color imperfections (under/ overexposure, saturation), and blur [1].

Blurring is a form of bandwidth reduction of an ideal image owing to the imperfect image formation process. It can be caused by relative motion between the camera and the original scene, or by an optical system that is out of focus. When aerial photographs are produced for remote sensing purposes, blurs are introduced by atmospheric turbulence, aberrations in the optical system, and relative motion between the camera and the ground. Such blurring is not confined to optical images; for example, electron micrographs are corrupted by spherical aberrations of the electron lenses, and Computed tomography scans suffer from X-ray scatter.

The field of image restoration (sometimes referred to as image deblurring or image deconvolution) is concerned with the reconstruction or estimation of the uncorrupted image from a blurred one. Essentially, it tries to perform an operation on the image that is the inverse of the imperfections in the image formation system. In the use of image restoration methods, the characteristics of the degrading system are assumed to be known a priori.

The paper concentrates on using algorithm of Lagrange multipliers and other different methods for removing the removal of blur in an X-ray caused by uniform linear motion. We assume that the linear motion corresponds to an integral number of pixels and is aligned with the horizontal (or vertical) sampling.

For comparison, we used two commonly used filters from the collection of least-squares filters, namely Wiener filter and the constrained least-squares filter [2]. Also we used in comparison the iterative nonlinear restoration based on the Lucy-Richardson algorithm [3], [4].

This paper is organized as follows. In the second section we present process of image formation, problem formulation and the method for reconstruction of the blurred image. We observe certain enhancement in the Improvement in Signal to Noise Ration (*ISNR*) compared with other standard methods for image restoration, which is confirmed by the numerical examples reported in the last section.

II. METHOD FOR RECONSTRUCTION OF THE BLURRED IMAGE

A. Image Formation

We assume that the blurring function acts as a convolution kernel or point-spread function $h(n_1, n_2)$ and the image restoration methods that are described here fall under the class of linear spatially invariant restoration filters. It is also assumed that the statistical properties (mean and correlation function) of the image do not change spatially. Under these conditions the restoration process can be carried out by means of a linear filter of which the point-spread function (PSF) is spatially invariant, i.e., is constant throughout the image. These modeling assumptions can be mathematically formulated as follows. If we denote by $f(n_1, n_2)$ the desired ideal spatially discrete image that does not contain any blur or noise, then the recorded image $g(n_1, n_2)$ is modeled as [2]:

$$g(n_1, n_2) = h(n_1, n_2) * f(n_1, n_2) = \sum_{k_1=0}^{N-1} \sum_{k_2=0}^{M-1} d(k_1, k_2) f(n_1 - k_1, n_2 - k_2). \quad (1)$$

[1]Igor Stojanovic is with the Faculty of Computer Science, University Goce Delcev, Toso Arsov 14, 2000 Stip, Macedonia, E-mail: igor.stojanovik@ugd.edu.mk

[2]Ivan Kraljevski is with Faculty for ICT, FON University, bul. Vojvodina bb, 1000 Skopje, Macedonia; E-mail: ivan.kraljevski@fon.edu.mk

[3]Slavcho Chungurski is with Faculty for ICT, FON University, bul. Vojvodina bb, 1000 Skopje, Macedonia; E-mail: chungurski@fon.edu.mk

The objective of the image restoration is to make an estimate $f(n_1, n_2)$ of the ideal image, under the assumption that only the degraded image $g(n_1, n_2)$ and the blurring function $h(n_1, n_2)$ are given.

B. Problem Formulation

The problem can be summarized as follows: let H be a $m \times n$ real matrix. Equations of the form:

$$g = Hf, g \in \Re^m; f \in \Re^n; H \in \Re^{m \times n}, \tag{2}$$

describe an underdetermined system of m simultaneous equations (one for each element of vector g) and $n = m + l - 1$ unknowns (one for each element of vector f). Where the index l indicates linear motion blur in pixels and $n = m + l - 1$.

The problem of restoring an X-ray that has been blurred by uniform linear motion, usually results of camera panning or fast object motion can be expressed as, consists of solving the underdetermined system (2). A blurred image can be expressed as:

$$\begin{bmatrix} g_1 \\ g_2 \\ g_3 \\ \mathrm{M} \\ g_n \end{bmatrix} = \begin{bmatrix} h_1 & \Lambda & h_l & 0 & 0 & 0 & 0 \\ 0 & h_1 & \Lambda & h_l & 0 & 0 & 0 \\ 0 & 0 & h_1 & \Lambda & h_l & 0 & 0 \\ \mathrm{M} & \mathrm{M} & \mathrm{M} & \mathrm{M} & \mathrm{M} & \mathrm{M} & \mathrm{M} \\ 0 & 0 & 0 & \Lambda & h_1 & \Lambda & h_l \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ \mathrm{M} \\ f_m \end{bmatrix} \tag{3}$$

The element of matrix H are defined as: $h_i = 1/l$ for i=1, 2,..., l.

The objective is to estimate an original row per row f (contained in the vector f^T), given each row of a blurred g (contained in the vector g^T) and a priori knowledge of the degradation phenomenon H. We define the matrix F as the deterministic original X-ray, its picture elements are F_{ij} for i=1,…, r and for j=1,…, n, the matrix G as the simulated blurred can be calculated as follows:

$$G_{ij} = \frac{1}{l} \sum_{k=0}^{l-1} F_{i,j+k} \quad \text{for } i=1,\ldots, r, j=1,\ldots, \mathrm{m} \tag{4}$$

with $n = m + l - 1$, where l is the linear motion blur in pixels. Equation (4) can be written in matrix form as:

$$G = \left(HF^T\right)^T = FH^T. \tag{5}$$

Since there is an infinite number of exact solutions for f or F in the sense that satisfy the equation $g = Hf$ or $G = FH^T$, an additional criterion that find a sharp restored matrix is required.

C. Method for Reconstruction Based on Lagrange Multipliers

Solution also is defined as the vector in the solution space of the underdetermined system $g = Hf$ (2) whose first n components has the minimum distance to the measured data, i.e. $\left\| \hat{f} - g \right\| \to \min$, where $\hat{f}$ are the first n elements of f. We can express vector $\hat{f}$ as $\hat{f} = Pf$, with P a $m \times n$ matrix which projects the vector f on the support of g [5], [6]:

$$P = \left[\begin{array}{c|ccc} & 0 & \Lambda & 0 \\ & 0 & \Lambda & 0 \\ I_m & \mathrm{M} & \mathrm{M} & \mathrm{M} \\ & 0 & \Lambda & 0 \end{array} \right] \tag{6}$$

The original optimization problem is now:

$$find \min_f \left\| Pf - g \right\|^2, \tag{7}$$

subject to the constraint $\left\| Hf - g \right\|^2 = 0$. Applying the technique of *Lagrange multipliers*, this problem can be alternatively formulated as an optimization problem without constraints:

$$V(f) = \lambda \left\| Hf - g \right\|^2 + \left\| Pf - g \right\|^2 \to \min \tag{8}$$

if λ is large enough. The solution of this problem is easy computing the partial derivative of criterion V respect to the unknown f:

$$\begin{aligned} &\frac{\partial}{\partial f} V(f) = 2\lambda H^T (Hf - g) + 2P^T (Pf - g) = 0 \\ &\hat{f} = \left[\lambda H^T H + P^T P\right]^{-1} \left[\lambda H + P\right]^T g. \end{aligned} \tag{9}$$

Matrix form of the solution of (9) is:

$$\hat{F} = G\left[\lambda H + P\right]\left[\left(\lambda H^T H + P^T P\right)^{-1}\right]^T. \tag{10}$$

III. Experimental Results

In this section we have tested the method based on Lagrange multipliers of X-ray images and present numerical results and compare with two standard methods for image restoration called least-squares filters: Wiener filter and constrained least-squares filter and the iterative method called Lucy-Richardson algorithm.

The experiments have been performed using Matlab programming language on an Intel(R) Core(TM)2 Duo CPU T5800 @ 2.00 GHz 32-bit system with 2 GB of RAM memory running on the Windows Vista Business Operating System.

In image restoration the improvement in quality of the restored image over the recorded blurred one is measured by the signal-to-noise ratio (*SNR*) improvement. The *SNR* of the

recorded (blurred and noisy) image is defined as follows in decibels [7]:

$$SNR_g = 10\log_{10}\left(\frac{\text{Variance of } f(n_1,n_2)}{\text{Variance of } g(n_1,n_2) - f(n_1,n_2)}\right)\text{(dB)}. \tag{11}$$

The *SNR* of the restored image is similarly defined as:

$$SNR_{\hat{f}} = 10\log_{10}\left(\frac{\text{Variance of } f(n_1,n_2)}{\text{Variance of } \hat{f}(n_1,n_2) - f(n_1,n_2)}\right)\text{(dB)}. \tag{12}$$

Then, the improvement in *SNR* is given by:

$$ISNR = SNR_{\hat{f}} - SNR_g = 10\log_{10}\left(\frac{\text{Variance of } g(n_1,n_2) - f(n_1,n_2)}{\text{Variance of } \hat{f}(n_1,n_2) - f(n_1,n_2)}\right)\text{(dB)}. \tag{13}$$

The improvement in *SNR* is basically a measure that expresses the reduction of disagreement with the ideal image when comparing the distorted and restored image. Note that all of the above signal-to-noise measures can only be computed in case the ideal image is available, i.e., in an experimental setup or in a design phase of the restoration algorithm.

The simplest and most widely used full-reference quality metric is the mean squared error (*MSE*) [8], computed by averaging the squared intensity differences of restored and reference image pixels, along with the related quantity of peak signal-to-noise ratio (*PSNR*). These are appealing because they are simple to calculate, have clear physical meanings, and are mathematically convenient in the context of optimization. The advantages of *MSE* and *PSNR* are that they are very fast and easy to implement. However, they simply and objectively quantify the error signal. With *PSNR* greater values indicate greater image similarity, while with *MSE* greater values indicate lower image similarity. Below *MSE*, *PSNR* are defined:

$$MSE = \frac{1}{rm}\sum_{i=1}^{r}\sum_{j=1}^{m}\left|f_{i,j} - \hat{f}_{i,j}\right|^2 \tag{14}$$

and,

$$PSNR = 20\log_{10}\left(\frac{MAX}{\sqrt{MSE}}\right)\text{(dB)}, \tag{15}$$

where *MAX* is the maximum pixel value.

The X-ray image making provides a crucial method of diagnostic by using the image analysis. Fig. 1, Original Image, shows such a deterministic original X-ray image. Fig. 1, Degraded Image, presents the degraded X-ray image for *l=40*. Finally, from Fig. 1, Lagrange multipliers Image, Wiener Restored Image, Constrained LS Restored Image and Lucy-Richardson Restored Image, it is clearly seen that the details of the original image have been recovered. These figures demonstrate four different methods of restoration, the method of Lagrange multipliers, Wiener filter, Constrained least-squares (LS) filter, and Lucy-Richardson algorithm, respectively.

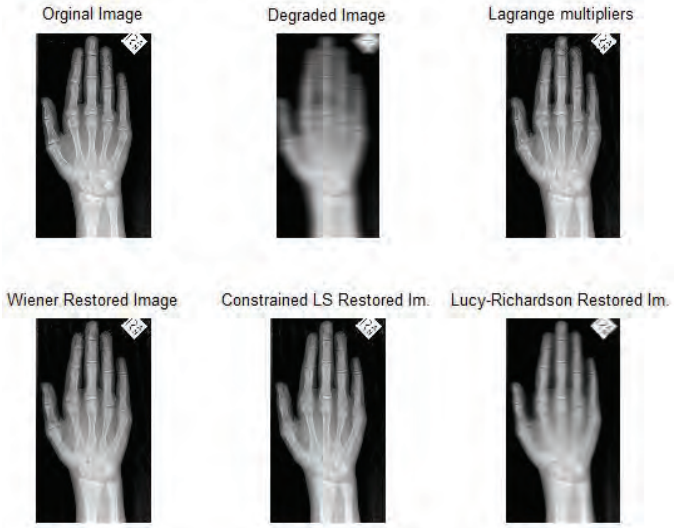


Fig. 1. Restoration in simulated degraded X-ray image for length of the blurring process, *l*=40

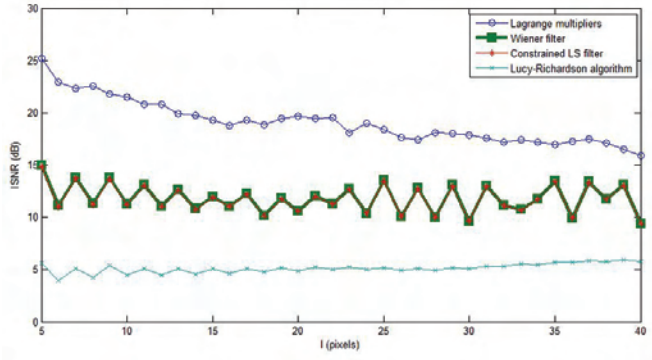


Fig. 2. Improvement in signal-to-noise-ratio vs. length of the blurring process in pixels

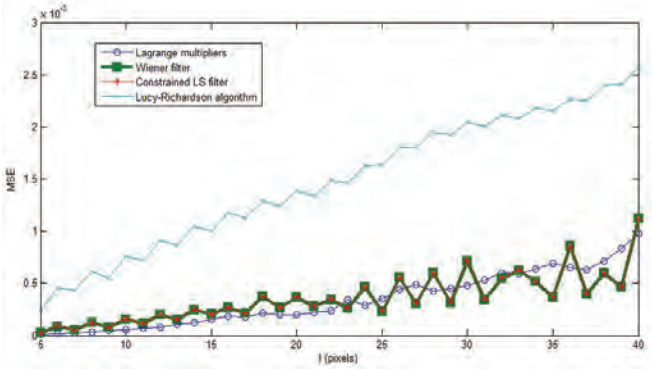


Fig. 3. Mean squared error vs. length of the blurring process in pixels

The difference in quality of restored images between the three methods is insignificant, and can hardly be seen by human eye. For this reason, the *ISNR* and *MSE* have been chosen in order to compare the restored images obtained by

the proposed method, the Wiener filter methods, the Constrained least-squares filter method and the Lucy-Richardson algorithm.

Fig. 2 and Fig. 3 shows the corresponding *ISNR* and *MSE* value for restored images as a function of *l* for the proposed method and the mentioned classical methods. The figures illustrate that the quality of the restoration is as satisfactory as the classical methods or better from them (l<40 pixels). Realistically speaking, large motions do not occur frequently in radiography.

The results present in Fig. 4 – 6 refer to another original image.

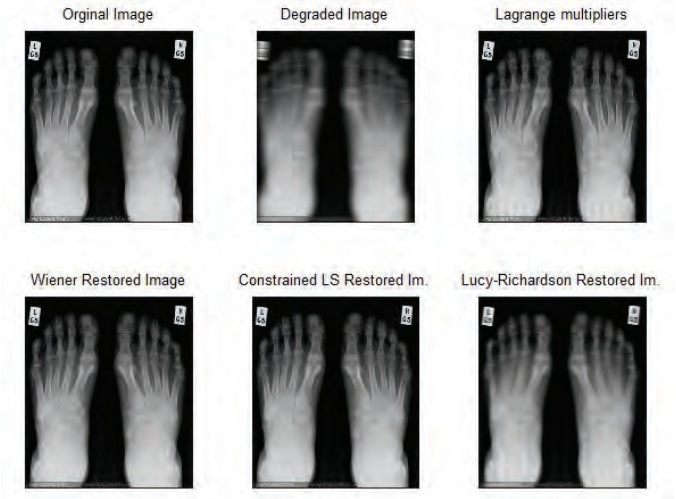


Fig. 4. Restoration in simulated degraded X-ray image for length of the blurring process, l=30

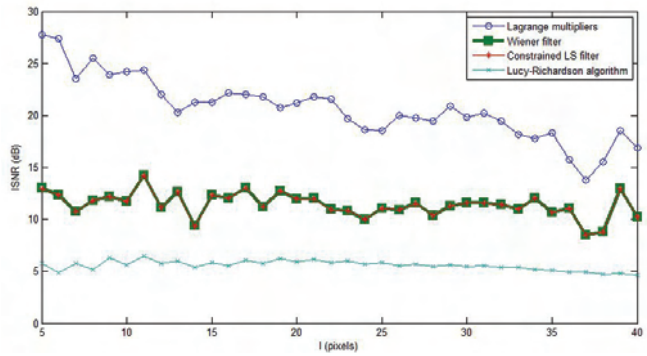


Fig. 5. Improvement in signal-to-noise-ratio vs. length of the blurring process in pixels

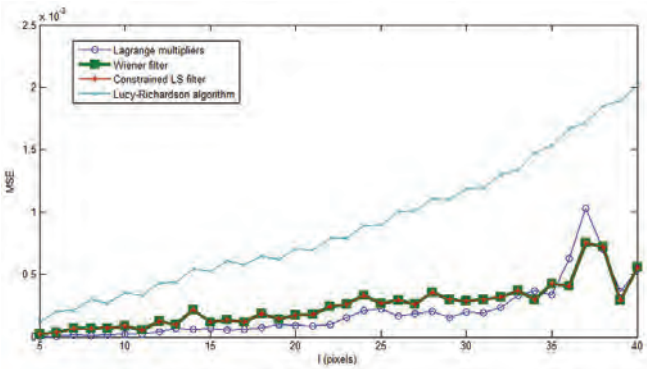


Fig. 6. Mean squared error vs. length of the blurring process in pixels

IV. Conclusion

We introduce a computational method, based on algorithm of Lagrange multipliers, to restore an X-ray that has been blurred by uniform linear motion.

We are motivated by the problem of restoring blurry images via well developed mathematical methods and techniques based on the Lagrange multipliers in order to obtain an approximation of the original image.

By using the proposed method, the resolution of the restored image remains at a very high level, although the main advantage of the method was found on the improvement *ISNR* that has been increased considerably compared to the other methods and techniques.

In this study, we present the results by comparing our method and that of the Wiener filter, Constrained least-squares filter and Lucy-Richardson algorithm, a well established methods used for fast recovered and high resolution restored images.

Obviously the proposed method is not restricted to restoration of X-ray images.

References

[1] J. Biemond, R. L. Lagendijk, and R. M. Mersereau, "Iterative methods for image deblurring", Proc. IEEE, 78(5):856–883, 1990.

[2] Al Bovik, *The essential guide to the image processing*, Academic Press, 2009.

[3] Rafael C. Gonzalez, Richard E. Woods, *Digital Image Processing*, 2nd Edition, Prentice-Hall, 2002.

[4] Rafael C. Gonzalez, Richard E. Woods, Steven L. Eddins, *Digital Image Processing Using MATLAB*, Prentice-Hall, 2003.

[5] Domingo Mery, Dieter Filbert, "A Fast non-Iterative Algorithm for the Removal of Blur Caused by Uniform Linear motion in X-Ray Images", 15th World Conference on Non-Destructive Testing, Computer Processing and Simulation , Rome, 15-21 October 2000.

[6] Spiros Chountasis, Vasilios N. Katsikis, and Dimitrios Pappas, "Applications of the Moore-Penrose Inverse in Digital Image Restoration", Mathematical Problems in Engineering Volume 2009 (2009).

[7] Ahmet M. Eskicioglu and Paul S. Fisher, "Image Quality Measures and Their Performance", IEEE Transactions on Communications, vol. 43, pp. 2959-2965, Dec. 1995.

[8] Zhou Wang and Alan C. Bovik, "Mean Squared Error: Love It or Leave It? A New Look at Signal Fidelity Measures", IEEE Signal Processing Magazine, vol. 26, no. 1, pp. 98-117, Jan. 2009.

Design of Gyrator Active Filters with Equalized Maxima of the Amplifier Output Voltages

Ivan S. Uzunov[1]

***Abstract* – New simpler method for equalizing the output voltage maxima of the amplifiers in gyrator active filter is proposed. The method is based on a proper variation of the gyrator transconductances. First is outlined the general approach and then it is applied consistently for deriving a set of equivalent circuits for replacing of each impedance of the LC ladder prototype by its gyrator counterparts.**

***Keywords* – active filters, gyrators, dynamic range, OTA**

I. INTRODUCTION

The dynamic range optimization of the active filters is a long-time studied problem [1], [2], maintained by the continuous demands for reducing the supply voltages in the integrated circuit (IC) and by the increased frequency range of operation. The dynamic range is two-sided problem: equalizing the maxima of the amplifier output voltages and minimizing the filter output noise. The first task is treated for 40 years (see e.g. [3], [4]) and there are satisfactory solutions for the most classes of the active filters. Exclusions are the filters, derived from passive ladder LC filters by substituting of the inductors with gyrator active circuits. The main difficulty is the preserving the structure of the passive filter, which keeps the nodal AC voltages in the gyrator filters the same as they are in the passive one. The problem was long time mentioned as unsolved [1] and a solution was proposed recently [4] as a consequence of more general method for optimizing the dynamic range of OTA-C filters.

This paper proposes an alternative approach, which is simpler and requires fewer calculations. It is based on a model of a floating inductor replaced by two gyrators and a capacitor. The necessary data for the equalization are received by single analysis of the initial LC ladder and then the parts of the LC ladder are directly replaced by their active counterparts, which elements can be calculated very easy.

II. THE PROPOSED METHOD

An admittance Y connected across between two gyrators in cascade with arbitrary transconductances is equivalent to longitudinal impedance Z in cascade with an artificial element called scalor (Fig. 1) [5], [6], which two port equations are $V_1 = K_V V_2;\ I_1 = K_I(-I_2)$. The relationships between the elements in Fig. 1 are

$$Z = Y/(g_{f1}g_{b1});\ K_V = g_{b2}/g_{f1};\ K_I = g_{b1}/g_{f2}\,. \qquad (1)$$

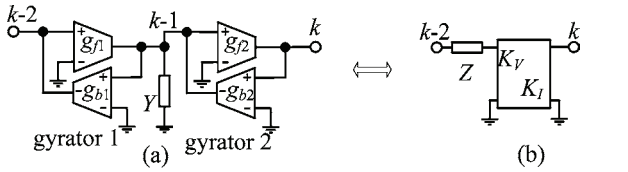


Fig. 1. An admittance between two gyrators (a) and its model (b).

The scalor can be moved freely when it is in cascade with other elements. When it is moved from output to the input of the circuit, it multiplies all "over-jumped" impedances by K_I/K_V; all over-jumped voltages by $1/K_V$ and all over-jumped currents by $1/K_I$. When the scalor motion is in opposite direction, the multiplying coefficients are reciprocal. This property keeps unchanged the relationship between input and output voltages of the whole circuit.

The voltages and currents in the LC ladder can be determined by AC analysis. First step in the design of the gyrator filter is to insert two scalors with reciprocal parameters before the last node (Fig. 2(b)). The scalors do not change the circuit due to the reciprocity of their parameters. Then the first scalor is moved leftward and placed before node 2 (Fig. 2(c)).

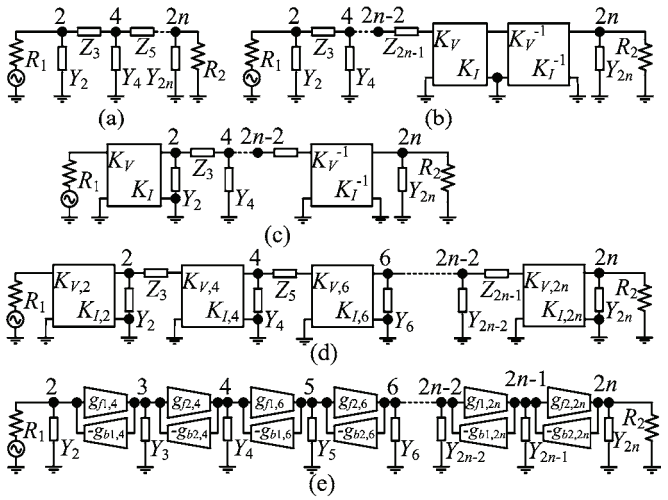


Fig. 2. (a) The LC ladder prototype. (b) Placing two scalors before the last node. (c) Moving the first scalor before the node 2. (d) The circuit after adding all scalors. (e) Replacement the series impedances and associated scalors by gyrator pairs and cross admittances.

The next steps are similar: New scalor pairs with mutually reciprocal parameters are placed before the second scalor in Fig 2(c) and the first scalors from the pairs are moved leftward and placed before nodes 4, 6, etc. At the end there will be a scalor before every node of the ladder (Fig 2(d)). All

[1]Ivan S. Uzunov is with the Faculty of Telecommunications, Technical University of Sofia, 8 Kl. Ohridski blvd., 1000 Sofia, Bulgaria, e-mail: iuzunov@tu-sofia.bg.

second scalors from the pairs, connected in cascade before node 2*n*, are equivalent to a scalor with parameters $K_{V,2n} = 1/(K_{V,2}K_{V,4}...K_{V,2n-2})$, $K_{I,2n} = 1/(K_{I,2}K_{I,4}...K_{I,2n-2})$.

The scalor moving multiplies all over-jumped voltages by $1/K_V$, which allows to use the K_V parameters for equalizing the maxima of the voltages in the nodes 2, 4, 6, etc. to the maximum of the output voltage. The node 2 is over-jumped only by the scalor $[K_{V,2}\ K_{I,2}]$ and the maxima of V_2 and V_{2n} will be equal if $K_{V,2} = \max|V_2|/\max|V_{2n}|$. The other ladder nodes are over-jumped by more scalors, intended for equalizing the voltage maxima of the previous nodes. The last scalor, passed through node *k* before adding the scalor associated with this node, is the scalor, making $\max|V_{k-2}| = \max|V_{2n}|$. Thus, the equalizing of the maxima at the nodes before *k* changes V_k to $V_k\max|V_{2n}|/\max|V_{k-2}|$. Then K_V of the scalor associated with node *k* must be:

$$K_{V,k} = \max|V_k|/\max|V_{k-2}| \qquad (2)$$

After equalizing the voltage maxima in all ladder nodes, every longitudinal impedance has an associated scalor (Fig. 2(d)). They can be replaced by gyrator pairs and cross impedances according Fig. 1 as it is shown in Fig. 2(e). This step creates new nodes (with odd numbers in Fig. 2(e)), which voltage maxima must be equalized too.

The voltage at node *k*-1 in Fig. 1(a) is equal to $(g_{f1}V_{k-2} - g_{b2}V_k)/Y$ and it can be transformed in the following way, if take into account formulas (2):

$$V_{k-1} = \left(g_{f1}V_{k-2} - g_{b2}V_k\right)/\left(g_{f1}g_{b1}Z\right) = \left(V_{k-2} - K_VV_k\right)/\left(g_{b1}Z\right). \quad (3)$$

The voltage K_VV_k is the voltage V_k' at node *k* before inserting the scalor associated with the impedance *Z*, when the nodes *k*-2 and *k* are connected by the impedance *Z* only. Thus the voltage V_{k-1} is equal to I_Z'/g_{b1}, where I_Z' is the current through *Z*. This current is affected by all scalors passed through it for equalizing the voltage maxima at the nodes before *k*-2 and it is equal to the current I_Z through the impedance *Z* in the LC prototype, divided by the K_I's of all scalors, passed through node *k*-2. The requirement for equal maxima of the voltages V_{k-1} and V_{2n} gives the following expression for $g_{b1,k}$

$$g_{b1,k} = \max|I_Z|/\left(K_{I,2}K_{I,4}...K_{I,k-2}\max|V_{2n}|\right). \qquad (4)$$

The first scalor $[K_{V,2}\ K_{I,2}]$ can be processed in a simpler way. It can be moved leftward again. Then it jumps over the input resistance R_1, multiplying its value by $K_{I,2}/K_{V,2}$, and over the signal source and disappears. The multiplication of the source voltage by $1/K_V$ can be compensated by additional amplifier if necessary.

Not all inductors in the circuit can be replaced by gyrators and capacitors during the above steps since: 1) The LC filter could has inductors in the cross branches. 2) Fig. 3 shows two cases of longitudinal LC branches in the prototype, in which the replacement the inductors only is not desirable. In fact, if a longitudinal branch in the ladder is of degree 2 or higher, it must be replaced at once by gyrators according Fig. 1, which creates new inductors. Even single longitudinal capacitors (in high-pass filters) create cross inductors when the voltage maxima in the ladder nodes are equalized.

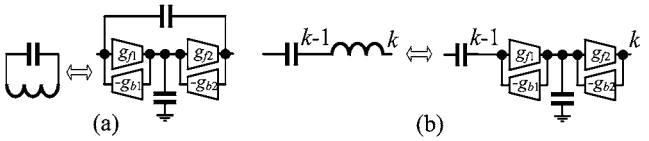


Fig. 3. Two transformations, which can not be applied: (a) The gyrator circuit is not equivalent to the tank if g_m's differ [7], [8]; (b) The voltage maximum in node *k*-1 can not be equalized.

A single grounded inductor is transformed by the classical way (Fig.4(a)) and the condition for equal voltage maximum in the new node *m* to the other voltage maxima is $g_b = \max|I_L|/\max|V_{2n}|$. A new scalor is inserted in a grounded series LC tank (Fig. 4(b)) for equalizing the voltage maxima in the intermediate nodes. Then the inductor and the scalor are replaced by gyrator pair and a capacitor C_L and the following conditions hold for having the voltage maxima in the nodes m_1 and m_2 equal to the others:

$$K_V = \max|V_m|/\max|V_{2n}|; \quad g_{b1} = \max|I_{LC}|/\max|V_{2n}| \qquad (5)$$

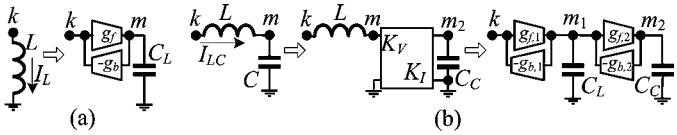


Fig. 4. Replacement of (a) single grounded inductor; (b) grounded series LC tank.

Grounded inductors and series LC tanks can appear in the structure in Fig. 2(e) in two ways: they could exist in the cross branches in the LC prototype (then *k* is even); or they can be a result from transformation of floating capacitors or parallel LC tanks according Fig. 1. The transformation in the second case converts impedances in admittances and vice verse, voltages into currents and vice verse as well. This is illustrated in Fig. 5. The relationships between element, voltage and current values in these equivalent circuits can be derived easily from the basic gyrator properties

$$\begin{aligned} &I_L = g_{f1}V_C;\ I_{LC} = g_{f1}V_{LC};\ V_m = I_L/g_{b1};\\ &L_C = C/\left(g_{f1}g_{b1}\right);\ C_L = LC/L_C \end{aligned} \qquad (6)$$

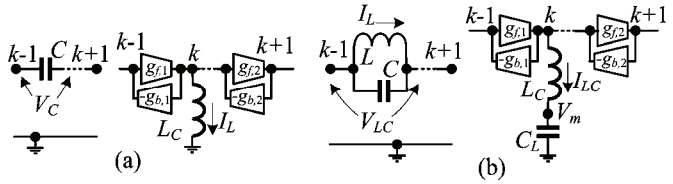


Fig. 5. Transforming of longitudinal elements: (a) capacitor; (b) parallel LC tank.

Of course, the element values and the voltages and currents in the right-hand sides of equations (5) and (6) are affected also by the scalors passed through these elements.

III. Catalogue for Replacement of the Basic Branches in LC Ladder Circuits

The method outlined above allows to create a catalogue containing the basic branches in an LC ladder circuit, their gyrator counterparts and the corresponding design equations

ensuring equal voltage maxima at all amplifier outputs. Fig. 6 gives the basic ladder blocks and their gyrator equivalents.

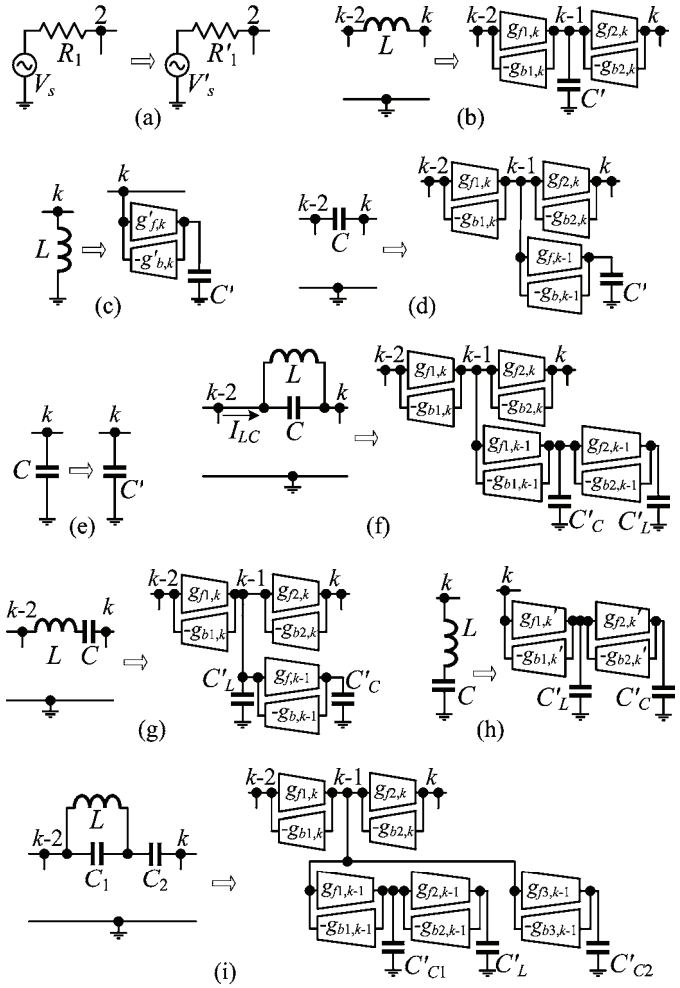


Fig. 6. Basic fragments of LC ladder filters and their gyrator counterparts: (a) signal source and its internal resistance; (b) longitudinal inductor; (c) cross inductor; (d) longitudinal capacitor; (e) cross capacitor; (f) longitudinal parallel LC tank; (g) longitudinal series LC tank; (h) cross series LC tank (i) third order longitudinal LC circuit.

The design equations for the circuits in Fig. 6 are:

Fig. 6(a):

$$R_1' = R_1 \frac{\max|V_o|}{\max|V_2|} K_{I,2};\ V_s' = V_s \frac{\max|V_o|}{\max|V_2|}$$

Fig. 6(b): C' must be chosen arbitrarily and

$$g_{b1,k} = \frac{\max|I_L|}{K_{I,2}K_{I,4}...K_{I,k-2}\max|V_o|};\ g_{f1,k} = \frac{C'}{L}\frac{\max|V_{k-2}|}{\max|I_L|};$$

$$g_{f2,k} = \frac{g_{b1,k}}{K_{I,k}};\ \ g_{b2,k} = g_{f1,k}\frac{\max|V_k|}{\max|V_{k-2}|}$$

Fig. 6(c): C' must be chosen arbitrarily and

$$g'_{b,k} = \frac{\max|I_L|}{K_{I,2}K_{I,4}...K_{I,k}\max|V_o|};\ g'_{f,k} = \frac{C'}{L}\frac{\max|V_k|}{\max|I_L|};$$

Fig. 6(d): C' and $g_{f1,k}$ must be chosen arbitrarily and

$$g_{b1,k} = \frac{\max|I_C|}{K_{I,2}K_{I,4}...K_{I,k-2}\max|V_o|};$$

$$g_{f2,k} = \frac{g_{b1,k}}{K_{I,k}};\ \ g_{b2,k} = g_{f1,k}\frac{\max|V_k|}{\max|V_{k-2}|};$$

$$g_{f,k-1} = \frac{C'}{C}\frac{\max|I_C|}{\max|V_C|};\ g_{b,k-1} = g_{f1,k}\frac{\max|V_C|}{\max|V_{k-2}|}$$

Fig. 6(e):

$$C' = \frac{C}{K_{I,2}K_{I,4}...K_{I,k}}\frac{\max|V_k|}{\max|V_o|}$$

Fig. 6(f): C'_C and C'_L must be chosen arbitrarily and

$$g_{b1,k} = \frac{\max|I_{LC}|}{K_{I,2}K_{I,4}...K_{I,k-2}\max|V_o|};\ g_{f2,k} = \frac{g_{b1,k}}{K_{I,k}};$$

$$g_{f1,k} = \frac{C'_L K_{I,k-1}}{L}\frac{\max|V_{k-2}|}{\max|I_L|};\ g_{b2,k} = g_{f1,k}\frac{\max|V_k|}{\max|V_{k-2}|};$$

$$g_{f1,k-1} = \frac{C'_C}{C}\frac{\max|I_{LC}|}{\max|V_{LC}|};\ g_{b1,k-1} = \frac{C'_L K_{I,k-1}}{L}\frac{\max|V_{LC}|}{\max|I_L|};$$

$$g_{f2,k-1} = \frac{g_{b1,k-1}}{K_{I,k-1}};\ g_{b2,k-1} = \frac{C'_C}{C}\frac{\max|I_L|}{\max|V_{LC}|};$$

Fig. 6(g): C'_C and C'_L must be chosen arbitrarily and

$$g_{b1,k} = \frac{\max|I_L|}{K_{I,2}K_{I,4}...K_{I,k-2}\max|V_o|};\ g_{f1,k} = \frac{C'_L}{L}\frac{\max|V_{k-2}|}{\max|I_L|};$$

$$g_{f2,k} = \frac{g_{b1,k}}{K_{I,k}};\ \ g_{b2,k} = g_{f1,k}\frac{\max|V_k|}{\max|V_{k-2}|}$$

$$g_{b,k-1} = g_{f1,k}\frac{\max|V_C|}{\max|V_{k-2}|};\ g_{f,k-1} = \frac{C'_L C'_C}{g_{b,k-1}LC}$$

Fig. 6(h): C'_C and C'_L must be chosen arbitrarily and

$$g'_{b1,k} = \frac{\max|I_L|}{K_{I,2}K_{I,4}...K_{I,k}\max|V_o|};\ g'_{f1,k} = \frac{C'_L}{L}\frac{\max|V_k|}{\max|I_L|};$$

$$K'_{I,k} = \frac{1}{K_{I,2}K_{I,4}...K_{I,k}}\frac{C}{C'_C}\frac{\max|V_C|}{\max|V_o|};$$

$$g'_{f2,k} = \frac{g'_{b1,k}}{K'_{I,k}};\ \ g'_{b2,k} = g'_{f1,k}\frac{\max|V_C|}{\max|V_k|}$$

Fig. 6(i): C'_{C1}, C'_{C2}, $g_{f1,k}$ and $K_{I,k-1}$ must be chosen and

$$g_{b1,k} = \frac{\max|I_{C2}|}{K_{I,2}K_{I,4}...K_{I,k-2}\max|V_o|};\ g_{f2,k} = \frac{g_{b1,k}}{K_{I,k}};$$

$$g_{b2,k} = g_{f1,k}\frac{\max|V_k|}{\max|V_{k-2}|};\ C'_L = \frac{g_{f1,k}L}{K_{I,k-1}}\frac{\max|I_L|}{\max|V_{k-2}|};$$

$$g_{f1,k-1} = g_{f1,k}\frac{\max|V_L|}{\max|V_{k-2}|};\ g_{f2,k-1} = \frac{g_{b1,k-1}}{K_{I,k-1}};$$

$$g_{f1,k-1} = \frac{C'_{C1}}{C_1}\frac{\max|I_{C2}|}{\max|V_L|};\ g_{b2,k-1} = g_{f1,k-1}\frac{\max|I_L|}{\max|I_{C2}|};$$

$$g_{f3,k-1} = \frac{C'_{C2}}{C_2}\frac{\max|I_{C2}|}{\max|V_{C2}|};\ g_{b3,k-1} = g_{f1,k}\frac{\max|V_{C2}|}{\max|V_{k-2}|}$$

In these equations V_o is the output voltage. The voltages and currents in the ladder elements are marked by V_Z and I_Z, where Z is L, C or LC depending on the case. V_k and V_{k-2} are

the ladder nodal voltages. All voltage and current maxima in the formulas are from the LC prototype.

The catalogue and the corresponding formulas can be applied directly and the whole procedure with introducing and moving the scalors is hidden. From this point of view the K_I parameters, which are chosen arbitrarily, can be thought like some design constants.

In all cases is assumed that K_I parameters are freely chosen. This is not valid for the last longitudinal branch in the LC ladder. Its K_I (let say $K_{I,2n}$) is reciprocal to the product of all previous K_I's according the theory in section II:

$$K_{I,2n} = 1 \Big/ \prod_{k=1}^{n-1} K_{I,2k} \ . \qquad (7)$$

IV. EXAMPLE

The proposed method will be illustrated by a low-pass 5th order elliptic filter with normalized passband and stopband bounds of 1Hz and 1.2Hz respectively, passband ripples of 0.5dB and minimum stopband attenuation of 35.5dB. The circuit of the LC prototype is shown in Fig 7(a) and its normalized element values are: $R_1 = R_2 = 1\Omega$; $C_1 = 1.44$F; $L_2 = 0.975$H; $C_2 = 0.346$F; $C_3 = 1.77$F; $L_4 = 0.613$H; $C_4 = 1.07$F; $C_5 = 1.03$F. The AC analysis of the prototype is done by PSpice at 2V magnitude of the input source and it returns the following maxima of the voltages and currents of interest: max$|V_6| = 1$V; max$|V_2| = 1.724$V; max$|V_4| = 1.866$V; max$|V_{LC2}| = 3.342$V; max$|V_{LC4}| = 2.419$V; max$|I_{LC2}| = 2.175$A; max$|I_{LC4}| = 1.418$A; max$|I_{L2}| = 3.35$A and max$|I_{L4}| = 3.949$A.

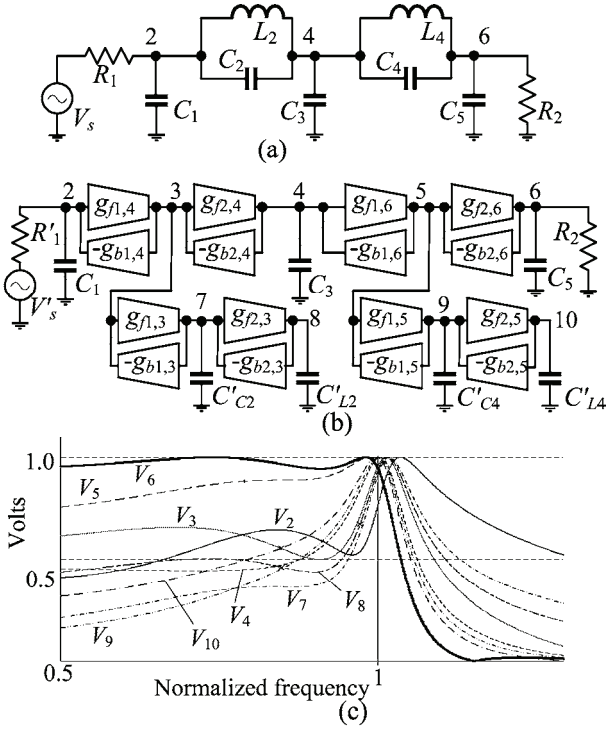


Fig. 7. (a) LC prototype for the example; (b) the gyrator filter; (c) amplifier output voltages in the gyrator circuit (PSpice simulation).

The LC prototype includes blocks, which gyrator counterparts are given in Fig. 6(a), 6(e) and 6(f). The gyrator circuit is given in Fig 7(b) and its elements are calculated by using the corresponding formulas in Section III. The signal source and the input terminating resistance are changed to $V'_s = 1.16$V, $R'_1 = 0.58\Omega$. The new values of the capacitors C_1 and C_3 are $C_1 = 2.483$F and $C_3 = 3.303$F. The capacitor C_5 and the output terminating resistance are unchanged. The elements of circuit, replacing the tank L_2C_2, have values: $g_{f1,4} = 2.174$S, $g_{b1,4} = 2.175$S, $g_{f2,4} = 2.175$S, $g_{b2,4} = 2.354$S, $g_{f1,3} = 2.739$S, $g_{b1,3} = 4.215$S, $g_{f2,3} = 2.737$S, $g_{f2,3} = 2.737$S. The following parameters are chosen for their calculation: $C'_{C2} = 1.456$F, $C'_{L2} = 2.675$F, $K_{I,2} = K_{I,4} = 1$, $K_{I,3} = 1.54$. The circuit, replacing the tank L_4C_4, is calculated in similar way: $K_{I,6} = 1$, $g_{f1,6} = 1.419$S, $g_{b1,6} = 1.418$S, $g_{f2,6} = 1.418$S, $g_{b2,6} = 0.76$S, $g_{f1,5} = 0.66$S, $g_{b1,5} = 1.839$S, $g_{f2,5} = 0.66$S, $g_{b2,5} = 1.837$S, with the following choice of the other values and parameters $K_{I,5} = 2.785$, $C'_{C4} = 1.204$F, $C'_{L4} = 0.661$F. A PSpice simulation confirms of the equality of the maxima of all amplifier output voltages (Fig 7(c)).

V. CONCLUSION

A modification of the method for design of active filters based on direct element substitution in LC prototype with gyrator circuits is proposed. The basic advantage of the new method is its simplicity. The active filter is received by block-by-block replacement in the LC prototype with corresponding active circuits and applying of simple formulas for calculation of the elements. The formulas are modified in such a way, which ensures equal maxima of all amplifier output voltages. This is the first step for complete optimization of the filter dynamic range. Some of the elements and the parameters are chosen freely in the design procedure and they can be used in the next minimization of the output noise.

ACKNOWLEDGEMENT

This work is supported by TU- Sofia – Grant 102ni065-07.

REFERENCES

[1] T. Deliyannis, Y. Sun, J. K. Fidler, *Continuous-Time Active Filter Design*, Boca Raton, CRC-Press, 1999.

[2] R. Schaumann, M. E. Van Valkenburg, *Design of Analog Filters*, Oxford University Press, 2001.

[3] S. Halfin, An Optimizatin Method for Cascaded Filters, *The Bell System Techn. Journal*, 49, No. 2 (February 1970), pp. 185-190.

[4] Y. Palaskas, Y. Tsividis, Dynamic Range Optimization of Weakly Nonlinear, Fuly Balanced, Gm-C Filters with Power Dissipation Constraints, *IEEE Trans. on Circuits and Systems – II: Analog and Digital Signal Processing*, vol. 50, No 10, Oct. 2003, pp. 714-727.

[5] S. I. Koutzarov, I. S. Uzunov, Equivalent Circuit for an Ungrounded Gyrator if Realized by Two Real Gyrators, *AEÜ*, vol. 26, pp. 408-409, September 1972.

[6] I. S. Uzunov, Theoretical Model of Ungrounded Inductance Realized With Two Gyrators, *IEEE Trans. on Circuits and Systems – II: Express Briefs*, vol. 55, No. 10, Oct. 2008, pp. 981-985.

Comparative Analysis of Adaptive Color KLT and YCrCb for Representation of Color Images

Peter N. Ivanov[1] and Roumen K. Kountchev[2]

Abstract – **In this paper are represented the results of the experimental research on a new algorithm for Adaptive Color KLT and the comparison with the YCrCb color system for the representation of color images. The obtained results include several groups of color test images. The results of the made experiments are represented both in graphical and table form. They are showing the advantages of the new algorithm for Adaptive Color KLT in comparison with the YCrCb color system.**

Keywords – **color systems, adaptive color KLT, YCrCb color system, format for color images representation.**

I. INTRODUCTION

The transformation of the primary color space RGB is very important part of the image processing. In general the components of the RGB color system are correlated, the most important feature of the Adaptive Color KLT is that it permits the decorrelation of the components. There are many applications of the new algorithm– in the domain of image processing, image coding, color image recognition and many others.

There are two types of color spaces – deterministic and statistical. The deterministic transformations such as YCrCb, YUV, YIQ, CMYK [2, 3 ,6 ,7 ,8 ,9, 11, 12] are calculated using fixed coefficients and require less computations but the disadvantage of those color systems is that they are not adapted for each individual image that is being transformed. In the other type – the statistical transforms such as the Adaptive Color KLT [1] the generated color space is adapted to the statistical properties of each image or group of images that is being transformed but the disadvantage is that it requires more computations than the deterministic color systems. That gives better quality of the restored image, less correlation of the components, etc [4, 5].

The purpose of this paper is to introduce a new format for color image representation based on the color KLT (ACT) and the experimental results obtained by comparing it with the already existing format YCrCb on which are based many image formats like JPEG and JPEG2000 [3, 11]. The goal is to achieve better quality of the restored images based on the color transforms only. For the experiments we used different sets of images of various size and contents to prove that the new format works with all kind of images.

This paper is organized in the following manner – in part two we give the detailed algorithm description of the new format for representation of color images, in part three we give the experimental results and the conclusion is given in part four.

II. ALGORITHM DESCRIPTION AND ANALYSIS

The proposed algorithm is a complete analytical solution to the problem of the color transform based on the KLT. It is based of the method presented in [1]. The algorithm is simplified so that to reduce the necessary computations of the color transform.

Transforming an RGB image into the new color format is made by the following steps following the proposed algorithm presented in Fig 2.1, blocks (1)-(9), which is the forward algorithm for the Adaptive Color KLT:

Step 1: Calculation of the primary color vectors $\overset{\rho}{C}_s$ for each pixel from the original RGB image, where s is the current pixel and S the total number of the pixels in the image, therefore $S = M \times N$, where M and N are the image height and width.

Step 2: Calculation of mean values of the colors R,G and B- Fig 2.1, block (2). The mean values are necessary for the computation of the covariance matrix in the next step.

Step 3: Calculation of the image covariance matrix:

$$[K_C] = \left[\frac{1}{S}\sum_{s=1}^{S} \overset{\rho}{C}_s \overset{\rho}{C}_s^t\right] - \overset{\rho}{m}_c \overset{\rho}{m}_c^t = \begin{bmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{23} \\ k_{31} & k_{32} & k_{33} \end{bmatrix}$$

Where the coefficients $k_{i,j}$ are calculated using Fig 2.1, block (3). The covariance matrix is a diagonal matrix so therefore the eigenvalues are always real numbers.

Step 4: Calculating the coefficients of the characteristic equation of the covariance matrix

$$\det|k_{ij} - \lambda\delta_{ij}| = \lambda^3 + a\lambda^2 + b\lambda + c = 0$$

using equations Fig 2.1, blocks (4) and (5).

Step 5: Calculation of the eigenvalues of the characteristic equation defined in the previous step. Given that the covariance matrix $[K_C]$ is a diagonal matrix the eigenvalues can be defined by the "Cardano" relations or the so called trigonometric equations [10] Fig2.1, block (6). Where we have the condition

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0 .$$

[1]Peter N. Ivanov is with the Faculty of Telecommunications, Technical University, Kliment Ohridski 8, 1000 Sofia, Bulgaria, e-mail peter.n.ivanov@gmail.com

[2]Roumen K. Kountchev is with the Faculty of Telecommunications, Technical University, Kliment Ohridski 8, 1000 Sofia, Bulgaria, e-mail: rkountch@tu-sofia.bg

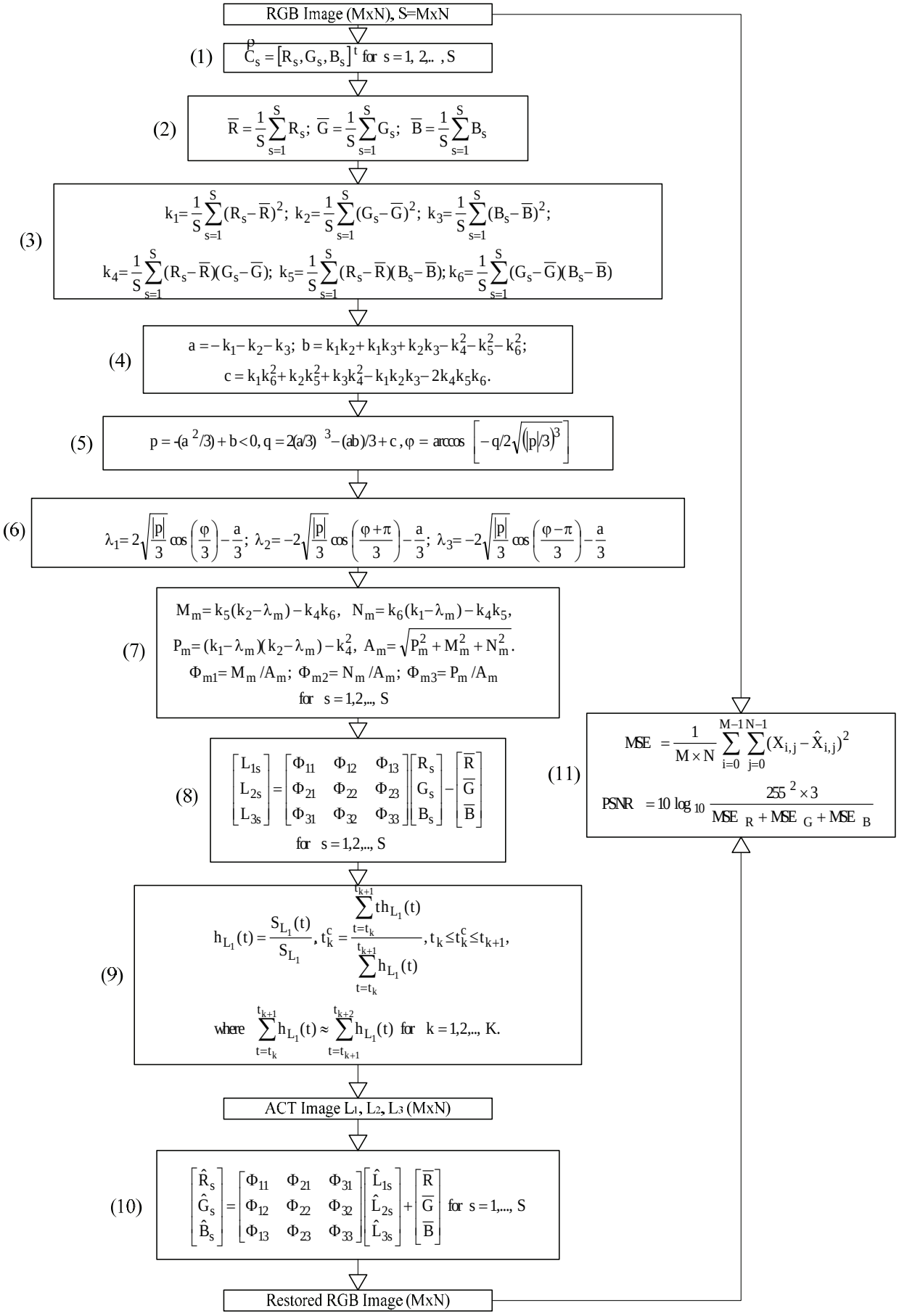


Fig 2.1 Block diagram of the algorithm for Adaptive Color KLT and the PSNR calculation

Step 6: Calculation of the eigenvectors of the covariance matrix $[K_C]$: Fig. 2.1, block (7). Here A_m, P_m, B_m, D_m are the coefficients used for the computations and $m = 1,2,3$. From the eigenvectors we form the transformation matrix $[\Phi]$:

$$[\Phi] = \begin{bmatrix} \rho\Phi_1^t \\ \rho\Phi_2^t \\ \rho\Phi_3^t \end{bmatrix} = \begin{bmatrix} \Phi_{11} & \Phi_{12} & \Phi_{13} \\ \Phi_{21} & \Phi_{22} & \Phi_{23} \\ \Phi_{31} & \Phi_{32} & \Phi_{33} \end{bmatrix}$$

Step 7: Performing the color transform using the already generated transformation matrix $[\Phi]$ to obtain the transformed color vectors $\rho L_s = [L_{1s}, L_{2s}, L_{3s}]^t$ using the equation Fig 2.1, block (8). Where again s is the current pixel that is being transformed and S is the total number of the pixels in the image.

Step 8: Perform an adaptive quantization of the obtained matrix $[L_1]$ to comply with the limits of 8 bits per pixel or 256 unique values in the matrix using the equations from Fig. 2.1, block (9). Here $h_{L_1}(t)$ is the histogram calculated for the first component of the Adaptive Color KLT, t_k^c is the center of gravity of the part of the histogram $h_{L_1}(t)$ between levels t_k and t_{k+1} of the component L_1 (k=1,2,..,K for K - number of quantization levels). After adaptive quantization we have three matrices $[\hat{L}_1], [\hat{L}_2], [\hat{L}_3]$ that comply with the limit of 8 bits per pixel or 24 bpp for each pixel in the image.

The restoration of an image already converted in the Adaptive Color KLT back into the primary format RGB is made with only one step using the equation from Fig. 2.1, block (10).

For the restoration of the original image in the primary color system RGB is needed the matrix $[\Phi]$, the quantization table and the mean values of the colors to be passed to the decoder. This information must be embedded to the file that contains the transformed color vectors. The new format explained in this part of the paper shows great results in terms of quality of the restored images. More detailed results are presented in the next section.

III. Experimental Results

As experimental dataset was used three different sets of images of different size – the "Kodak" image set plus the image "Lena" and "Barbara" 26 images in total, of size 512×768 or 768×512, Lena - 512×512 and Barbara - 640×512 pixels. The "cgraph" image set which comprises ten computer generated images of size 1024×768. The "natural" image set which is comprises 10 images of HD size, 1920×1024 pixels. All the images are in the format .bmp (primary format for the RGB color system) with 24 bits describing each pixel (bpp).

As an algorithm for YCrCb conversion was used the fix coefficients from the matrix equation [3]:

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.2989 & 0.5866 & 0.1145 \\ -0.1688 & -0.3312 & 0.5000 \\ 0.5000 & -0.4184 & 0.0816 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

For quality measurement was used the Peak Signal to Noise Ratio (PSNR) equation given on Fig 2.1, block (11). The PSNR gives the objective representation of the restored image quality [3].

$$MSE = \frac{1}{M \times N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (X_{i,j} - \hat{X}_{i,j})^2$$

Where the MSE is the Mean Square Error of one component of the RGB image. The value $X_{i,j}$ is the original value from the original image and $\hat{X}_{i,j}$ is the restored value (from the restored image) in the point (i, j) of the image for the given component R, G or B. Therefore there are three different MSEs which represent the error in the image. So to find the total error we must add them $(MSE_R + MSE_G + MSE_B)$ [3].

$$PSNR = 10 \log_{10} \frac{255^2 \times 3}{MSE_R + MSE_G + MSE_B}$$

a) b)

c) d)

Fig. 3.1

In Fig. 3.1 are shown some images from the test image sets: a) cgraph7 (1024×768, 24 bpp); b) natural4 (1920×1024, 24bpp); c) kodim05 (768×512, 24 bpp); d) kodim22 (768×512, 24 bpp).

Image Set	ACT PSNR [dB]	YCrCb PSNR [dB]
Kodak + Lena +Barbara	48.44	46.57
Cgraph	48.97	46.12
Natural	48.59	46.01

Table 3.1. Mean PSNR values for all the image sets

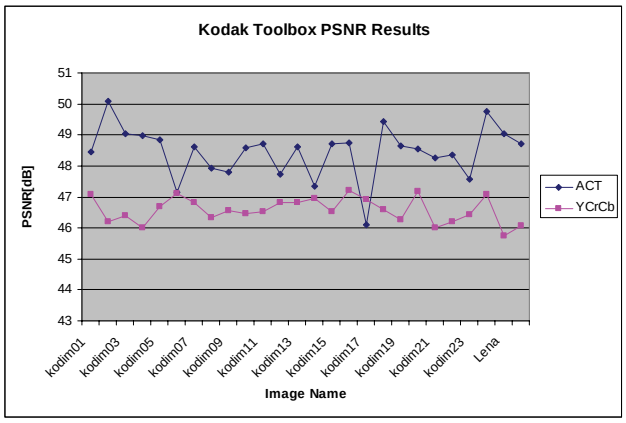


Fig 3.2 Kodak image set PSNR

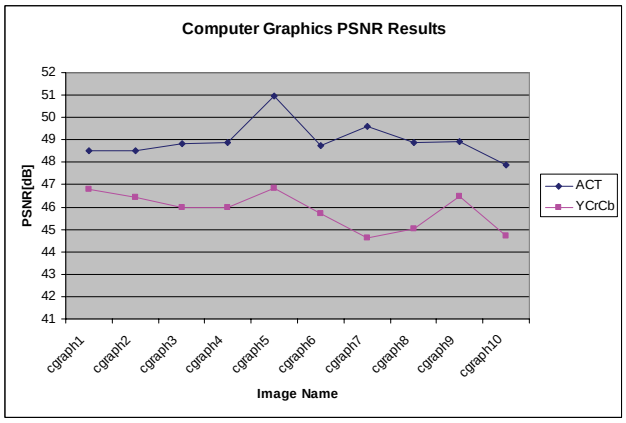


Fig 3.3 "Cgraph" image set PSNR

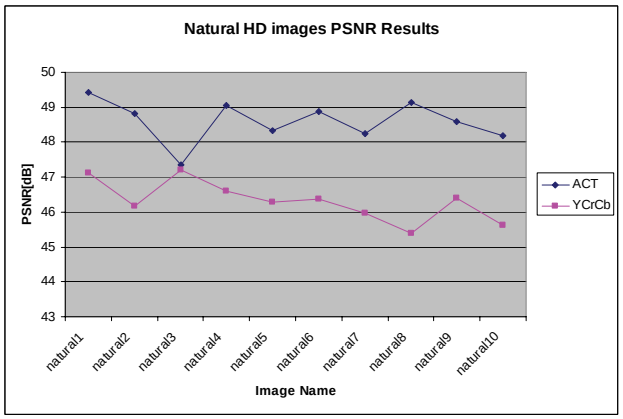


Fig 3.4 "Natural" image set PSNR.

From Table 3.1 we see that the average difference between the proposed format for representation of color images and the already existing YCrCb in no less than 2 [dB] for image set. This is something that can give many advantages in the domains of image processing and/or image coding.

From the charts Fig. 3.2, 3.3 and 3.4 we can see that the quality of the restored images is very high – plus 48 [dB] – without any visual loss in the image. The losses in the proposed algorithm result from the "casts" of the data types in the program. There are no losses from compression or any other kind of lossy actions on the image data.

IV. Conclusion

There are two types of color transforms – deterministic (RGB, YCrCb, YUV, YIQ, CMYK) [2, 3, 6, 7, 8, 9] and statistical (ACT – Adaptive Color KLT Transform) [4, 5]. The deterministic transforms are defined by fixed equations with fixed coefficients which are not changing for each image. The statistical transforms (ACT) the components are adaptive for each image that is being transformed. Therefore the transform matrix generated by the algorithm is adapted to the statistical information of the image that is being transformed. The algorithm gives very high quality of the restored image as we have seen in the previous part of the paper. The high quality of the output images can be carried with other algorithms such as image coding and compression. As a result of that we can say that the proposed algorithm has many applications – in the domain of image coding, compression, processing, etc.

Acknowledgement

This paper was supported by the National Fund for Scientific Research of the Bulgarian Ministry of Education and Science (Contract - BY-И-305/2007).

References

[1] R. Kountchev, R. Kountcheva. New Method for Adaptive Karhunen-Loeve Color Transform. Proc. of 9th Intern. Conf. on Telecommunications in Modern Satellite, Cable and Broadcasting services (TELSIKS'09), October 7-9, 2009, Nish, Serbia, pp. 209-216.

[2] D. Carevic, T. Caelli. Region based coding of color images using K-L transform. Graphical Models and Image Processing. Vol. 59, No.1, 1997, pp. 27-38.

[3] W. Pratt. Digital Image Processing. Wiley Interscience, New York, 2007.

[4] M. Fleury, A. C. Downton and A. F. Clark. Karhunen – Loeve Transform – Image Processing, University of Essex, Wivenhoe Park, UK 1997.

[5] R. Dony. Karhunen-Loève Transform. Book Chapter in The Transform and Data Compression Handbook, Ed. K. Rao and P. Yip, Boca Raton, CRC Press LLC, 2001.

[6] International Color Consortium - http://www.color.org.

[7] G. Wyszecki, W. Stiles. Color Science. John Wiley & Sons, New York, 1982.

[8] P. Hao, Q. Shi, Reversible Integer KLT for progressive-to-lossless compression of multiple component images, ICIP, Vol. 1, 2003, pp. 633-636.

[9] S. Sangwine, R. Horne, The colour image processing handbook, Chapman & Hall, 1998.

[10] G. Korn, T. Korn. Mathematical Handbook for Scientists and Engineers, Mc Graw-Hill Book Company, NY, 1968.

[11] R. Gonzalez, R. Wood. Digital Image Processing, Prentice Hall, 2001.

[12] M. Fairchild, Color and Image Appearance Models. John Wiley & Sons, 2005.

Comparison of the Structures of the Inverse Difference and Laplacian Pyramids for Image Decomposition

R. Kountchev[1], R. Kountcheva[2]

***Abstract* – In this work are compared and evaluated the computational structures of the Inverse Difference Pyramid, IDP (already presented in earlier publications of the authors) and the famous Laplacian Pyramid, LP for discrete images decomposition. On the basis of the comparison of the substitution graphs which represent the recursive calculation of the 3-layer decompositions and of the evaluation of their structures complexity, are outlined the basic IDP advantages for pipeline image processing. The results obtained could be used for design of IDP coders for image compression, aimed at real-time applications.**

***Keywords* – Pyramidal decompositions, Laplacian Pyramid, Inverse Difference Pyramid**

I. INTRODUCTION

The contemporary computer world involves the management and processing of huge amounts of visual information: still images, video, and multimedia. The efficient storage and compression of this information requires the use of various techniques for image representation. The primary form for digital image presentation, which is not compressed, is the *matrix* [1]. This approach ensures unchanged image quality, but, its storage requires significant resources, and the processing – high computational power. The secondary forms for image representation are obtained from the primary one and could be pyramids, multi-dimensional vectors, orthogonal transforms, tree structures, algebraic models, models for visual information perception, etc. The main attention in this paper will be given to the pyramidal representations and their efficiency.

In general, the *pyramidal* representation [1-4] describes the image with progressively increased resolution, which corresponds to the layers of the Gaussian-Laplacian Pyramid. The derivatives of this representation are the Reduced Sum/Difference pyramid; the S-transform pyramid, the Hierarchy-Embedded Differential Pyramid; the Least Square Pyramid, the Morphological Pyramid, etc. This group of pyramids is called over complete [6] because the needed memory is larger than that for the non-compressed image.

[1]Roumen K. Kountchev is with the Faculty of Telecommunications, Technical University, Kl. Ohridsky 8, 1000 Sofia, Bulgaria, E-mail: rkountch@tu-sofia.bg

[2]Roumiana At. Kountcheva is with T&K Engineering, 1712 Sofia, Pob 12, Bulgaria, E-mail: kountcheva_r@yahoo.com

The Orthogonal pyramids are non-over complete. They are usually based on Wavelets [1] or Contourlets [7] functions and have higher efficiency and computational complexity than pyramids from the first group. The *spectral* image representation [1-4] is based on orthogonal transforms of different kind: statistical (Karhunen-Loeve Transform, Principle Component Analysis, Independent Component Analysis, Singular Value Decomposition) and determined (Discrete Fourier Transform, Discrete Cosine Transform, Walsh-Hadamard Transform-WHT, Hartley Transform, Lapped Orthogonal Transform, etc.). In this group could be included the new algebraic image transform [8] based on 2D angular windowing functions, which is suitable for the synthetic shape local phase and orientation evaluation. The transforms from the first group have higher computational complexity than these from the second one. Another approach for image representation is the *perceptual* one [9], based on anisotropic filtration controlled by the Human Visual System (HVS) visual attention model. The *knowledge-based* models for image representation are used mostly in the systems for Visual Information Retrieval. The main approach for image representation used is a pyramid model of 4 layers, which contain correspondingly: the primary matrix, the features vectors, the description of the relations between the features and the semantic image structure.

The general class of linear transform decomposes the image into various components by multiplication with a set of transform functions. Some examples are the Discrete Fourier and Discrete Cosine Transforms, the Singular Value Decomposition, and finally, the Wavelet Transform, of which the Laplacian Pyramid and other subband transforms are simple ancestors.

In this paper the IDP decomposition [5] is compared with the Laplacian pyramid (LP), because LP is the fundamental technique in this area.

The paper is arranged as follows: Section 2 is devoted to the analysis of the quantization noise on the IDP and LP decompositions; in Section 3 are presented the graphs, representing the calculation rules of IDP and LP and Section 4 is the Conclusion.

II. ANALYSIS OF THE QUANTIZATION NOISE ON THE IDP AND LP DECOMPOSITIONS

The still image is initially represented as a matrix [B] of size N×N ($N=2^n$) and elements B(i,j). In order to simplify the IDP decomposition analysis [5] is assumed that in the transform of each sub-image is retained one spectrum coefficient only. In case that this coefficient is with spatial frequency (0,0) the IDP is represented by the relation:

$$B(i,j)=\overline{B}(i,j)+\sum_{p=1}^{n-1}\overline{E}_{p-1}^{k_p}(i,j)+E_{n-1}(i,j), \qquad (1)$$

$$k_p=1,2,..,4^p, \quad i,j=1,2,..,N.$$

where

- for level $p = 0$

$\overline{B}(i, j)=I_0(\overline{B})=\overline{B}$

$\overline{B}=M_0[B(i,j)]=\left(2^n\times 2^n\right)^{-1}\sum_{i=1}^{2^n}\sum_{j=1}^{2^n} B(i,j)$

$E_0(i, j) = B(i, j) - \overline{B}(i, j),$

- for level $p = 1,2,.., n - 1$

$\overline{E}_{p-1}^{k_p}(i,j)=I_p(\overline{E}_{p-1}^{k_p}),$

$\overline{E}_{p-1}^{k_p}=M_p[E_{p-1}^{k_p}(i,j)]=(2^{n-p}\times 2^{n-p})^{-1}\sum_{(i,j)\in W_{k_p}}\sum E_{p-1}(i,j),$

$E_{p-1}(i,j)=E_{p-2}(i,j)-\overline{E}_{p-2}^{k_{p-1}}(i,j),$

- for level $p = n$

$E_{n-1}(i,j)=E_{n-2}(i,j)-\overline{E}_{n-2}^{k_{n-1}}(i,j).$

W_{k_p} is a window of size $2^{n-p}\times 2^{n-p}$ and k_p is the number of the mean difference $\overline{E}_{p-1}^{k_p}$ or interpolated $\overline{E}_{p-1}^{k_p}(i,j)$ image in the IDP layer p.

The IDP components are quantizated starting from the layer p = 0 up to layer p = n. It is assumed that the influence of the quantization noises could be represented by linear additive model. Then Eq. (1) is transformed into:

$$B'(i,j)=\overline{B}'(i,j)+\sum_{p=1}^{n-1}\overline{E}'^{k_p}_{p-1}(i,j)+E'_{n-1}(i,j), \qquad (2)$$

where:

$\overline{B}'(i,j) = I_0\{Q_0^{-1}[Q_0(\overline{B})]\}; \quad \overline{B} = M_0[B(i,j)],$

$\overline{E}'^{k_p}_{p-1}(i,j) = I_p\{Q_p^{-1}[Q_p(\overline{E}^{k_p}_{p-1})]\}, \ \overline{E}^{k_p}_{p-1} = M_p[E^{k_p}_{p-1}(i,j)],$

$E'_{n-1}(i,j) = Q_n^{-1}\{Q_n[E_{n-1}(i,j)]\}.$

Here $Q_p(.)$ and $Q_p^{-1}(.)$ are the corresponding operators for quantization/dequantization of the components in the decomposition layer p. With the mark $(.)'$ are indicated the terms, restored after dequantization, which contain additive noise components obtained in result of the quantization.

The dequantizated IDP component for the decomposition layer p=0 is represented by the sum:

$$\overline{B}'(i, j) = \overline{B}(i, j)+ \varepsilon_0(i, j), \qquad (3)$$

where $\varepsilon_0(i, j)$ is the noise ingradient.

For the next decomposition layers p = 1,2,...,n is obtained correspondingly:

$$\overline{E}'^{k_p}_{p-1}(i, j)= \overline{E}^{k_p}_{p-1}(i, j)+M_p\{\varepsilon_{p-1}(i, j)+M_{p-1}[\varepsilon_{p-2}(i, j)+... + M_1[\varepsilon_0(i, j)]]\}+\varepsilon_p(i, j). \qquad (4)$$

Тук $e_p(i,j)$ is the quantization noise for the IDP component p.

The restored image (Eq. 2) could be expressed using the original one and Eqs. 3 and 4):

$$B'(i, j) = B(i, j) + \varepsilon_\Sigma(i, j) \qquad (5)$$

where $\varepsilon_\Sigma(i, j)$ is the total quantization error for the layer p=n:

$$\varepsilon_\Sigma(i, j) = \sum_{p=0}^{n}\varepsilon_p(i, j) + M_1[\varepsilon_0(i, j)] + M_2\{\varepsilon_1(i, j) + + M_1[\varepsilon_0(i, j)]\} + ... + 2M_{n-1}\{\varepsilon_{n-2}(i, j) + + M_{n-2}[\varepsilon_{n-3}(i, j) + ... + M_1[\varepsilon_0(i, j)]\} + \varepsilon_{n-1}(i, j) \qquad (6)$$

For the LP decomposition [2] the total quantization error for the layer p=0 (which corresponds to IDP layer p=n), is defined as follows:

$$\varepsilon_\Sigma^{PL}(i,j) =\sum_{p=0}^{n}\varepsilon_p(i,j) + F[\varepsilon_n(i,j)] + F\{\varepsilon_{n-1}(i,j) + + F[\varepsilon_n(i,j)]\} + ...+ F\{\varepsilon_1(i,j) + F[\varepsilon_2(i,j) + ... + F[\varepsilon_n(i,j)]\} \qquad (7)$$

where F(.) is an operator, representing the filtration of the corresponding LP component.

The comparison of the quantzation noise distribution on IDP and LP decomposition layers shows that for the IDP the noises for the layers with icreasing numbers p=1,2, ... are much lower than these for the LP layers with decreasing numbers p = n-1, n-2,...

The noise relation for layers p=1 for IDP and p=n-1 for LP could be represented as follows:

$$\varepsilon_{1(i,j)} + M_1[\varepsilon_0(i, j)] << \varepsilon_{n-1}(i, j) + F[\varepsilon_n(i, j)], \qquad (8)$$

because

$$\frac{F[\varepsilon_n(i, j)]}{M_1[\varepsilon_0(i, j)]} = (2^{n-2}\times 2^{n-2})\frac{\sum\limits_{(i,j)\in W(2\times 2)}\sum\varepsilon_n(i, j)}{\sum\limits_{(i,j)\in W_{k_1}(2^{n-1}\times 2^{n-1})}\sum\varepsilon_o(i, j)} >> 1 \qquad (9)$$

Similarly, for IDP and LP layers p = 2 and p = n-2 correspondingly this relation is:

$$\varepsilon_2(i, j) + M_2\{\varepsilon_1(i, j) + M_1[\varepsilon_0(i, j)]\} << << \varepsilon_{n-2}(i, j) + F\{\varepsilon_{n-1}(i, j) + F[\varepsilon_n(i, j)]\}, \qquad (10)$$

because

$F\{\varepsilon_{n-1}(i,j) + F[\varepsilon_n(i,j)]\} >> M_2\{\varepsilon_1(i,j) + M_1[\varepsilon_0(i,j)]\}$, etc.

Then in result of the comparison of Eqs. (6) and (7) follows, that

$$\varepsilon_\Sigma(i, j) << \varepsilon_\Sigma^{PL}(i, j), \qquad (11)$$

i.e. the influence of the quantization noise on the restored image is much lower in IDL than in LP.

III. Graphs, representing the IDP and LP Decompositions

On Figs.1 and 2 are shown the graphs, representing the calculations of IDP and PL for image of size 4×4 pixels (N = 4). The comparison of these graphs shows that:

- The information in the IDP and LP components is identical, but arranged in inverse order;

• The volume of the accumulated quantization noises in the restored image is lower in IDP than in LP. This conclusion follows from the analysis of the noise distribution for both pyramids in accordance with Eqs. (2-10);

• IDP better corresponds to the requirement for progressive image transfer (PIT) [9] than LP, because the calculation of the IDP components starts from the pyramid top and continues to its base, while for the PL this sequence is in inverse order;

• The IDP graph (Fig.2) shows that after calculating the layer p = 0 the input data could be substituted with the so obtained, etc. – for the next layers (p = 1,2...). From this follows that the memory needed for the IDP pipline implementation is two times smaller than that for LP.

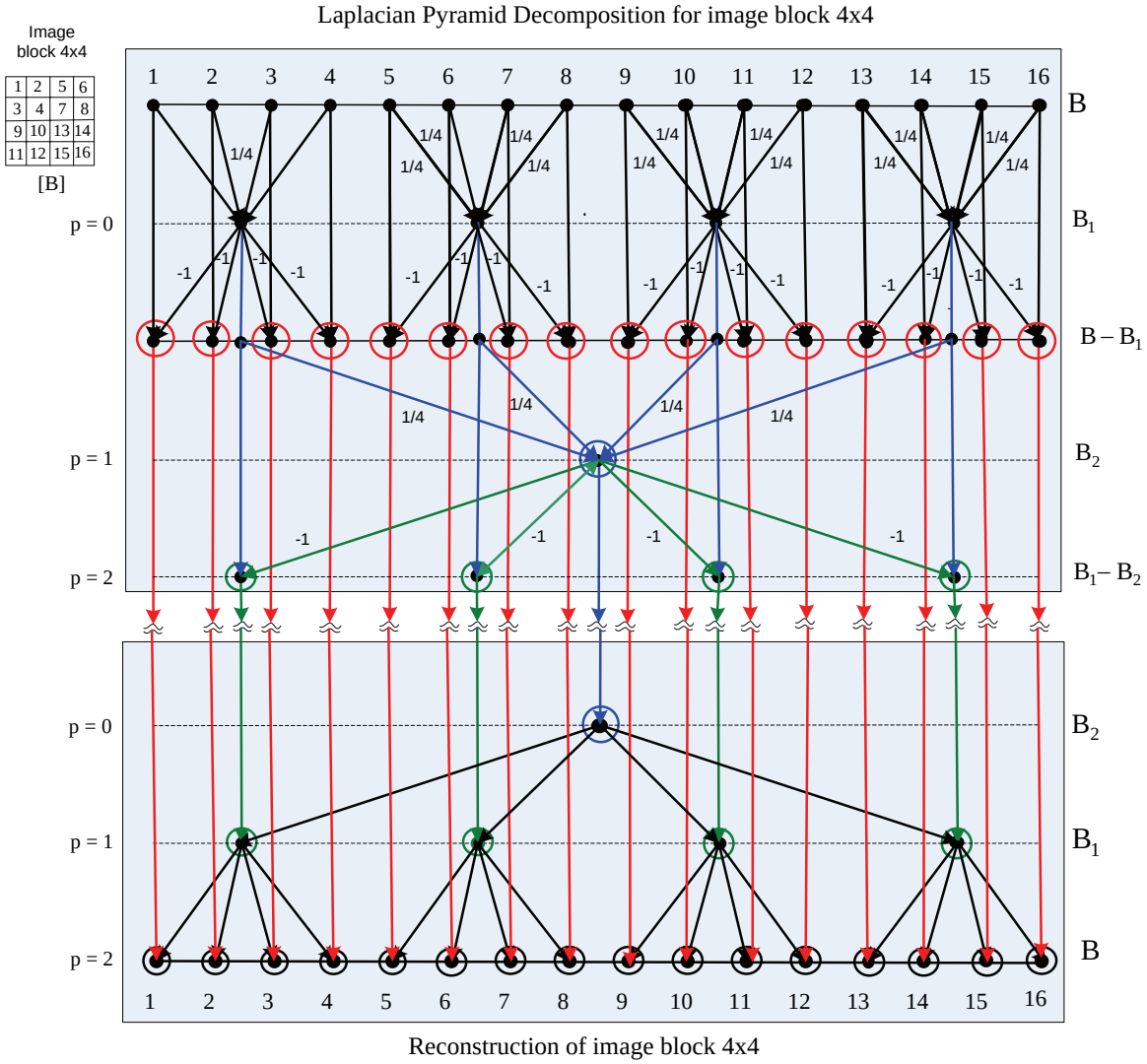


Fig. 1. LP structure graph

$$[B(i, j)] = [\overline{B}(i, j)] + [\overline{E}_0^{k_1}(i, j)] + [E_1(i, j)]; \quad k_1 = 1,2,3,4.$$

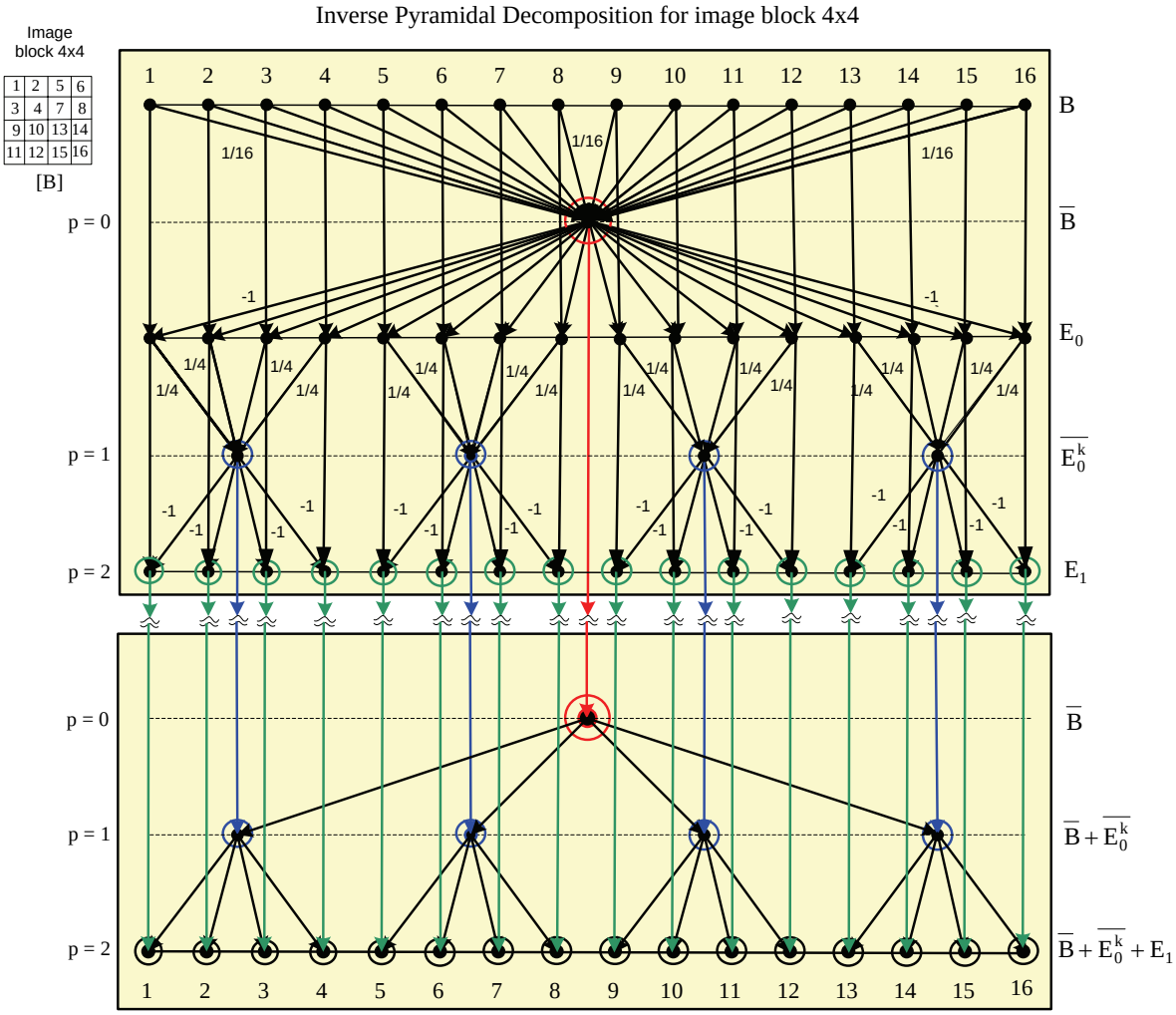


Fig.2. IDP structure graph

ACKNOWLEDGEMENT

This work was supported by the National Fund for Scientific Research of the Bulgarian Ministry of Education and Science, Contract VU-I 305.

REFERENCES

[1] R.Gonzalez, R. Woods, *Digital Image Processing*, Prentice Hall, 2001.

[2] P. Burt, E. Adelson. "The Laplacian Pyramid as a Compact Image Code". IEEE Trans. on Communications, Vol. Com-3l, No. 4, April 1983, pp. 532-540.

[3] Al Bovik. *The Essential Guide to Image Processing*. Ch. 6, Multiscale Image Decomposition and Wavelets. Academic Press, NY, 2009, pp. 123-142.

[4] Z. Liu, Y. Ho, K. Tsukada, K. Hanasaki, Y. Dai, L. Li. "Using multiple orientational filters of steerable pyramid for image registration". Information Fusion 3, 2002, Elsevier, pp. 203-214.

[5] R. Kountchev, V. Haese-Coat, J. Ronsin. "Inverse Pyramidal decomposition with multiple DCT". Signal Processing: Image Communication, Elsevier, Vol. 17, Feb. 2002, pp. 201-218.

[6] V. Goyal, M. Vetterli, N. Thao, "Quantization of over complete expansions", Proc. IEEE Data Compression Conf., Utah, USA, 1995, pp.13-22.

[7] M. Do, M. Vetterli, *Contourlets in Beyond Wavelets*, J. Stoeckler, G. Welland (eds.), Academic Press, New York, 2002.

[8] D. Zang, G. Sommer. "Algebraically Extended 2D Image Representation". 17th Intern. Conf. on the Application of Computer Science and Mathematics in Architecture and Civil Engineering, Germany, July 2006, pp. 1-10.

[9] M. Mancas, B. Gosselin, B. Macq. "Perceptual Image Representation", EURASIP Journal on Image and Video Processing, 2007, pp. 1-9.

SESSION SP II

Signal Processing II

Face Identification with Modified K-NN Classifier Based on Linear and Angular Distance

Agata H. Manolova[1] and Roumen K. Kountchev[2]

***Abstract* – Visual recognition systems use characteristic portions of the human body for identification purposes – like the face. This paper proposes a novel face identification method which exploits the global discriminative features of the human face. In this method, global features are extracted from the whole face images by keeping the low-frequency coefficients of 2D Mellin-Fourier transform. This transform ensures that the face representation is invariant to scale, rotation, translation and significant contrast and illumination changes. After that, to the feature vector of each face, a modified K-Nearest Neighbours classifier with Euclidean and cosine angle distance is applied. We evaluate the proposed method using one large-scale face database. Experiments show that the results of our method are very good for identifying a given human face.**

***Keywords* – Face identification, K-NN classifier, cosine angle distance, 2D Mellin-Fourier transform.**

I. INTRODUCTION

The field of biometrics involves identifying people by measuring parts of their bodies. It now promises to find wide acceptance as a convenient and secure alternative to typed passwords, mechanical keys, or written signatures for access to computers, facilities or vehicles, and identification for financial transactions. Personal access control systems have been implemented using visual recognition for identification of individuals. Typical of this type of access control are face recognition systems and fingerprint recognition systems. Image recognition, and particularly face recognition, is becoming an increasingly popular feature in a variety of applications. Face recognition applications can be used by security agencies, law enforcement agencies, the airline industry, the border patrol, the banking and securities industries and the like.

There are two predominant approaches to the face recognition problem: geometric (feature based) and photometric (view based) methods. The feature-based techniques use edge information, skin color, motion, symmetry, feature analysis, snakes, deformable templates and point distribution. Image-based techniques include neural networks, linear subspace methods like eigen faces. The most well studied algorithms for face identification in the literature are the Principal Component Analysis, the Independent Component Analysis, the Linear Discriminant Analysis, the Elastic Bunch Graph Matching, Kernel Methods, the Active Appearance Model and the Support Vector Machines. Given these numerous theories and techniques that are applicable to face recognition, a detailed evaluation and benchmarking of these algorithms is given in [1] and [2].

We have developed an automatic approach for one-to-one face matching that is capable to recognize an input facial image in a given database of images. This method for face identification is based on the extraction of the low-frequency coefficients from the whole face image and a modified K-Nearest Neighbors algorithm with Euclidean and angular distance, applied on the feature vectors for comparison. The current work is divided in three steps:

The first part consists of image pre-processing for the needs of the algorithm;

The second step is the development of the algorithm for face identification - the extraction of the feature vector of each face image of the database and the matching algorithm;

The last part includes the implementation of the method, applied on a face image database with 450 JPEG images with 31 faces with different lighting/expressions/backgrounds from the Computational Vision at CALTECH [3].

II. DEVELOPMENT OF THE FACE IDENTIFICATION ALGORITHM

A. Face Image pre-processing

For the needs of the proposed algorithm, there are some necessary pre-processing steps for the face images before the matching step is applied. The first step of the pre-processing is the development of an application for segmentation of the face by a skin color segmentation method combined morphological filtering as proposed in [4]. We also apply a filtering by object dimensions to be able to separate the face from any other non relevant objects with skin color. Then a rectangular frame is cut around the segmented area with the face. After that the segmented rectangular frame is transformed in greyscale i.e. a conversion from the RGB color space to the YCbCr model where we keep the Y component. The second important condition is that all the images in the database must have the same height / weight m ($m = 2^n$ and $n \in IN$) so we apply the standard nearest neighbor interpolation algorithm on the rectangular frame. An example of this application is shown on Fig. 1.

B. The feature vector extraction

[1]Agata H. Manolova is with the Faculty of Telecommunications at the TU – Sofia, 8 Kliment Ohridski Blvd, Sofia 1000, Bulgaria, e-mail: amanolova@tu-sofia.bg

[2]prof. Roumen K. Kountchev is with the Faculty of Telecommunications at the TU - Sofia, 8 Kliment Ohridski Blvd, Sofia 1000, Bulgaria, e-mail: rkountch@tu-sofia.bg .

The next step is the feature vector extraction for each segmented face and the creation of the feature vector database.

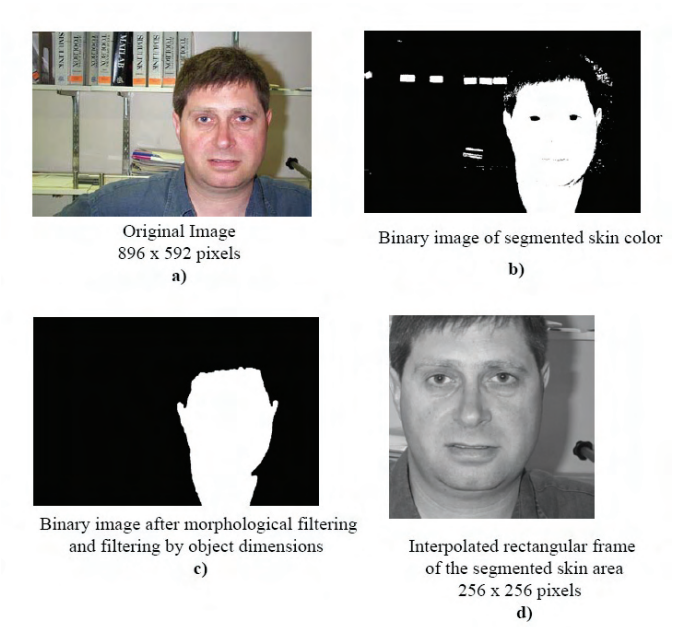


Fig. 1 Example of the pre-processing stage of the algorithm.

The idea of the method is to represent the image by a small number of meaningful coefficients that describe it at various resolution levels, and which also encode spatial relationships between various image components, by virtue of the space-frequency localization property of the wavelet transform. Meaningful coefficients can be extracted through vector quantization procedures, or by selectively picking those with the largest magnitude. Although these methods successfully address the multiple-feature integration problem, they suffer from a major weakness, namely their lack of invariance to translation and rotation, and, to a lesser extent, to scale changes. If a set of images of the same subject contains object in motion, it will be therefore difficult to recognize the object if a standard algorithm is used.

Face recognition depends heavily on the particular choice of features so we use the 2D Mellin – Fourier transform witch allows us to extract the global features from the whole face by keeping the low-frequency coefficients of 2D Mellin-Fourier transform, which we believe encodes the holistic facial information, such as facial contour. This transform ensures also, that the face representation is invariant to scale, rotation, translation and significant contrast and illumination changes [5], [6]. Experimental results on real gray-level images show that it is possible to recover an image to within a specified degree of accuracy and to classify objects reliably even when a large set of descriptors is used [7]. The feature vector extraction algorithm includes the following steps:

1. To the halftone image of Fig. 1 d) a 2D Fourier transform is applied then the magnitude specter is centered and a binary square mask (32 x 32) is applied to retain only the low-frequency coefficients at the center (Fig. 2 a) and b)). The first 2D FFT achieves translation invariance ;
2. To achieve invariance to rotation and scale transformation we perform further non-linear and non-invertible operations on the power spectrum. The log-polar transform is then used on the retained coefficients (Fig 2 c));
3. To the resulting log-polar image a second 2D Fourier transform is applied. The feature vector is constructed with a selected number of coefficients of the second magnitude spectrum using a binary mask with a special form – square 20 x 20 at each corner and first column at the left side of the image (Fig. 2 d) e));
4. A matrix with the feature vectors for all the face images of the database is constructed.

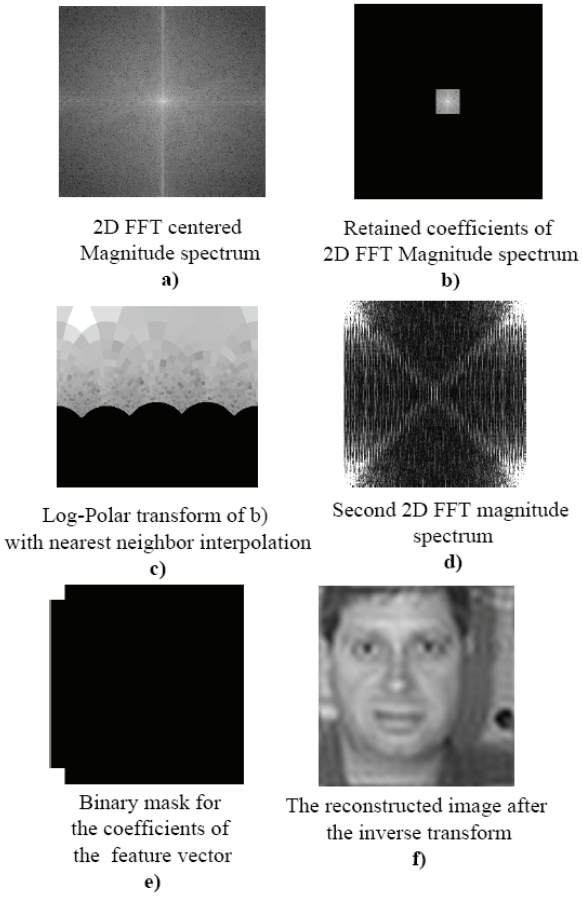


Fig. 2 Steps from 1 to 4 for the construction of the feature vector for the image from Fig. 1.

C. Face identification algorithm

The identification process represents nearest neighbourhood queries in high dimensional data space. The first step of the proposed algorithm for face identification is to calculate the Modified Euclidean Distance between each of the feature vectors in the database. It will allow a faster distance computation and combined with the modified angular distance, the whole computational time of the matching algorithm will be reduced.

The square Euclidean distance (D_E) between two vectors $\vec{X}$ and $\vec{Y}$, $\vec{X} = \{x_i, i = 1,...,N\}$ and $\vec{Y} = \{y_i, i = 1,...,N\}$, then D_E is represented as follows:

$$D_E(\vec{X},\vec{Y}) = \sum_{i=1}^{N} x_i^2 + \sum_{i=1}^{N} y_i^2 - 2\sum_{i=1}^{N} x_i y_i =$$

$$= \left\|\vec{X}\right\|^2 + \left\|\vec{Y}\right\|^2 - 2\sum_{i=1}^{N} x_i y_i \qquad (1)$$

To construct the modified square Euclidean Distance we consider two cases:

1. if $x_i \geq 0$ and $y_i \geq 0$ then

$$\sum_{i=1}^{N} x_i y_i \leq x_{\max} \sum_{i=1}^{N} y_i \qquad (2)$$

where $x_{\max} = \max\{x_i\}, i = 1,...,N$.

Then the modified distance is calculated as follows:

$$d_E(\vec{X},\vec{Y}) = \left\|\vec{X}\right\|^2 + \left\|\vec{Y}\right\|^2 - 2x_{\max} \sum_{i=1}^{N} y_i \qquad (3)$$

where $d_E(\vec{X},\vec{Y}) \leq D_E(\vec{X},\vec{Y})$.

2. if x_i and y_i are positive and negative then:

$$x_i' = x_i + |p| \qquad (4)$$

$$y_i' = y_i + |p| \qquad (5)$$

where $p = \min(x_1,...,x_N, y_1,...,y_N)$.

Then the modified distance is calculated:

$$d_E(\vec{X},\vec{Y}) = \left\|\vec{X}\right\|^2 + \left\|\vec{Y}\right\|^2 - 2x'_{\max} \sum_{i=1}^{N} y_i' \qquad (6)$$

For large vectors the computation of the square Euclidean distance is much faster because the number of multiplications and summations is reduced.

So after calculating the modified square Euclidean distance between each of the feature vectors in the database, we choose the 21 nearest neighbors to each face image. The second step of the algorithm is the calculation of the modified Cosine angle distance between these 21 neighbors. The Cosine angle distance (CAD) between two vectors $\vec{X}$ and $\vec{Y}$ is represented as follows:

$$D_\theta(\vec{X},\vec{Y}) = \arccos\left[\frac{\sum_{i=1}^{N} x_i y_i}{\left(\sum_{i=1}^{N} x_i^2\right)\left(\sum_{i=1}^{N} y_i^2\right)}\right] \qquad (7)$$

Using Eq. (2) we can rewrite the Eq. (7) as follows (modified CAD):

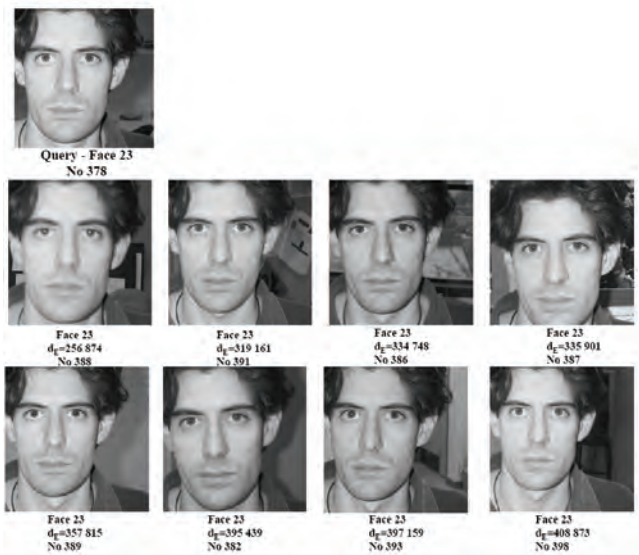
$$d_\theta(\vec{X},\vec{Y}) = \arccos\left[\frac{\sum_{i=1}^{N} y_i}{N x_{\max}\left(\sum_{i=1}^{N} y_i^2\right)}\right] \qquad (8)$$

One important property of CAD is that it gives a metric of similarity between two vectors unlike Manhattan distance and Euclidean distance, both of which give metrics of dissimilarities. Also $d_\theta \in [0,1]$. This makes it easy to combine distance between two images using multiple features. The normalization implied by the denominator in Eq. (8) prevents that two images having similar distributions of terms appear distant from each other just because one is much longer than the other. In fact, the cosine similarity seems to not outperform the conventional Euclidean distance when high dimensional spaces are concerned [8].

So we rearrange the 21 image feature vectors with modified CAD and with K-NN classifier where K= 9, we match the query face to the face witch is most frequently occurring in this neighborhood.

III. EXPERIMENTAL RESULTS

We apply the proposed identification algorithm on a face image database with 450 JPEG images with 31 faces with different lighting/expressions/backgrounds/ and male or female from the Computational Vision at the California Institute of Technology, USA. The number of face images is unevenly distributed. In the database 12 faces are present with 1, 5 or 7 images per face; the other 19 faces are present with around 20 images per face. The identification process will be performed only on these 19 faces; the others will be used in the search database. After the pre-processing in Section II the face images are with dimensions 256 x 256 pixels. After the Mellin-Fourier transform, each feature vector in has 1016 components. So the database for identification has 450 feature vectors with 1016 elements each. On Fig. 3 an identification result with a query image belonging to Face 23 with the square Euclidean distance is shown.



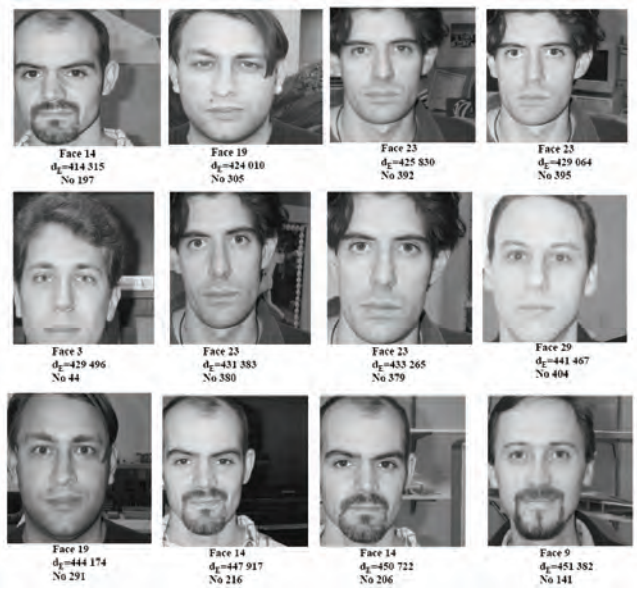


Fig. 3 Identification with a query image. The closest 20 images by square Euclidian distance are presented with their number in the database.

As mentioned in [8] the results with CAD are very similar to the results of the matching with the Euclidean distance. But the CAD is easier and faster to compute. The results of the identification for the 19 faces of the database are summarized in Table 1 – the confusion matrix after the classification with K-NN with K=9 using CAD. Only the diagonal with the rate of successful identification of the total confusion matrix is presented.

TABLE I
CONFUSION MATRIX –

Rate of successful identification in %	
Face 1	61.9%
Face 2	90%
Face 4	95.5%
Face 5	100%
Face 6	82.6%
Face 7	90%
Face 9	90.5%
Face 13	70%
Face 14	85.7%
Face 15	96%
Face 16	54.5%
Face 18	79%
Face 19	55%
Face 20	37.9%
Face 21	100%
Face 22	75%
Face 23	95.5%
Face 30	100%
Face 31	86.4%

IV. CONCLUSION

As shown on Table 1, the overall success rate of identification for all the faces is the database is 77.33%. Almost all of the faces are identified except the Face 20. For this face most the images were taken from far away and the face was not as visible as the others. The proposed method for the extraction of the feature vector of the face gives very accurate description of the characteristics of the particular face and with the very simple K-NN classifier we are able to identify each face in the database. The discrimination between female and male faces is excellent.

The method was developed in Matlab 7. Thanks to the modified Euclidean and Cosine angle distance, the search in all the database takes around 3 seconds on a standard PC configuration. The extraction of the feature vector takes around 2 seconds for each image.

The next step will be to test the algorithm on other databases not only with faces but with other objects. Also the method will be tested according to the FERET Evaluation Methodology [2]. This evaluation methodology is used by all the researchers to test the success rate for identification of faces of a given algorithm.

ACKNOWLEDGEMENT

This work was financed by a project grant of the National Fund for Scientific Research of the Bulgarian Ministry of Education and Science by the contract VU-I-305.

REFERENCES

[1] K. Delac, M. Grgic, S. Grgic, Statistics in Face Recognition: Analyzing Probability Distributions of PCA, ICA and LDA Performance Results, Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis, ISPA 2005, Zagreb, Croatia, 15-17 September 2005, pp. 289-294.

[2] P.J. Phillips, H. Moon, S.A. Rizvi, P.J. Rauss, The FERET Evaluation Methodology for Face-Recognition Algorithms, IEEE Trans. on Pattern Recognition and Machince Intelligence, Vol. 22, No. 10, October 2000, pp. 1090-1104.

[3] http://www.vision.caltech.edu/index.html

[4] A. Elgemmal, C. Muang, D. Hu, "Skin detection – a short tutorial" in S.Z. Li and A. Jain [eds], *Encyclopedia of Biometrics*, Springer Verlag, 2009.

[5] R. Milanese, M. Cheruliez, "Invariant content-based image retrieval using the Fourier-Mellin transform, OFES grant 95.0493, EU ACTS project AC089.

[6] Richard Alan Peters II, "On the Computation of the Discrete Log-Polar Transforms", IEEE Transactions on Image Processing (in review), 14.01.2008.

[7] S. Derrode, F. Ghorbe, "Robust and efficient Fourier-Mellin transform approximations for Gray-level image reconstruction and complete invariant description", Computer Vision and Image Understanding, Elsevier Science Inc, Volume 83 , Issue 1 (July 2001), pp. 57 – 78.

[8] G. Qian, S. Sural, Y. Gu, S. Pramanik, "Similarity between Euclidean and cosine angle distance for nearest neighbor queries." In Proc. of the 2004 ACM symposium on applied computing, pp. 1232-1237, 2004.

Single Eye Gaze Tracking With Active Pan-Tilt Camera

Stanislav V. Panev[1], Ognian L. Boumbarov[2] and Plamen P. Petrov[3]

***Abstract* – In this paper an algorithm is proposed for estimation of human gaze direction by means of a geometrical model of the eye based on distances between the pupil centre and the two eye corners and artificial infra-red switching lightning. For user comfort an active pan-tilt camera is employed, which makes possible free head movements instead requiring the user to be still.**

***Keywords* – eye, gaze, tracking, active camera, pan tilt camera.**

I. INTRODUCTION

Nowadays human-computer interaction takes one of the main positions of scientific researches and developments. Estimation of gaze direction is one very interesting part of it, thus a lot of effort is made in attempting to make computers see where we look at. This way interacting with computer will be much easier for the user because we perceive most of the surrounding information by our eyes.

A large number of systems for eye tracking are available. Many of them are based on contact lenses, electrodes, specialized hardware, and infrared emitters. Such systems could easily be separated into three groups, according to the approach used: methods, using the electric potential of the human skin [1], methods that involve contact lenses [2] and methods that involve image analysis.

In general image based approaches are divided into two groups – appearance-based and model-based.

Appearance-based approaches directly treat an eye image as a high dimensional feature. Baluja and Pomerleau use a neural network to learn a mapping function between eye images and gaze points (display coordinates) using 2,000 training samples [3]. Tan et al. take a local interpolation approach to estimate unknown gaze point from 252 relatively sparse samples [4]. Recently, Williams et al. proposed a novel regression method called S3GP (Sparse, Semi-Supervised Gaussian Process), and applied it to the gaze estimation task with partially labeled (16 of 80) training samples [5]. Appearance-based approaches can make the system less restrictive, and can also be very robust even when used with relatively low-resolution cameras.

Model-based approaches use an explicit geometric model of the eye, and estimate its gaze direction using geometric eye features. For example one typical feature is the pupil glint vector ([6], [7]), the relative position of the pupil centre and the specular reflection of a light source. Model-based approaches typically need to precisely locate small features on the eye using a high-resolution image and often require additional light sources but can be very accurate.

Several approaches exist for eye-tracking by using images. Some of them are largely dependent on active light-sources, such as infra-red emitters [8]. Other approaches, on the other hand, do not use the information from an explicit light-source, but rather use information for improving the quality of the image. Still other approaches even exclude the usage of active lighting, but rather rely on natural light-sources.

The rest of the article is organized as follows: in Section 2 an approach is described for human gaze tracking which relies on detecting pupil and the two eye corners locations. In order to estimate the pupil's centre coordinates a double switching infra-red lightning is used which turns on and off consecutively in two neighboring frames the camera on- and off-optical axis LEDs. In order to estimate the location of the two eye corners a deformable template of the eye is used. This template consists of two parabolas which depicts the two eye lids. The two cross points of the two parabolas are the inner and outer eye corners. Assuming that the eye is a spherical body we can estimate the gaze direction. In Section 3 is described the control system of a Pan-Tilt unit which is employed to centre the image of the eye in the frame. Thus the user has the freedom to move its head instead of standing still. In Section 4 we present our experimental results. In Section 5 we conclude and discuss the future development of our algorithm.

II. GAZE TRACKING ALGORITHM DESCRIPTION

A. Assumptions

Before presenting the operations implemented to realize the gaze tracking procedure some basic assumptions should be taken in mind.

First of all the working light spectrum is in the near-visible infra-red domain. The visible light is cut by a infra-red filter. This way a lot of noise is eliminated coming from the surrounding light sources and the contrast of the pupil is much better.

Only one eye should be presented in the frame. This way a higher resolution of the eye image is achieved with low resolution cameras, which increases the accuracy of tracking and algorithm confusion is avoided. The rest of the image is filled with the face skin.

All the images taken from the camera are in the grayscale space.

[1]Stanislav V. Panev is with the Faculty of Telecommunications, 8 Kl.Ohridski, 1000 Sofia, Bulgaria, E-mail: s_panev@tu-sofia.bg

[2]Ognian L. Boumbarov is with the Faculty of Telecommunications, 8 Kl.Ohridski, 1000 Sofia, Bulgaria, E-mail: olb@tu-sofia.bg

[3]Plamen Petrov is with the Faculty of Machine Technology, 8 Kl.Ohridski, 1000 Sofia, Bulgaria, E-mail: ppetrov@tu-sofia.bg

B. Pupil segmentation in image

Pupil segmentation is implemented according to the method described by Y. Ebisawa in [9]. It uses the ability of the eye's retina to reflect the light beam which penetrates in it exactly in the same direction where it came from. Hence if two light sources are employed, one on and one off the camera optical axis, the pupil in frame when the on-axis light is on will be bright. Respectively in the frame shot during the off-axis light is on the pupil will be dark. The rest of the image will have almost the same intensity in the two frames because the lights are close enough to the camera and object. So if the pixel values of the two frames are subtracted the greatest difference will have these pixels which belongs to the pupil (Fig. 1).

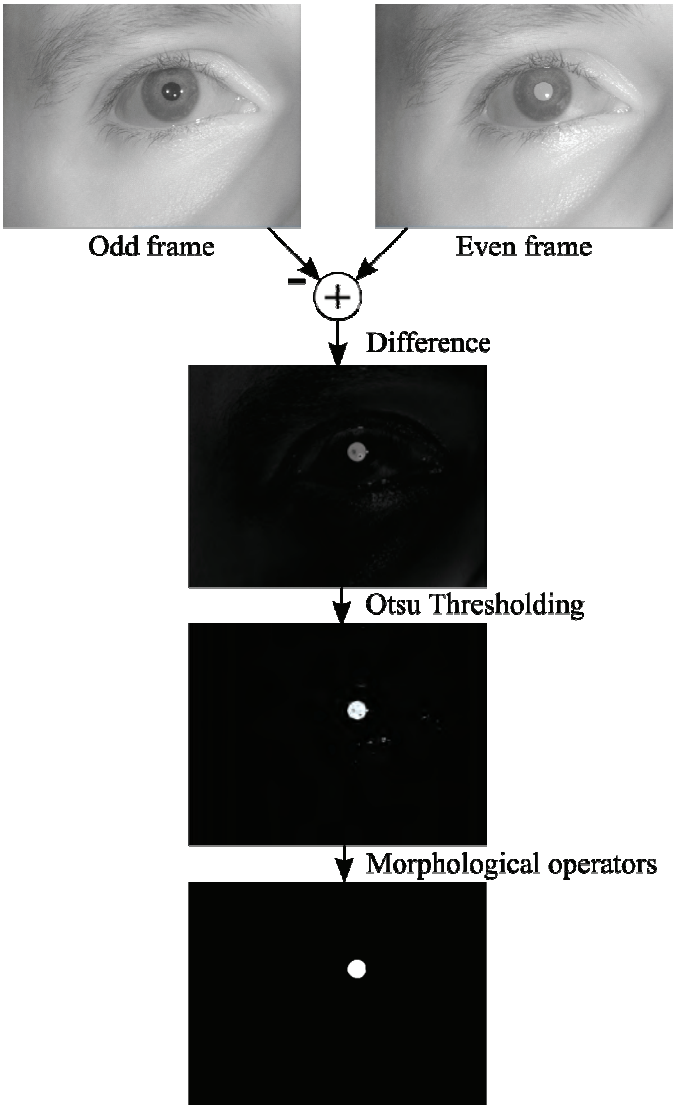


Fig. 1. Pupil segmentation procedure.

As can be seen from Fig. 1 after binarizing the difference image with optimal threshold based on Otsu algorithm, there are some little white dots which can be assumed as noise. By using combination of morphological operators erosion and dilation all these noise is suppressed and the white spot in the final image shows where exactly the pupil is in the image. This way finding the position of the pupil in the image is very fast and doesn't require complex tracking algorithms. As a result the center of the white spot is estimated.

C. Eye corners detection

The detection of the eye corners is achieved by using deformable template of the eye lids showed on Fig. 2. The template which has been used consists of two parabolas which represent the top and bottom eye lids. Hence the two points where the two parabolas intersect with the other one are the two eye corners.

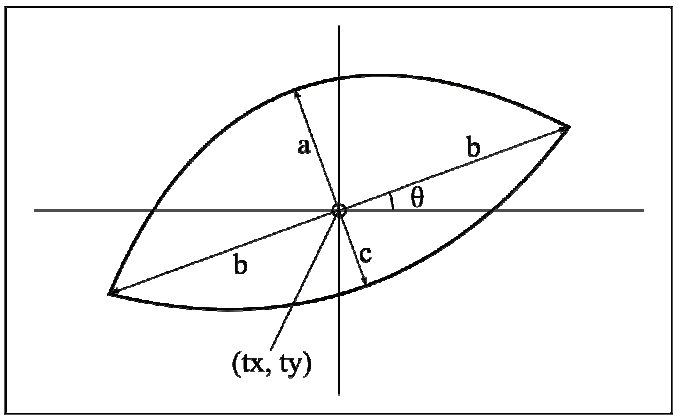


Fig. 2. Deformable template of eye.

The method for template fitting is described by Jyh-Yuan Deng in [10] and uses so called regional forces. These forces can be obtained by calculating the window force f_W (Fig. 3) of a small window from the image according to the following equation:

$$f_W = \frac{\sum each\ pixel\ value\ in\ the\ window}{number\ of\ the\ pixels\ in\ the\ window} - \frac{255}{2} \qquad (1)$$

If we slide this window along a known curve (the two parabolas) and integrate the window force for all the positions, the result will be the *regional force* which "tries" to deform the template such way that the forces will be equal to zero. As the force is has a linear effect this means that the regional forces will try to move or scale the template but not rotate. That's why the concept for the *regional torque* should be introduced (Fig. 3). If a pivot point is chosen as a center of the rotation, for example the center of the pupil which is already known, it is easy to estimate the torque for rotation of the template with the following equation:

$$\tau_{clockwise} = \sum_i \text{each } f_{w_i} w_i \text{ in the left} - \sum_j \text{each } f_{w_j} w_j \text{ in the right} \quad , \qquad (1)$$

where w_i is the weighting factor equivalent to the effective distance to the pivot point for calculating the torque.

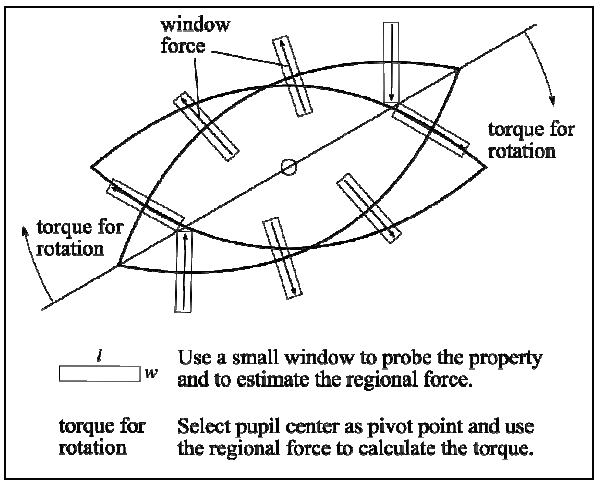


Fig. 3. Regional torque for rotation

D. Gaze estimation geometric model

In this subsection a description of the geometric eye model which we use to estimate gaze direction is given. There is nothing particularly novel about this model. We assume that the eyeball is spherical and the inner and outer eye corners have been estimated, in our case using a deformable template. The algorithm is split into two steps [11]:

1. Estimate the center and the radius of the eyeball in the image from the eye corners and the head pose.
2. Estimate the gaze direction from pupil location, the center and the radius of the eyeball.

The first of these two steps requires the following anatomical constants (Fig. 4):

- R_0 : The radius of the eyeball in the image when the scale of the face is 1.
- (T_x, T_y): The offset in the image between the mid-point of the two eye corners and the center of the eyeball.
- L : The depth of the center of the eyeball relative to the plane containing the eye corners.

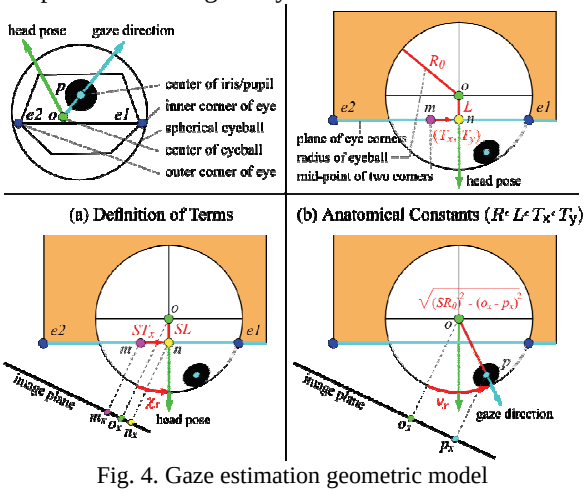


Fig. 4. Gaze estimation geometric model

The center and the radius of the eyeball are computed using the following three steps:

1. The mid-point (m_x, m_y) between the inner corner $(e1_x, e1_y)$ and the outer corner $(e2_x, e2_y)$ is estimated:

$$\begin{pmatrix} m_x \\ m_y \end{pmatrix} = \begin{pmatrix} \frac{e1_x + e2_x}{2} \\ \frac{e1_y + e2_y}{2} \end{pmatrix}. \quad (2)$$

2. The scale of the face S is computed by

$$S = \frac{\sqrt{(e1_x - e2_x)^2 + (e1_y - e2_y)^2}}{\cos\varphi_x} \quad (3)$$

3. The center of the eyeball (o_x, o_y) is then computed as the mid-point (m_x, m_y) plus two corrections:

$$\begin{pmatrix} o_x \\ o_y \end{pmatrix} = \begin{pmatrix} m_x \\ m_y \end{pmatrix} + S\begin{pmatrix} T_x\cos\varphi_x \\ T_y\cos\varphi_y \end{pmatrix} + SL\begin{pmatrix} \sin\varphi_x \\ \sin\varphi_y \end{pmatrix} \quad (4)$$

Gaze direction (θ_x, θ_y) can then be estimated as follows:

$$\begin{pmatrix} \sin\theta_x \\ \sin\theta_y \end{pmatrix} = \begin{pmatrix} \frac{p_x - o_x}{\sqrt{R^2 - (p_y - o_y)^2}} \\ \frac{p_y - o_y}{\sqrt{R^2 - (p_x - o_x)^2}} \end{pmatrix} \quad (5)$$

The anatomical constants R_0 , (T_x, T_y) and L are pre-computed in an offline training phase.

III. Pan-Tilt Camera Motion Algorithm Description

The viewing angle of the camera is comparatively small (8,24° in horizontal and 6,18° in vertical direction for ¼" camera sensor and focal length 25 mm) which is a consequence of the assumptions depicted in Section 2.1. This means that there is a real possibility parts of the image of the eye or even the whole of it to go out of the frame if the user moves itself a little in some direction. That's why for its comfort and robustness of the tracking procedure an mechanical Pan-Tilt unit for camera motion is employed. The aim of this mechanical system is to position the image of the eye in the center of the frame.

As the inertia of the camera is comparatively small (its weight is about 100 – 200 gr.), it can be ignored to simplify the law of movement. Besides we assume that the user is seated and the movement limits of the head will be in the interval ±10º from the zero position of the camera and the movement won't be rabid and abrupt. Hence the control law can be simplified to a purely proportional law by eliminating the integral and differential part of a PID controller.

The input data for the control system are the coordinates of the pupil center (x_i, y_i) in the image in relation to a coordinate system which is placed in the center of the frame. Therefore the system has to aspire to move the camera such direction that the pupil coordinates are 0. The angles of rotation of the camera can be estimated from:

$$\begin{aligned} \wp_x &= arctg\left(\frac{w_{ccd}}{w_p f}.x_i\right) \\ \wp_y &= arctg\left(\frac{h_{ccd}}{h_p f}.y_i\right) \end{aligned}, \qquad (6)$$

where w_{ccd} and h_{ccd} are the physical width and height of the camera's CCD sensor respectively, w_p and h_p are the size of the frame in pixels in vertical and horizontal direction respectively and f is the focal length of the lens.

The control of the camera rotation is done through its angular velocity ω. Thus as far the pupil is from the image center as faster the camera will rotate its position. On Fig. 5 a block diagram of the camera control system is depicted.

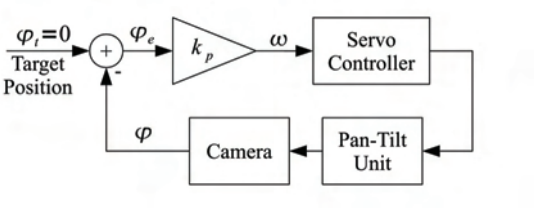


Fig. 5. Block diagram of Pan-Tilt Unit control system

IV. Experimental Results

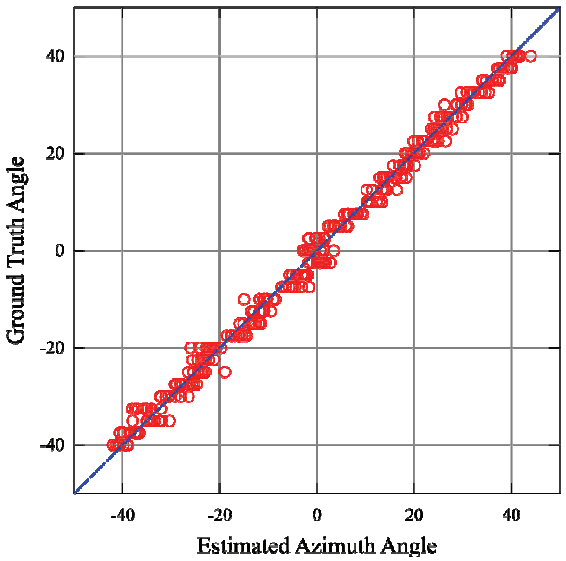


Fig. 6. Azimuth angle estimation of an object

The described above system for gaze estimation has been tested in real conditions and gave quite satisfying results. An experimental estimation of the azimuth angle of the gaze direction is shown on Fig. 6. The calculated mean squares error is 1,56º.

V. Conclusion

In this paper an approach to gaze direction estimation of a single eye image was presented. It uses a deformable template and double switching infra-red lighting to estimate the position of the pupil and the eye corners. Hence by the means of an eye geometrical model the direction of the user gaze is estimated. The system behaves notably robust and accurate.

Acknowledgement

This work was supported by the National Ministry of Science and Education of Bulgaria under contract BY-I-302/2007: "Audio-video information and communication system for active surveillance cooperating with a mobile security robot".

References

[1] J. Gips, P. Olivieri, "EagleEyes: An Eye Control System for Persons with Disabilities," The Eleventh Intern. Conference on Technology and Persons with Disabilities, 1996, pp. 128-132.

[2] L. Young and D. Sheena, "Methods & Designs: Survey of Eye Movement Recording Methods," Behavioral Research Methods & Instrumentation 7(5), 1975, pp. 397-429.

[3] S. Baluja and D. Pomerleau, "Non-intrusive gaze tracking using artificial neural networks," Advances in Neural Information Processing Systems (NIPS) 6, 1994, pp. 753–760.

[4] K. H. Tan, D.J. Kriegman, and N. Ahuja, "Appearance-based eye gaze estimation," In: Proceedings of the Sixth IEEE Workshop on Applications of Computer Vision (WACV 2002), 2002, pp. 191–195.

[5] O. Williams, A. Blake, and R. Cipolla, "Sparse and semi-supervised visual mapping with the S3GP," Proceedings of the 2006 IEEE Conference on Computer Vision and Pattern Recognition, 2006, pp. 230–237.

[6] T. E. Hutchinson, K. P. White Jr., W.N. Martin, K.C. Reichert, and L.A. Frey, "Human-computer interaction using eye-gaze input," IEEE Transactions on Systems, Man and Cybernetics 19(6), 1989, pp. 1527–1534.

[7] R.J. Jacob, "What you look at is what you get: eye movement-based interaction techniques," In: Proceedings of the SIGCHI conference on Human factors in computing systems, 1990, pp. 11–18.

[8] C. Morimoto, D. Koons, A. Amir, and M. Flickner, "Pupil detection and tracking using multiple light sources," IVC18(4), 2000, pp. 331-335.

[9] Y. Ebisawa, "Improved Video-Based Eye-Gaze Detection Method", IEEE Transactions on Instrumentation and Measurement, Vol. 47, No. 4, August 1998

[10] JY. Deng, F. Lai, "Region-based Template Deformation and Masking for Eye-feature Extraction and Description", Pattern Recognition, The Journal of Pattern Recognition Society, Vol. 30, No. 3, pp. 403-419, 1997

[11] T. Ishikawa, S. Baker, I. Matthews, T. Kanade, "Passive Drive Gaze Tracking with Active Appearance Models", In Proceedings of the 11th World Congress on Intelligent Transportation Systems, October 2004

Discrete Model of Isotropic Excitable Media for Image Processing

Ognyan I. Zhelezov[1]

***Abstract*-In this paper is proposed one model of discrete excitable media for wave propagation. Model provides a constant wave velocity along all directions, which gives possibility for effective parallel image processing. The state of element of discrete media is described by the vector S=(X,Y,O). During the n-tm step of wave propagation every element receives state vectors of it's neighbour elements and calculates the minimal Euclidean distance to the source of excitation. During the process of changing the state of elements, every element will switch to active state after a time, that depends of it's Euclidean distance to the source of excitation. Thus, having only local relations, discrete media provides a spherical front of wave propagation.**

***Keywords*: texture generation, shapes modulation, covering of surface**

I. INTRODUCTION

Wave processes in excitable media (WPEM) are non-linear wave processes. For continuum they are described by partial differential equations of type:

$$\frac{du}{dt} = D.\frac{d^2u}{dx^2} + f(u) \qquad (1)$$

Equation (1) describes change of state of element u as a function of partial co-ordinates x,y and time t. It is derivative of equation, which describes process of diffusion – spatial and time changing of concentration, known as Fick's second laws of diffusion [2]. Discrete model of excitable media is cellular automata, in wich the state of a particular cell in the next time step depends on the state of the neighbours cells at the current time. Discrete models of WPEM gives possibilty for effective parallel image processing: recovering of contours, smoothing of contours, obtaining sceleton, etc.

II. BINARY CELLULAR MODEL OF WPEM

In the simplest binary model of excitable discrete media every element has only two possible states (we can denote them with 0 and 1). Algorithm of wave propagation can be described as follows:

If at the given time step one element has state 0 and at least one neighbour element has step 1, in the next time step this element switch to state 1. in opposite case element keeps it's state.

This model, despite it's simplicity, gives possibility to examinate wavefront of WPEM. However, models of excitable medium need more then two states of elements – at least one additional state – refractory state -- the element has recently been excited and has not yet been through the refractory period. Refractory period can continue one or more steps of wave propagation.

III. TYPE OF WAVEFRONT IN DISCRETE PERIODICAL SET OF ELEMENTS

Algorithm of binary WPEM, described above, defines the changing element's state rule as a simple logical function on neighbour elements state. To define precisely this function, it's need to define wich are the neighbour elements of given element. Most often descrete medium is considered obtained from continuum using rectangular mesh for discretization.

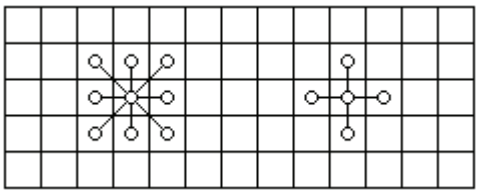


Fig. 1. 8-Connectivity and 4- Connectivity between elements of rectangular mesh

This definition of connectivity in a digital pattern need clarification: which pixels are neighbor pixels? Answer of this question (the need of which is a specific problem of digital images) is related to definition of neighborhood of elements in rectangular mesh. The figure below shows two types of connection;

- 4- Connectivity – neighbors are the elements, which share an edge
- 8- Connectivity – neighbors are the elements, which share an vertex

Difference between two types of connectivity deterninates difference between wavefronts of WPEM from point source. Wavefront from point source for WPEM with 4-connectivity for k steps of wave propagation is presented as follows:

$$Y4(n1,n2,k) = \delta[|n1+|n2|-k] , \qquad (2)$$

Wavefront from point source for WPEM with 8-connectivity for k steps of wave propagation is presented as follows:

[1] Ognyan. I. Zhelezov is with the Faculty of Computer Science and Automation , Studentska STR. 1 9000 Varna , Bulgaria, E-mail: ognyanz@yahoo.com

$$Y8(n1,n2,k)= \delta[\max(|n1|,|n2|)-k] \quad (3)$$

where:
n1 and n2 are co-ordinates of element in discrete media.
κ – discrete time step.
δ{…} – discrete impuls function,
max(|n1|,|n2|) dnotes the biggist of two co-ordinates, given as absolute values.

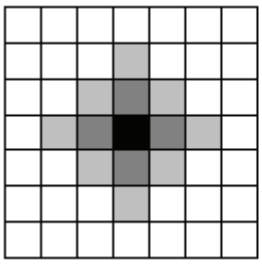


Fig. 2a - Wavefront from point source for 4 connectivity

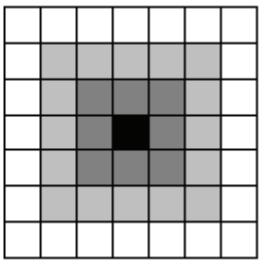


Fig. 2b - Wavefront from point source for 8 connectivity

Fig. 2 shows propagation of WPEM from source point in rectangular mesh of elements, having 4 and 8 connectivity. Image on Fig. 2a presents the state of discrete media after two steps of WPEM for 4 connectivity between elements, while image on Fig. 2b presents the state of discrete media after two steps of WPEM for 8 connectivity between elements. Comparing of wavefronts shows, that on the direction of axes n1 and n2 of coordinate system velocity of wavefront of WPEM for four-connectivity and eight-connectivity is the same – wavefront moves to the next discrete element on every step of wave propagation. However, along the rest directions, velosities are different: high velocity has propagation of wavefront на of WPEM for eight-connectivit.

The common link between the two wavefronts is their zero curvature, which shows anisotropic nature of examinated discrete excitable media. This vastly restrict its possibilities for image processing. In example using of WPEM for scalling of contour images gives as a result at the same time change of dimensions and change of form of the contour, because the contour line moves with the different velocity along different directions.

Fig. 3a shows one digital image, which defines initial state of elements in rectangular mesh. White color present elements, which are in initial (non-excited) state. Black color present elements, which are in active (excited) state. Gray color present refracter state (state, after which element switch to inactive state regardless the state of neighbour elements

As can be seen, contour, obtained after 10 steps of WPEM with 4 connectivity beteen elements not only scale the given contour, but change the form of contour. The cause of this result, as was told above, is thedifferent velocity of wave along different directions. Exactly this type of connectivity gives hier velosity along direction of axes X and Y of coordinate system, that's why contour curve is more moved along these directions

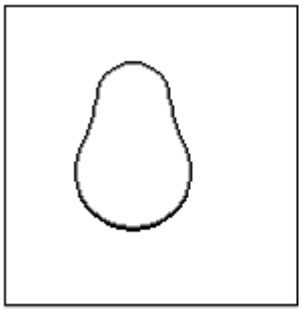


Fig. 3a – Given contour image

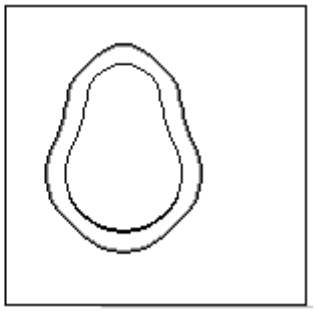


Fig. 3b - Given contour image and contour image, obtained after 10 steps wave process of with 4 connectivity

The simular result we will get using WPEM with 8 connectivity beteen elements. Fig. 4b shows the given contour and the contour, obtained after 10 steps of wave process in mesh with 8 connectivity. Result again is change of firm of contour, the cause of thedifferent velocity of wave along different directions. Exactly this type of connectivity gives hier velosity along direction of axes X and Y of coordinate system and along their's angular bisectors, that's why contour curve is more moved along these directions

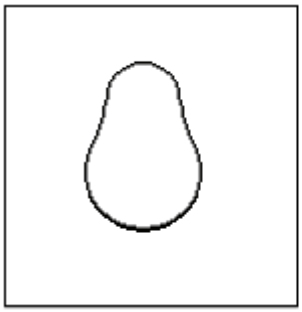


Fig. 4a – Given contour image

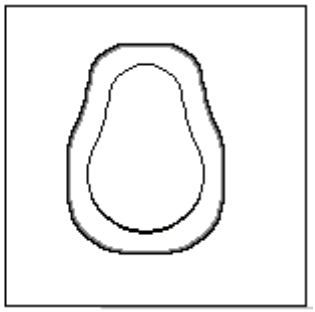


Fig. 4b - Given contour image and contour image, obtained after 10 steps wave process of with 8 connectivity

This report offers isotropic excitable discrete media, which can be used for some tasks of image processing.

IV. Model of isotropic excitable discrete media

The cause for the anisotropic behavior of the regular discrete mesh is the regular and local type of connectivity between elements. To have the same velocity in all directions, it is needed to have wavefront from point source, which looks as follows:

$$Y(n1,n2,m1,m2,k) =$$

$$\delta\left(k-\sqrt{(n1-m1)^2+(n2-m2)^2}\right) \qquad (4)$$

where:
n1 and n2 are co-ordinates of element in discrete media.
κ – discrete time step.
δ{…} – discrete impuls function,
m1 and m2 are co-ordinates of point source.

To obtain this type of wavefront it is needed to have the time of change the state of elements proportional to Euclidean distance of that element to the source of excitation. It is easy to obtain this for discrete media, in which every element is connected to all others. If some element in moment 0 is in excitable state, every of all others elements can calculate it's own moment, in which have to change its state. Problem is , that this type of discrete mesh has too mush connections – it has quadratic relationships between the number of elements N and number or connections R

The state of element in discrete mesh can be presented by vector S=(X,Y,O) where X и Y are the distances to source point on direction of co-ordinate axes OX и OY , measured in number of elements, and O is the output signal of current element.

The steps of proposed algorithm are:

1) Wave process for changing the state of elements in mesh.
2) Change of output signals of elements as a function of its state and of its distance to the source.

Changing the state of elements is performed as follows: At the time of clock n of discrete wave process every one element receives from it's neighbour elements their state vectors and calculates it's minimal Euclidean distance to the source point of exitation. At the time of next step of algorithm – changing the state of elements, every element switch to active state after the time, proportional to Euclidean distance D of that element

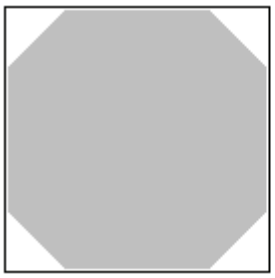
$$D = \sqrt{X^2 + Y^2} \qquad (5)$$

to the source of excitation This way, having only connections between neighbour elements, proposed discrete media ce provide sзrerical form of wavefront.

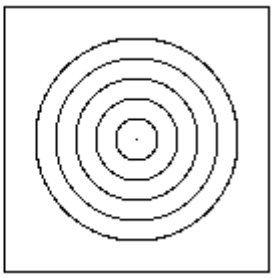
V. PRACTICAL RESULTS.

Has been created a program, which demonstrate using of algorithm, described above, for scaling of contour images.

Fig. 5 shows the results of wave process from source point in proposed discrete media. Fig. 5a shows area of elements, for which has been calculated the distance to the source point after 100 steps of wave process.



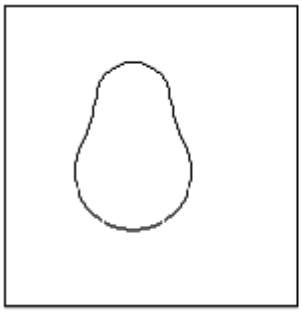
Фиг. 5a – Area of elements, whichdistance to source point has been calculated



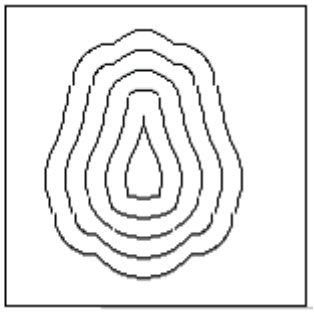
Фиг. 5b - Wavefronts from source point

Fig. 5b shows wavefront subsequently on 5-th, 10-th, 15-th, 20-th и 25-th step of process of change the output singnal of elements. As can be seen, wave fronts are spherical (circles) for all steps of wave propagation.

Fig. 6a shows given contour image, while Fig. 6b – given contour image and wavefronts on distances 10 и 20 pixels, obtained using proposed algorithm.



Фиг. 6a – Given contour image



Фиг. 6b – Given image and wavefronts on distance 10 and 20 pixels

This example demonstrates possibilities to obtain a contour curves on the given distance to the object in digital image.

VI. CONCLUSION

The proposed algorithm for wave propagation in discrete excitable media provide a spherical front of wave propagation from source point and identical velocity along all directions of wave propagation. However, it's importatn to take in consideration, that discrete media, wich works as it ti described above is not a linear system, which means, that it does not satisfied the superposition principle. That's why wave, obtained from two source points is not the sum of waves, wich can be obtained from this source points applied separately. This gives as a result specific distortion of wavefront, as can be seen in Fig. 6b. One possible decision of this problem is using of multilayer medium, in which at least neighbour source points "works' on different layers. In this model, one additional layer can make superposition of waves of waves of all layers, in which partial waves are propagated

REFERENCES

[1] Зыков, В. С., Моделирование волновых процессов в возбудимых средах. М. “Наука” 1984г.

[2] Железов О, Метод за изглаждане на контури. Научна сесия “Дни на науката - Варна”, 1998г

[3] Железов О, Възстановяване на контурни изображения посредством автовълнов процес, списание “Автоматика и информатика”, брой 5-6, 1998г

[4] Железов О, Генериране на осветеност на обекти посредством автовълнов процес, сборник доклади от конференция “Морски научен форум”, том 2, 1998г

[5] Рабинер Л, Б. Гоулд, Теория и применение цифровой обработки сигналов. М. “Мир” 1978г.

[6] Kung S, Whitehouse H and Kailath editors, VLSI and modern signal processing, Prentice hall inc. New Jersey, 1989

[7] Павлидис Т. , Алгоритмы машинной графики и обработки изображений. пер. Н.Гуревич, М. “Радио и связь”, 1986г.

[8] Kaplan Daniel, Leon Glass, Understanding nonlinear dynamics, Textbooks in Mathematical Sciences Series. New York: Springer-Verlag. 1995

[9] http://en.wikipedia.org/wiki/Excitable_medium Wikipedia – Excitable Medium

Algorithm for Short Term Conflict Detection, Using Multi-Objective Optimization of the Decision Making Process

Ivan Korobko and Polina Marinova,

Abstract: **This paper presents a review of short term conflict detection methods that are in use in current algorithms for conflict detection and gives some arguments using multi-objective optimization. Evaluation of different methods for multi-objective optimization and their applicability for making decision of effectiveness of conflict detection algorithm.**

Key words: **short term conflict ,multi-objective optimization, Pareto optimization, vector synthesis, function of parameters.**

I. INTRODUCTION

Taking into consideration safety and the increasing of the efficiency of ATC, the questions for conflict alerting in due time become a hot topic. The key element in the conflict prediction and uncertainty modeling, common for aircraft motion, because of the wind, commands of ATC, Navigation and control. This paper presents short –term conflict detection in the case of optimizing criteria for estimated value of the distance between two aircrafts in each moment of their flight. The probabilistic method is used for calculation of the estimate aircraft position taking into account all potential trajectories. In this method uncertainties are modeled to describe potential variations in the future trajectory of the aircraft. All situation are described by their "possibility of conflict".

II.BASIC APPROACHES TO MULTI-OBJECTIVE OPTIMIZATION

A)Possibility of conflict

The "probabilistic drift" has a Gaussian density with increasing covariance matrix. The validity of the prediction decreasing during the time interval. That model has a lot of disadvantages, but a probabilistic approach provides an opportunity for a balance between relying on either a single-trajectory model or a set of worst- case maneuvers. Short-term conflicts are time limited up to 10 min.

Conflict detection uses current aircraft position to describe estimate future trajectory of the aircraft. The ordinary propagation of the trajectory is for 5-10 min. During that time the aircraft keeps its initial flight path especially straight forward.

Prof. Ivan E.Korobko, Technical University- Sofia, Bulgaria, e-mail: ikor@abv.bg

Eng.Polina I.Marinova, Technical University –Sofia, Bulgaria, e-mail: polina.marinova@sofco.bg

Conflicts are detected, by checking the distance ($\vec{R}(x,y,z,t)$) between the flight paths at time steps of Δ seconds.

Since the algorithm does not compute the actual time when the first loss of separation occurs, the choice of Δ is crucial.

(3) The major contributor to the uncertainty in the aircraft motion is wind, for which a consistent physical model is not available yet. However, since the prediction error can be modeled as the sum of a large number of independent random perturbations in disjoint time intervals, it is expected to be Gaussian. This hypothesis was indeed verified by empirical data, which also suggested that the uncertainty can be decomposed in to two components: an along track component whose variance grows with time, and a cross track component whose variance remains constant.

According to the minimum vertical and horizontal separation is designed a protected zone for each aircraft. Protected zone is a cylinder with 2000 ft or 4000ft height, depends on the flight level, and radius of the bases- 5 nm. In the center of that cylinder is situated the aircraft. The conflict probability is equal to the probability of intersecting of PZs of two aircrafts and for its calculation the predicted future positions of the two aircrafts must be used. If the conflict probability is divided in two components, the conflict will appear at every moment when any of the separations is broken.

$$P(C) = P(C)_{H.} P(C)_{V} \quad (1)$$

Conflict detection as a process is similar to radar signal detection. The conflict detection algorithm has the following steps:

1)detection and processing with the data signal for ($\vec{R}(x,y,z,t)$)

2)calculation of the predicted value of ($\vec{R}(x,y,z,t)$)

3)comparison between the calculated value of R and the threshold. Making decision there is a conflict or not for given future period.

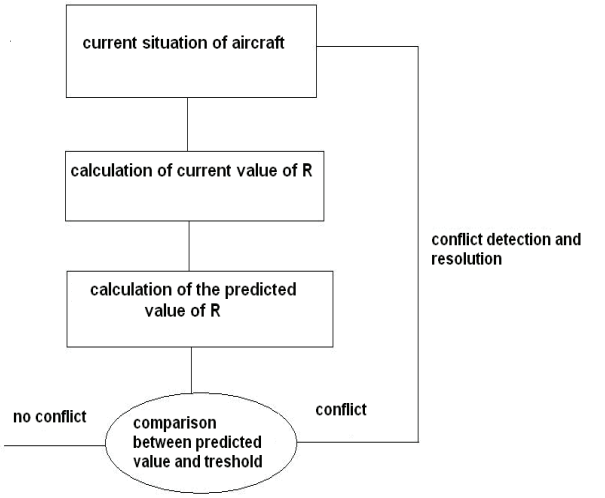


fig.1 Conflict detection algorithm

Conflict detection is entirely addressed to future moment of time- the correctness of the detection depends on the correctness of the prediction. Obviously, the existence of a conflict depends on the type and parameters of the chosen predicting algorithm. The conflict detection is a random process, determined by the probabilistic algorithm for prediction calculation. At the first step of the algorithm, the process of data receiving is also the random type process.

Analysis of the results of every type of error, provide the information about weight coefficient of evaluation, for each error. The optimization of conflict prediction algorithm is a multi-objective optimization.

Parameters of the optimization:

1)current value of R

2)time interval of the prediction

3)data innovation period (frequency of innovation)

4)measuring errors

- meteorological phenomena
- electro-magnetic environment

5)size of the Protected Zone

6)dynamic model of the aircraft

7)interference between the predicted position of the pair of aircrafts.

The quality of prediction of ($\vec{R}(x,y,z,t)$) affected by a lot of operators.

The aim of the present algorithm is to realize the most real estimation of R. For the given value of R, is it optimal or not, can be decided, according to the parameters, which determined the explicit situation. For every parameters is calculated the value of the target function and it must be maximum. By this way regarding the chosen parameter (criteria) the optimum is reached.

. f_i – local criteria

$f=(f_1,f_2,\ldots,f_m)$ – vector criteria

when m=1 – one-object task

$m \geq 2$ – multi-objective task

each solution –x – determined entirely the vector - f(x)

$$Y = f(X) = \{ y \in E^m \mid y = f(x), x \in X \} \quad (2)$$

All methods for optimization used in the algorithms are Pareto methods. In the multi-objective tasks every solution $x \in X$, is determined by its evaluation y=f(x), that's why the choice of the optimal solution is equal to the choice of a optimal evaluation, from the variety of all possible evaluations.

B)Probability state

Depending on the characteristics of criteria and the structures of varieties of possible solutions determined different types necessary and sufficient conditions for a specific solution to be optimal or not. All characteristics which prevent accurate conflict detection are called perturbation. In the context of the games theory "perturbations" are treated as enemies. The win of one player is a loss for other player. That type of game is antagonistic game with null sum.

$$H_{II}(s) = -H_I(s), s \in S$$
$$H_I(s) + H_{II}(s) = 0 \quad (3)$$

C) Min max method for game theory

The theorem of min max, declares that every antagonistic game has optimal strategy, called col point. The necessary and sufficient is:

$$\max_i \min_j a_{ij} = \min_j \max_i a_{ij} \quad (4)$$

Win function of the first player is mathematical estimation:

$$H(x.y) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_i y_j \quad (5)$$

For determining the optimal strategies is used col point and the methods of domination and linear programming. Min max method gives solution in any case, but this solution is not always optimal according to Pareto rules. The availability of solution in any case is an advantage for the method, for the actual problem, where it is an important matter. But the method has one dangerous aspect, there is no guarantee that the particular solution is optimal.

D) Medium risk criterion

The medium risk criterion is an approach to the problem, which treats all factors by the same way. Analyzing and evaluating the distance between a pair of aircrafts(R) and comparing it with the threshold the method gives decision is there a conflict or not.

R_o (:R_{ox}, R_{oy}, R_{oz}.)

$R_{пр}$ =R+n(t)

The following situation are possible, as a result of the process:

1) When $R_o < R_{min}$ ($R_y < R_{ymin}$ or $R_x < R_{xmin}$ or $R_z < R_{zmin}$) -$P(x_1)$- conflict possibility

D(detect)- probability of successful detection.-$D=P(R_o<R_{min})$

$P_1= P(x_1).D$

M(miss)- probability of missed detection

$M=P(R_0>R_{min})$

$P_2 = P(x_1).M$

2)When $R_o > R_{min}$($R_x > R_{xmin}$ and $R_y > R_{ymin}$ and $R_z > R_{zmin}$) - $P(x_0)$-probability of no conflict
R(reject)- probability of right reject detection
$R=P(R_0>R_{min})$
$P_3= P(x_0).R$
F(false)- probability of unnecessary alert
$F=P(R_0<R_{min})$
$P_4= P(x_0).F$
$P(x_1).D + P(x_1).M + P(x_0).R+ P(x_0).F =1$
The conditional risk, when R is evaluated, is combination of the probabilities of missed detection and unnecessary alert. P(R) –priori distribution of the probability calculated estimate value of R_o to be exact.
If to situation is given a price for the error, depending on the risk, so that right solutions have null price and falt solutions- max.
The medium error price is a mathematical estimation:.

$$\overline{K} = k_1 D.P(x_1) + k_2 M.P(x_1) + k_3 R.P(x_0) + k_4 F.P(x_0) \quad (6)$$

When k_1 =0 and k_3 =0

$\overline{K} = k_2 M.P(x_1) + k_4 F.P(x_0)$ - medium risk is a sum of error probability.
The medium risk criterion is one of the most common criteria. The exceptions of these criteria are the criterion of the ideal observer, criterion of the posterior probability and etc. For the criterion is needed output data for the characteristics of perturbation. That data is not always available, because of that the criterion is not reliable and appropriate for the specific problem.

E)Criterion of the ideal observer

All errors lead to same results. $k_2 = k_4 = 1$

and $\overline{K} = k_2 M.P(x_1) + k_4 F.P(x_0) - \min$ -error probability

$\overline{K} = k_1 D.P(x_1) + k_3 R.P(x_0) - \max$ - probability for right evaluation
Optimal by the criterion of the ideal observer is that method of evaluation of R, which provide min probability for error and at the same time max probability of right decision. The solution of the problem can be simplified by refusing to take into consideration the different result of the errors.

F) Criterion of the maximum likelihood.

$P(x_1) = P(x_0)$ - conflict probability and probability of no conflict are equal
$P_{zp} = M + F - \min$
Disadvantages of these approaches are the same as the ones of medium risk.

G)Multi-target (vector) optimization.

In the process of searching the best solution there is need to have a chance to change some of parameters, describing the state of the object. The managing parameters must be changed independently and promote varieties of the problem solution. Mathematical model of the object join all the parameters by a system of functions.

$$Y_i = f_i(x^u), i = 1, 2, ...n \quad (7)$$

Input parameters $\mathbf{x^u}$ include: x- vector of managing parameters and $\mathbf{x^c}$ – vector of constant parameters. Limitations of parameters form G_x – range from which the managing parameters can take values. Analysis of the obtained results –y, produce the selection of the optimal decision. Criterion of optimization is numerical presentation of estimate object state and it is used for comparing different states.

$$Q = Q(x) = Q(x_1, x_2, \ldots x_n) \quad (8)$$

Design and work of the conflict detection algorithm is appraised by a complex index.

$$Q_i (i = 1, 2, \ldots n1) \quad (9)$$

The optimal level of the index are achieved at different values of managing and design parameters. In general case, range of solutions- Q_i is open, but if some restrictions are made the range will become covered (closed) inside which will be situated the optimal solution. For every point in G_0 - field, can be found better solution, except the points on the line. These kind of solutions is called Pareto- optimal (non-improvable). The final solution of a problem of multi-objective optimization is solution from Pareto- optimal solution. The main problem is which point is the optimal one. Some methods are proposed: scalarization of vector criterion, generalized loss function, utility, iterative approach and etc.

H)Compromise optimization using loss function.

Deciding to choose one or another Pareto – optimal point for a final solution will provide some loss of absolute extreme Q_i^* of the target function $Q_i(x)$.

$$\Delta Q_i(x) = \frac{Q_i^* - Q_i(x)}{Q_i^*}, i = 1, 2, \ldots m \quad \text{-relative loss} \quad (10)$$

Generalized target function:

$$\phi(x) = \sum_{i=1}^{m} [\Delta Q_i(x)]^2 W_i = \sum_{i=1}^{m} [\frac{Q_i^* - Q_i(x)}{Q_i^*}] W_i \to \min_{x \in G_x}, \sum_{i=1}^{m} W_i = 1 \quad (11)$$

The optimal solution by some criteria is that solution, optimal by Pareto, which has minimal value, according to minimizing criteria and maximal, according to maximizing. Giving the weight of the every criterion, ensure that there is one optimal solution, based on the given priorities.
It is acceptable and possible the chosen solution to be very close to the optimal, but not to be optimal.
$\Delta Q_i = Q_i^* - Q_i(x)$ -min – diversion of the chosen solution from the optimal
In order to minimize the diversion for all criteria it is necessary they to be normalized. When the criteria functions are transformed it must be taken into account, that the transformations have to be started from a common point and same order of priorities. The transformations must be monotonous and inside the $[0;1]$.

$$0 \le \omega_i(q_i(x)) \le 1 \quad (12)$$

Using consistent analysis approach each acceptable solution $x \in A$ and $0 < \omega_i(x) < 1$ there is a vector $\rho \in \Phi^+$ which covers:

$$\rho = \langle \rho_i \rangle \in \Phi^+ = \langle \rho_i : \rho_i > \forall i \in I, \sum_{i \in I} \rho_i = 1 \rangle \quad (13)$$

The transfromation of the vector type criteria into scalar type can be made by different ways: linear, minimizing, maximizing etc..

As a result of the review is presented the following optimizing algor4ithm of conflict detection process.

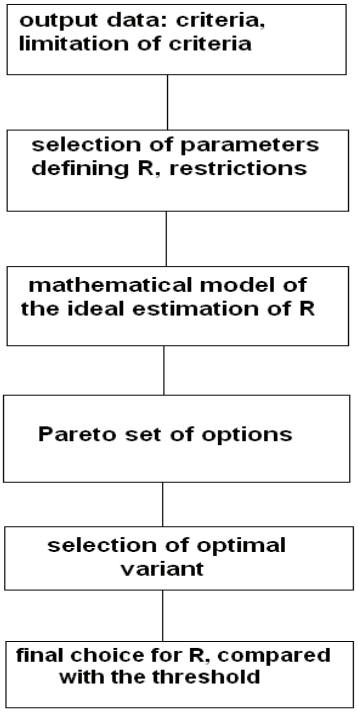


fig.2 Optimizing algorithm

Proposed algorithm uses:

$$q_{i\text{н}}(x) = \frac{q_i(x) - \min\limits_{x \in X} q_i(x)}{\max\limits_{x \in X} q_i(x) - \min\limits_{x \in X} q_i(x)} \text{ - simple normalization. (14)}$$

III. CONCLUSION

That survey is an attempt, analyzing some of the approaches for obtaining estimation of parameter (R), to find the best solution for the optimization of conflict detection process, taking into consideration the particularity of the problem. The proposed methods for optimization, which provide necessary and sufficient conditions for Pareto optimization, have a strong advantage. They are more effective, because of the performing the priorities better. Vector-valued optimizing criterion guaranteed bigger part of demanded requirements of criterion selection. Using of vector-valued criterion, which choose one solution from a variety of Pareto optimal solution, gives assurance that the final one is optimal and it is the closest one to the ideal solution. At the same time we can be sure that all the criteria are taken into account, because by this way can be processed all the available information. For our particular problem is very important to find just one solution and to be sure that it is the most non-shifted one from the real value (we compare calculated by the system R with real R in the air).

REFERENCES

1. Korobko,I.E.,"Radar systems and devices" Sofia , 2002
2. Lygeros, J., Prandini, M.,"Aircraft and Wheather Models for Probabilistic Collision Avoidance in Airtraffic Control"
3. Dowek, G., Geser, A., Munoz,G.,"Tactical conflict detection and resolution in a 3D airspace"
4. Hu, J., Lygeros, J., Prandini, M.,Sastry, Sh., "Aircraft conflict prediction and resolution using brownian motion"
5. Henk A.P. Blom, G.J. Bakker," Conflict probability and incrossing probability in air traffic management" 2002, Las Vegas
6. L.C.Yang, J.K.Kuchar, " Prototype Conflict Alerting system for free flight" Paper aiaa-97-0220,35th aiaa aerospace Sciences Meeting and Exhibition , Reno,NV, January 6-10,1997
7. L.C.Yang, J.K.Kuchar," A review of conflict detection and resolution modeling methods," IEEE Transactions on Intelligent Transportation Systems, vol.1, December, 2000
8. Paielli, R.A. and Erzberger ,H., "Conflict Probability Estimation for Free Flight", Journal of Guidance, Control and Dynamics, Vol.20 № 3, 1997
9. Eurocontrol standard document for radar surveillance in en-route airspace and major terminal areas., 1997, March
10. Korobko,I. „ Radio systems"Korvet press, 2003
11. Vorobev,N.N „Game theory for economists cybernetic"Leningrad University 1974
12. Petrov,P.G, Ivanov, H.C.,Yordanova-Dukova ,E.A., "Conflict detection and resolution performance"
13. V.V, Podinovskii, V.D. Nogin ,"Solutions of mul6ti-objective problems, optimal by Pareto", Science,Moscow 1982

Normal Point Density Input Influence over a Finite One-Dimensional SOM

Ivo R. Draganov[1]

***Abstract* – In this paper an analysis is presented concerning the asymptotic state of the one-dimensional self-organizing map (SOM) with finite grid in the case of normal point distribution input. The SOM distortion measure is analyzed with its optimum found approximately. The results obtained are considered useful enough in wide variety of practical cases where fine tuning of the SOM is needed.**

***Keywords* – Self-Organizing Map, Normal Point Distribution, Distortion Measure.**

I. INTRODUCTION

It is well known fact that the area allocated for storing the most important feature set inside a self-organizing map (SOM) is proportional to the frequency of occurrence of that very same feature in the observations [1]. As the SOM structure tends to become very complex in the most of its real case applications often the magnification factor is used to describe heaping of feature vectors. It is simply the inverse of the point density around each neuron representing a cluster.

So far an investigation of the point density for the linear map is led in the presence of a very large number of codebook vectors over a finite area [2], [3]. It is revealed that the asymptotic point density is proportional to the probability of a certain feature vector occurring raised to some exponent depending of the number of neighbors including the winning neuron and some scalar factor.

In any case the initial neighbor function width may vary largely during the training process starting with huge values and ending with zero-order topology case – no neighbors except the winner are present. This boundary case is undesired since the learning process no longer maintains the order of the codebook vectors. The approximation accuracy of the probability of occurring for a feature and the minimum stability of ordering demanding more neighbor interactions are the two aspects to be balanced.

If we have no neighbors around the winner a simple scalar quantization case occurs. Then the power of the asymptotic function for the point density decreases, according to [3] bellow 1/3. Getting this power to higher values incrementally by trial and error approach seems a good solution but the following tendencies should be considered. If we try to approach the Bayesian classifier, i.e. to find the optimal classification border and the density functions of adjacent clusters are close to each other the latter could be replaced with any other pair of monotonic functions of densities. In such a case the practical SOM application is adaptable to simplification. The other important property is that when feature dimensionality is increased in the order of hundreds of components per vector the power is close to 1 [3]. Similar research on the change of this power is done in [4] when the neighbor function is Gaussian kernel and its normalized second moment is independent variable. The resulting range for the power value in this case is from 1/3 to 2/3. Analogous results are presented in [5].

What has not been investigated so far is the influence of the normal point density of the input over the asymptotic state of a finite one-dimensional SOM and its distortion measure. In [6] a typical practical challenge is given which can be solved by the approach presented here. In part 2 such analysis is presented and in part 3 some computational results are given. In part 4 a conclusion is made.

[1]Ivo R. Draganov is with the Faculty of Telecommunications, 8 Kliment Ohridski Blvd., 1000 Sofia, Bulgaria, E-mail: idraganov@tu-sofia.bg

II. SOM ANALYSIS WITH NORMAL POINT DENSITY INPUT

A. Asymptotic State of the One-Dimensional Finite-Grid SOM

Let one-dimensional feature space of x is considered. For our analysis to be correct the following assumptions should be granted: the number of points (feature vectors) must be large enough (e.g. by criteria given in [1]) and they must be stochastic variables so their differential probability for each cluster they fall into, i.e. the probability density $p(x)$ could be defined. The codebook vectors m_i usually form regular optimal configuration and thus can not be stochastic. Their number is typically low in any cluster as well.

Let suppose m_i and m_{i+1} are two neighboring points. A way of defining the point density is as $(m_{i+1} - m_i)^{-1}$ but it does not cover the samples around the boundaries of the clusters for which this density does not have meaning. So a better way of defining it is as the inverse of the width of the Voronoi set $[(m_{i+1} - m_i)/2]^{-1}$.

The input consists of samples $x(t) \in \Re, t = 0,1,2,...$ while the codebook is represented by $m_i(t) \in \Re, t = 0,1,2,..., i = 1,..., k.$ It is assumed $0 \le x(t) \le 1$. The one-dimensional SOM algorithm with at least one neighbor at each side according to [1] is given by:

$$m_i(t+1) = m_i(t) + \varepsilon(t)[x(t) - m_i(t)], \quad \text{for i} \in \mathrm{N_c},$$

$$m_i(t+1) = m_i(t) \quad \text{for} \quad \mathrm{i} \notin \mathrm{N_c},$$
$$c = \arg\min_{\mathrm{i}} \{|x(t) - m_i(t)|\}, \tag{1}$$
$$N_c = \{\max(l, c-1), c, \min(k, c+1)\},$$

where N_c is the neighbor set around node c and $\varepsilon(l)$ is the learning-rate factor. The Voronoi set V_i around m_i is defined as:

$$V_i = \left[\frac{m_{i-1}+m_i}{2}, \frac{m_i+m_{i+1}}{2}\right], V_1 = \left[0, \frac{m_1+m_2}{2}\right],$$
$$\mathrm{V_k} = \left[\frac{m_{k-1}+m_k}{2}, 1\right], \quad \text{for} \quad 1 < \mathrm{i} < \mathrm{k}, \tag{2}$$
$$U_i = V_{i-1} \cup V_i \cup V_{i+1}, U_1 = V_1 \cup V_2,$$
$$\mathrm{U_k} = V_{k-1} \cup V_k, \quad \text{for} \quad 1 < \mathrm{i} < \mathrm{k}.$$

In this case U_i is the set of such $x(t)$ which provoke changes in $m_i(t)$ during one learning step. Following (1) and (2) we get to the well known stationary equilibrium for m_i coinciding for the general case [1]:

$$m_i = E\{x \mid x \in U_i\}, \forall \mathrm{i}. \tag{3}$$

In other words every m_i becomes centroid of the probability mass for each U_i and then for 2 < i < (k-1) the limits for U_i are:

$$A_i = \frac{1}{2}(m_{i-2} + m_{i-1}),$$
$$B_i = \frac{1}{2}(m_{i+1} + m_{i+2}). \tag{4}$$

For i =1 and i = 2, A_i = 0, and for $i = k - 1$ and $i = k$, = 1.

The case investigated here concerns input data with the following distribution:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-x_0)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}, \tag{5}$$

where x_0 = 0 is set since $0 \le x \le 1$ and it is enough to examine only the right part of the Gaussian curve – the results for the left one could be obtained from the symmetry properties.

As (5) is too complex to be used in finding the centroids of the probability masses, Taylor series are used instead to the second order for simplicity:

$$p(x) = \sum_{i=0}^{\infty} \frac{p^{(i)}(0)}{i!} x^i \approx p(0) + p'(0) + \frac{p''(0)}{2} x^2, \tag{6}$$

$$p'(x) = -\frac{x}{\sqrt{2\pi}\sigma^3} e^{-\frac{x^2}{2\sigma^2}}, \tag{7}$$

$$p''(x) = \frac{1}{\sqrt{2\pi}\sigma}\left(\frac{x^2}{\sigma^4} - 1\right) e^{-\frac{x^2}{2\sigma^2}}, \tag{8}$$

and substituting (7) and (8) into (6) the approximate distribution is:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma}\left(1 - \frac{1}{2}x^2\right). \tag{9}$$

The stationary values of the m_i are defined by the set of nonlinear equations obtained by transition to analog domain:

$$m_i = E\{x \mid x \in U_i\} = \frac{\int_{A_i}^{B_i} x p(x) dx}{\int_{A_i}^{B_i} p(x) dx} = \tag{10}$$
$$= \frac{24\sqrt{2\pi}\sigma(B_i^2 - A_i^2) - 3c(B_i^4 - A_i^4)}{48\sqrt{2\pi}\sigma(B_i - A_i) - 8(B_i^3 - A_i^3)}, \forall \mathrm{i}$$

The solution of such a set could be done in the following way [1]. Let us have:

$$z = [m_1, m_2, ..., m_k]^T. \tag{11}$$

The equation to be solved is of the form:

$$z = f(z). \tag{12}$$

The first approximation then is $z^{(0)}$ and every subsequent approximation for the root is found by:

$$z^{(s+1)} = f(z^{(s)}). \tag{13}$$

The increase of the computational complexity using (13) is nonlinear with the number of grid points growing. So in [1] an expedient way of calculating the root is given and it consists of defining the point density q_i around m_i as the inverse of the length of the Voronoi set – $q_i = [(m_{i+1} - m_{i-1})/2]^{-1}$. As a result of that q_i can be expressed in the form $const.[p(m_i)]^{\alpha}$. Then passing from m_i to m_j it is true:

$$\alpha = \frac{\log(m_{i+1} - m_{i-1}) - \log(m_{j+1} - m_{j-1})}{\log[p(m_j)] - \log[p(m_i)]}. \tag{14}$$

For improved accuracy more values of the m_i are needed as we shall see in the next section.

B. Finding the One-Dimensional SOM Distortion Measure with Finite Grids

The objective function of the SOM is given by [1]:

$$E = \sum_i \sum_j \int_{x \in V_i} h_{ij} \|x - m_j\|^2 p(x) dx, \tag{15}$$

where V_i is the Voronoi set around m_i and h_{ij} is defined as:

$$h_{ij} = \begin{cases} 1, & \text{if} \quad |\mathrm{i} - \mathrm{j}| < 2 \\ 0, & \text{otherwise} \end{cases} \tag{16}$$

and i and j run over all the values defining h_{ij}.

Then (15) becomes:

$$E = \sum_i \sum_j \int_{C_i}^{D_i} (x - m_j)^2 p(x) dx =$$
$$= \sum_i \sum_j \frac{1}{\sqrt{2\pi}\sigma}\Big[-\frac{(D_i^5 - C_i^5)}{10} + \tag{17}$$
$$+ \frac{m_j(D_i^4 - C_i^4)}{4} + \frac{(2 - m_j^2)(D_i^3 - C_i^3)}{6} -$$
$$- m_j(D_i^2 - C_i^2) + m_j^2(D_i - C_i) \Big] = \sum_i \sum_j E_{ij}.$$

where N_i is defined in (1) and the borders C_i and D_i of the Voronoi set V_i are:

$$\begin{aligned} C_1 &= 0, \\ C_i &= \frac{m_{i-1} + m_i}{2} \quad \text{for} \quad 2 \le \mathrm{i} \le \mathrm{k}, \\ \mathrm{D_i} &= \frac{m_i + m_{i+1}}{2} \quad \text{for} \quad 1 \le \mathrm{i} \le \mathrm{k}\text{-}1, \\ \mathrm{D_k} &= 1. \end{aligned} \tag{18}$$

From (17) and (18) the gradient of E could be found as:

$$\frac{\partial E}{\partial m_i} = \frac{\partial}{\partial m_i} \sum_{j \in N_{i-1}} E_{i-1,j} + \frac{\partial}{\partial m_i} \sum_{j \in N_i} E_{ij} + \\ + \frac{\partial}{\partial m_i} \sum_{j \in N_{i+1}} E_{i+1,j} \quad , \tag{19}$$

but here will not be given due to its long form and since the calculations are trivial.

To obtain minimal distortion for the SOM optimal values for m_i should be found. A way of doing this is using the gradient-descent method according to:

$$m_i(t+1) = m_i(t) - \lambda(t)(\partial E / \partial m_i \mid_t), \tag{20}$$

where the scalar factor $\lambda(t)$ is typically in the range from 0.001 to 0.01 but even larger values of the order of 10 are tolerable [1]. Here E is of the fifth order, so at least one minimum can be found. The tendency is with the growth of $\lambda(t)$ the found minimum to become global not depending of the initial values of m_i.

III. Computational Results

As a simulation environment we use Matlab® R2009B over MS® Windows® XP® Pro SP3.

First α from (14) is found for different number of grid points. The more m_i are used the more accurate are the results. For $i = 4$ and $j = k - 3$ assuring negligible border effects 10, 25, 50, and 100 grid points are used. The same experiment is done with three other probability functions – linear, quadratic and linear-quadratic in [1], so here a direct comparison can be made. The results are given in Table I.

TABLE I
THE DERIVED ALPHA FOR FOUR DIFFERENT DISTRIBUTIONS OF THE INPUT

Grid points	Exponent α			
	Linear, [1]	Quadratic, [1]	Linear-quadratic, [1]	Normal
10	0.5831	0.5845	0.5845	0.5850
25	0.5976	0.5982	0.5978	0.5980
50	0.5987	0.5991	0.5987	0.5980
100	0.5991	0.5994	0.5990	0.5992

It is clearly seen from Table 1 that even for such different distributions as linear and quadratic from one side and the normal on the other, the resulting α is almost constant and very close to 0.6. Graphically the results are given on Fig. 1.

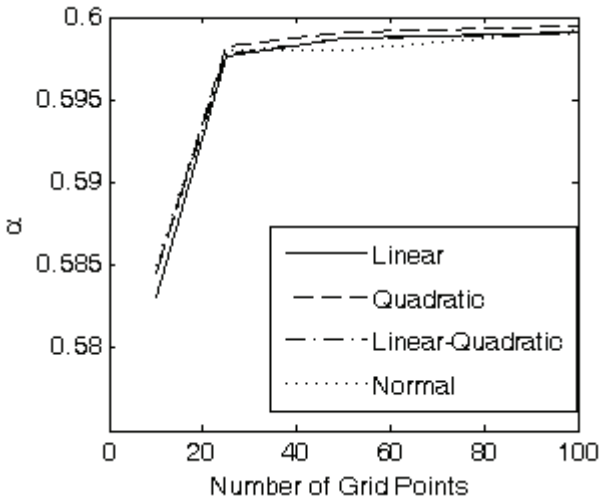


Fig. 1. Theoretically derived α as a function of the number of grid points for four different distributions of the input for the SOM

After computing the optimal m_i using (19) and (20) for the same number of grid points – from 10 to 100, it is possible to estimate α. Now when the minimal distortions have been assured this is done experimentally over real input data with quadratic and normal distribution around the winners. The results are given in Table II.

TABLE II
EXPERIMENTALLY ESTIMATED ALPHA FOR TWO DIFFERENT DISTRIBUTIONS OF THE INPUT AT THE MINIMAL DISTORTION MEASURE

Grid points	Exponent α	
	Quadratic, [1]	Normal
10	0.3281	0.2989
25	0.3331	0.3330
50	0.3333	0.3331
100	0.3331	0.3330

Obviously the values obtained here again do not depend on the number of grids and again are too close one to another despite the different form of the input distributions. The significant difference is with the results obtained by the theoretical derivations from (14). Now when we have the optimal m_i found it is seen that the exponent of the approximated state of the SOM is virtually equal to 1/3. This is actually a case coinciding with the optimal vector quantization [1] unlike the case of $\alpha = 0.6$ and confirms the correctness of the optimal m_i calculated. Graphically the results from Table II are given in Fig. 2.

In both cases, for $\alpha = 0.6$ and $\alpha = 0.3$, m_i can be considered as forming an elastic network expanding into the input feature space following the order of appearance of the consecutive

samples. The fact that the presented stochastic approximation (1) – (14) can not ensure the optimal case when $\alpha = 0.3$ does not mean that it is useless – actually saving the iterations from (20) and the preliminary fixation of m_i from (15)-(19) means considerable saving of computation time which is important in wide range of practical cases.

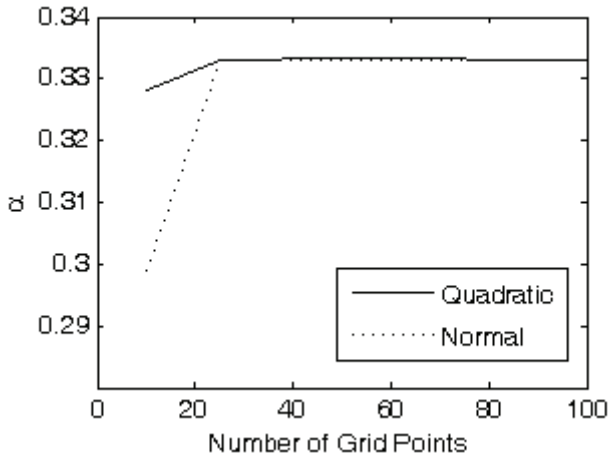


Fig. 2. Experimentally derived α as a function of the number of grid points for four different distributions of the input for the SOM at minimal distortion measure

One practical way of enhancing the work of a SOM is taking higher value for α, not only over 1 but even greater than 10, in the cases when the expanding network of nodes does not occupy at least half of the feature manifold. This step should be preceded by passing the training sample set in random order considerable number of times and only then should be used as an utter measure. In other cases proper action could be introducing normal distributed data into the area where the expansion is not wide enough and then to substitute it with the original one the number of epochs needed. In such situations the balance between the computational costs determined from (20) and the accuracy with its maximum defined by solving (19) should be carefully achieved.

IV. Conclusion

In this paper an approach for finding the stationary positions of the nodes of one-dimensional SOM has been presented in the case of normal density point input. The results are precise enough taking the advantage of very fast computation. Furthermore the distortion measure of the SOM using finite grid is calculated in the general case and it is shown that the positions of the nodes could be optimized iteratively at the cost of more computation time but which leads to results close to those from the optimal vector quantization.

The results achieved prove the correctness of the suggested approach which is considered useful in a large number of practical cases where the input data poses normal point density distribution.

Acknowledgement

This paper was supported by the National Fund for Scientific Research of the Bulgarian Ministry of Education and Science (Contract - BY-I-302/2007).

References

[1] T. Kohonen, *Self-Organizing Maps*, 3 ed., New York, Springer, 2001.

[2] H. Ritter and K. Schulten, “On the Stationary State of Kohonen's Self-Organizing Sensory Mapping”, Biological Cybernetics, Vol. 54, Issue 2, pp. 99-106, June 1986.

[3] H. Ritter, “Asymptotic Level Density for a Class of Vector Quantization Processes”, IEEE Transactions on Neural Networks, No. 2, pp. 173-175, 1991.

[4] D. Dersch and P. Tavan, “Asymptotic Level Density in Topological Feature Maps”, IEEE Transactions on Neural Networks, Vol. 6, No. 1, pp. 230-236, 1995.

[5] S. Luttrell, “Code Vector Density in Topographic Mappings: Scalar Case”, IEEE Transactions on Neural Networks, Vol. 2, No.4, pp. 427-436, 1991.

[6] R. Kountchev, D. Bojchev and D. Dobrev, “An Analysis of the Possibilities for Using Lattices for the Interpolation of the Exact Signal Vector Position of QAM Constellations”, In Proceedings of ICEST’2009, Veliko Tarnovo, Bulgaria, Vol.1, pp. 71-75, June 25-27, 2009.

Automatic Report Producing from Results of Optimization Tasks in Batch Signal Processing

Slavy G. Mihov[1]

Abstract – **In this paper is presented a program solution (created in Matlab) designed to execute similar computations over a set of input data files, store results from every single processing (text, graphics, execution time, etc.) and produce a perspicuous report (*.doc, *.pdf) with the obtained results included. This tool performs batch signal processing and removes the necessity of human interaction until all processing is over and the report is automatically generated, which tends to be very helpful for long time intervals of calculations.**

Keywords – **generate automatic report, batch processing.**

I. Introduction

Optimization tasks (particularly in signal processing) are a common problem with massive theory built for the purpose. Many times signal processing is complicated, time consuming and computationally expensive, especially in performing monotonous operations over packs of numerous signal files - batch processing.

In previous works, concerning algorithm development and evaluation for processing biological signals and estimating the results, a common use case was experimenting with various parameters of the processing algorithm to seek optimal results in some meaning. In [1] is investigated a wavelet denoising procedure for the purposes of hearing-aid signal processing. Particular database of recordings of human speech captured in various noisy environments (in several signal-to-noise ratios) is subjected to processing with a denoising algorithm. The wavelet denoising algorithm used, has numerous of parameters which adjust its performance, thus seeking the optimal algorithm profile turns in an optimization task with numerous independent arguments. An evaluation metric is used to estimate the quality of processing every single database recording. In this investigation, plausible results are obtained not until processing a large enough set of test recordings, which could have never been done (easily) without using a tool for batch signal processing and automatic report generation, generalizing the obtained results.

In the same manner [2] develops an algorithm for distinguishing atypical from typical heart beats in ECG signal for the purposes of automatic feature recognition. The evaluated algorithm uses an optimal linear transformation of the available leads in a multichannel ECG signal to derive a new lead with enhanced atypical beats. The pursue of this optimality is again an optimization task searching for a single set of coefficients for the linear transformation that produce the best result (according the optimization criterion). Evaluating the algorithm and proving its functionality [2, 3] is again subjected to processing large databases of different format ECG recordings. This could have also been a tedious task without the use of a tool for batch processing and automatically accumulating the results in the form of a report for the benefit of the human to estimate the processing.

The tool used for automatic report generation has improved more and more with every successive project. It gradually evolved from a couple of script lines, becoming more and more complex, to turn to a sophisticated and a irreplaceable in algorithm development and evaluation tool.

II. Tool for Batch Processing

For the purposes of investigation and algorithm development [1, 3] is created an environment for processing and visualizing signals in the form of input data. This development tool is supposed to greatly facilitate and automate the minor, and insignificant for the processing operations.

The primary goal of the batch processing tool is to facilitate gaining and storing in a convenient form the results from processing numerous test datasets (ECG, human speech, etc. recordings). As a result from execution, in the working directory are created groups of files containing information for the processing. For instance, for each test recording is created a text log file (*.log) containing statistic for: computation time (begin, end, continuance), optimization criterion used, optimal algorithm parameter set obtained and so on. The input signals and output results are stored in the form of diagrams in vector graphic format (*.eps, *.emf), convenient to use in documentation and consecutive processing. Amongst the valuable features of the tool for batch processing is the ability to automatically produce perspicuous report (*.pdf, *.doc) including data for all processing, parameters of computations and results in the form of signal graphics, text statistics and tables. This feature tends to be very convenient and saves much time in browsing the suitable data recordings in large databases.

A. Structure of the Tool

The tool is developed as a Matlab script program, which generally tends to be platform independent, executed by Matlab Common Runtime (MCR). However, due to the fact that some parts of its functionality is based on MS Windows specific program technologies (COM/OLE), the tool occurs to be platform dependent.

[1]Slavy G. Mihov is with the Faculty of Electronics, Technical University – Sofia, blv. Kliment Ohridski 8, 1000 Sofia, Bulgaria, e-mail: smihov@tu-sofia.bg

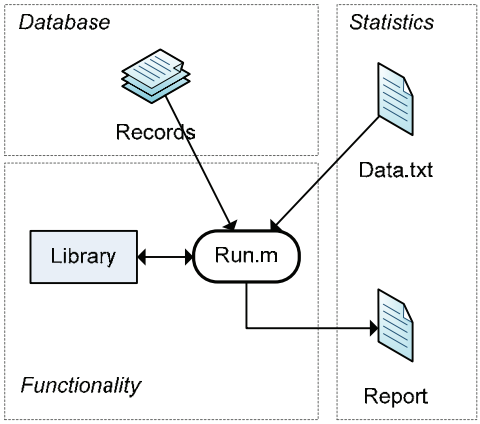


Fig. 1. Tool Structure

The program for batch processing (Run.m) consists of separate functional modules formed in a couple of script files (*.m). Figure 1 shows the general structure of the development tool and the connections between its five major parts:

Run.m is the core module of the tool for batch signal processing, formed as a sequential procedure consisting several processing steps. Its structure can be seen on Fig. 2.

Library is a set of library functions (read/write text/data files, functions for signal pre-processing, etc.), which functionality is used by the main program.

Records denotes the set of input data files (*.tmp) pending for processing. In [2, 3] these are multichannel (4, 8, 12) ECG recordings and in [1, 4] these are human speech recordings (clean and contaminated with noise in several SNRs).

Data.txt is a text file which contains specifications for the batch processing. It is used by the main program (Run.m) to define parameters such as: list of input data file names, paths, optimization criteria, constraints, optimization intervals, etc.

Report is the final report in the form of a MS Word document (*.doc) and a Portable Document Format (*.pdf). This report is automatically generated after all batch processing is completed and contains results in graphical, text and table form for convenience of the human reader.

Table 1 summarizes all files involved in the batch processing and automatic report generation. It lists the files in the work directory by extension, either used as an input or created during the execution of the optimization task.

TABLE I
FILES USED IN BATCH PROCESSING

File	Purpose	Description
*.tmp	input data	Miscellaneous data
*.log	execution log	Text log
*.eps / *.emf	output diagram	Graphical results
*.doc	output report	MS Word document
*.pdf	output	Portable Document

After the parameters for signal processing and optimization have been set in the text file Data.txt, the batch computations can be started by running subroutine Run.m. The last one organizes all calculations in batch mode, thus eliminates the necessity of human interaction and control until finalizing all operations which can be quite time consuming. As final result,

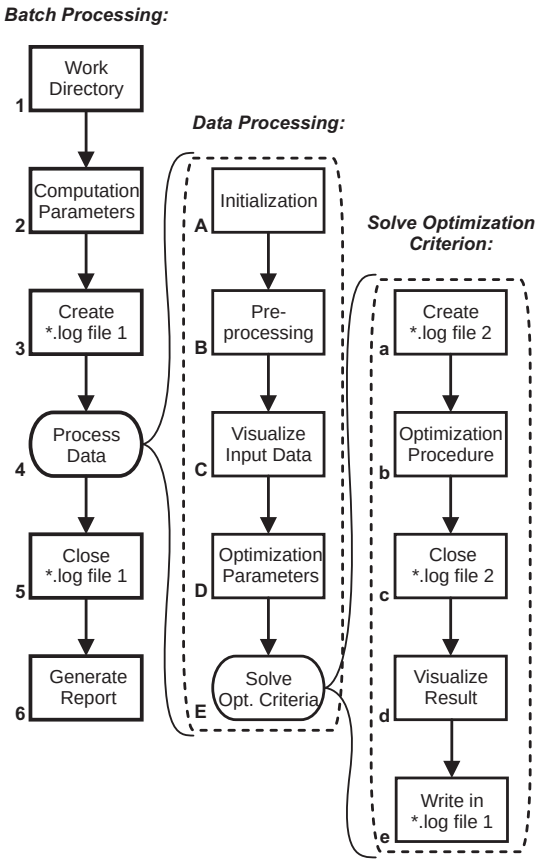


Fig. 2. Batch Execution Flow

in the working directory are created numerous files holding text and graphical information for the input data (signals) and the obtained results (Report).

B. Tools Specifics

The tool for batch processing and automatic report generation (Run.m) does not need any Graphical User Interface (GUI) to operate, provided that all necessary files (Records, Data.txt) are present. It can be run in a console, deployed to a remote machine to use its computation resources and easily be scheduled for execution in time. However, setting the parameters of the batch processing task (Data.txt) can be more convenient if this is done automatically in a GUI, which is already being developed for the optimization task in [3].

The essential work in organizing the batch processing is done by the subroutine Run.m, which main stages are shown on Fig. 2 as a nested sequence of operations. The process flow is relatively straightforward consisting the following steps:

1. Create working directory (.\Temp\) or clear already existing one to prepare it for storing the temporary work files and processing results. In case an old directory already exists, this suggests that batch processing has been done before. Its contents is being archived (*.rar) with the current date/time in order to preserve these previous results for potential use later.

```
%% Create Temp directory
if( ~exist('.\Temp', 'dir') )
    mkdir('.\Temp');
else
```

```
    disp('Backup existing results...');
    command = sprintf('rar a "Temp (%s).rar" Temp',
datestr(now, 'dd mmmm yyyy, HH.MM.SS'));
    [~, result] = system(command);
    disp(result);
    delete('.\Temp\*.eps');
    delete('.\Temp\*.emf');
    delete('.\Temp\*.log');
end
```

2. Load parameters for batch computations, from external text file Data.txt – file names, intervals for optimization, criteria for optimization, procedures for pre-processing (like subroutines for removing zero line shift, tremor artifacts and powerline interference in ECG) and so on. The text file Data.txt is being parsed and all input parameters are placed in the vector variable *param*.

```
%% Read record file names and intervals
fileName = '.\Data.txt';
fid = fopen(fileName, 'r');
textscan(fid, '%s %s %s %s %s', 1); % bypass line 1
param = textscan(fid, '%s %d %d %d %d');
fclose(fid);
```

3. Create master *.log file to store in text form the results of the performed computations; also starts a global timer to track execution time.

4. Successive iterative processing of every single input data file from the list in Data.txt.

5. Close master *.log file.

6. Generate automatic Report for all signal processing (Report.doc, Report.pdf) – document containing results and statistics in text, table and graphical for the performed algorithm optimization, obtained set of optimal parameters, input signals and output data.

The iterative processing of the succession of input recordings (step 4) consists of the following major stages:

A. Initialization of constants, loading input signal data from Records (*.tmp) and parameters of the optimization task.

```
%% Read data file
    fileName = strcat(char(param{1}(index)));
    [~, name] = fileparts(fileName);
    fid = fopen( fileName, 'r' );
    data = fread( fid, 'int16' );
    fclose(fid);

    %% Set typ and atyp intervals
    typ1 = param{2}(index);
    typ2 = param{3}(index);
    atyp1 = param{4}(index);
    atyp2 = param{5}(index);

    %% Initialize constants
    t = 0: 1/Q: N/Q; % Scale in sampling period, [s]

    %% Signal channels
    channels = reshape(data, N, numCh);
```

B. Pre-processing input data before starting the main optimization task. For ECG records [3, 5] these can be any sophisticated algorithms for eliminating zero line drift, tremor artifacts, powerline interference, etc. formed as program procedures in a Library for convenient use. Similarly, when processing human speech records, these can be any adequate algorithms for filtering, denoising, resampling and so on.

C. Visualizing input data and storing the diagrams of the signals in vector graphic format (*.eps, *.emf).

```
%% Plot signal leads
    f1 = figure('visible','off');
    plot(   t(axe1), S1(axe1), 'k', ...
            t(axe1), S2(axe1), 'g', ...
            t(axe1), S3(axe1), 'r' );

    print( f4, '-depsc2', '-tiff', '-r2400',
sprintf('.\\Temp\\Record-%sd.eps', name) );

    print( f4, '-dmeta', '-r2400',
sprintf('.\\Temp\\Record-%sd.emf', name) );
```

D. Define optimization parameters for the input signals and prepare constrained task for minimization – set intervals, criteria and constraints.

E. Perform optimization iteratively for every single optimization criterion of the processing algorithm, create logs for execution and store plots in vector graphic formats (*.eps, *.emf) in the working directory.

```
%% Optimization
    switch criterion
        case 1
          obj = @(x)Objective1(x, LTyp, LAtyp);
        case 2
          obj = @(x)Objective2(x, LTyp, LAtyp);
        case 3
          obj = @(x)Objective3(x, LTyp, LAtyp);
        otherwise
          disp('Unexpected criterion!');
    end

    options = optimset('Display', 'iter',
'Algorithm', 'active-set', 'MaxFunEvals', 1600);
    [x, fval] = fmincon(obj, [0 0 0 0 0 0 0 0], [],
[], [], [], [], [], @Constraint1, options);
```

III. Subroutine for Generating Report

Probably the most interesting part of the tool for batch signal processing and storing results for the performed optimizations is the subroutine for automatic report generation. It is formed as a separate subroutine, which can be invoked independently and it produces the report based on whatever input text and graphic files are present in the directory given. It consists of 4 sequential steps and below is shown its listing code.

The first step is to **Create a MS Word document** and add some text, graphics and formatting using VBA (Visual Basic for Applications) script commands invoked directly from Matlab environment [6]. Matlab's support for VBA is utilized by the command *actxserver()*, which creates a local OLE Automation server, and returns a handle to the default interface of a COM server with the programmatic identifier of MS Word.

The next steps are trivial and again use VBA commands to **Save**, **Print** (in *.pdf format) and **Close** the so created report document.

```
%% Generate report from *.eps/emf results.
function GenerateReport(directory, extension)

  filter = ['*' extension];

  % write to Word document
  Doc = actxserver('Word.Application');
  MS = invoke(Doc.Documents,'Add');

  set(Doc.Selection.Font, 'Name','Arial','Size',20);
  set(Doc.Selection.ParagraphFormat, 'Alignment',1);
  invoke(Doc.Selection, 'TypeText', 'Auto Report');
```

```
  invoke(Doc.Selection, 'TypeParagraph');

  files = dir(fullfile(directory, filter));
  for f = 1:length(files)
    fname = files(f).name;
    [~, name] = fileparts(fname);

    %Put figure in document
    invoke(Doc.Selection, 'TypeParagraph');
    graphicFile = [directory, sep, name, extension];
    invoke(Doc.Selection.InlineShapes, 'AddPicture',
graphicFile);
  end

  %Save word file
  word_name = 'Report (GUI).doc';
  full_name = [directory sep word_name];
  invoke(MS, 'SaveAs', full_name);

  % print report
  set(Doc, 'ActivePrinter', 'Adobe PDF');
  invoke(MS, 'PrintOut');

  % close document
  invoke(MS, 'Close');
  invoke(Doc, 'Quit');
  delete(Doc);

end
```

IV. Practical Examples

Using the presented tool for batch signal processing are accomplished several experiments in former projects, involving evaluating an algorithm of interest with a large set of input data files, estimating its performance and optimizing parameters of computation. For example, in the development of the algorithm for optimal linear transformation of multiple ECG leads [2, 3] is done automatic processing of 250 12-channel ECG recordings from database CSE (Common Standards for quantitative Electrocardiography) [7], task large enough to be out of scope of manual processing (thus no parallel between automatic and manual execution, in means of accelerating computation time, can be made).

On Fig. 3 is shown a sample page from a report (consisting of over 50 pages) automatically generated after processing the set of ECG recordings.

V. Additional Remarks

The substantial reduction of the time needed for manual processing, achieved by harnessing such specialized procedures as the one presented here, opens potentials for development and utilization of automatic tools for algorithms test, optimization and development. Such a technique is quite successfully adopted in the investigation the practical capabilities of wavelet transform analysis in speech signal denoising [1] and in the development of a novel algorithm for enhancing atypical heart beats in multichannel ECG signals [2, 3].

VI. Conclusion

In this paper is presented a program solution (created in Matlab) designed to execute similar computations over a set of input data files, store results from every single processing (text, graphics, execution time, etc.) and produce a perspicuous report (*.doc, *.pdf) with the obtained results included. This tool performs batch signal processing and removes the necessity of human interaction until all processing is over and the report is automatically generated, which tends to be very helpful for long time intervals of calculations. The method has already been used in two different projects, particularly for solving optimization tasks.

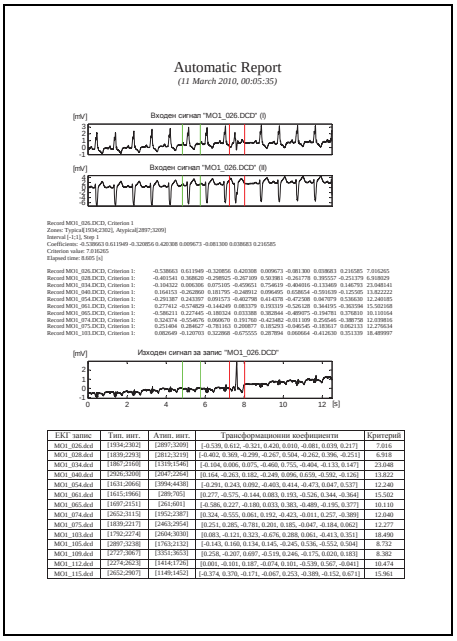


Fig. 3. Example Report Layout

References

[1] S. Mihov, R. Ivanov, A. Popov (2009). *Denoising Speech Signals by Wavelet Transform*. The 18th International Scientific and Applied Science Conference ELECTRONICS ET-2009, B. 1, ISSN 1313-1842, Sozopol, Bulgaria, June 14-17, pp. 69-72.

[2] S. Mihov, Ch. Levkov, G. Mihov (2009). *Algorithm For Optimal Linear Transformation of 4 Holter Leads for Emphasizing Difference Between Typical and Atypical QRS Complexes*. XLIV International Conference ICEST-2009, B. 1, Veliko Tarnovo, Bulgaria, June 25-27, pp. 403-406, 2009.

[3] Ch. Levkov, S. Mihov, G. Mihov (2009). *Algorithm for Optimal Linear Transformation of 12 Standard Leads for Emphasizing Difference between Typical and Atypical QRS Complexes*. The 18th International Conference ELECTRONICS ET-2009, B. 1, ISSN 1313-1842, Sozopol, Bulgaria, June 14-17, pp. 20-23.

[4] S. Mihov, D. Doychev, R. Ivanov (2009). *Practical Investigation of Specific Types of Noise Signals for the Purpose of Suppression in Hearing-Aid Devices.* XLIV International Conference ICEST-2009, B. 1, Veliko Tarnovo, Bulgaria, June 25-27, pp. 399-402, 2009.

[5] Ch. Levkov, S. Mihov. *Multilead signal preprocessing by linear transformation to derive an ECG lead where the atypical beats are enhanced: Matlab implementation.* Proceedings of TU-Sofia, Vol. 58, Book 2, pp. 24-30, 2008.

[6] http://www.mathworks.com/matlabcentral/

[7] Journal of the American College of Cardiology, Volume 10, p.1313-1321, 1987

Sound Insulation Measurement by Impulse Method

Dejan Ćirić[1], Borislav B. Budisavljević[2] and Aleksandar Pantić[3]

***Abstract* – One of the most important features of building construction elements from acoustical point of view is sound insulation. There are different alternatives how to measure this quantity. Generally, it could be measured in special laboratories or in situ where the elements and walls are mounted. Also, there are different methods for its measurements: classical method using stationary broad-band noise, and a new method based on impulse response measurement.**

This paper analyses some aspects of the application of a new impulse method for the measurement of sound insulation both in laboratory and in situ. The sound insulation determination is based on the measurements of impulse responses both in a source and receiver room. For the processing of these responses, the software module is developed. The influence of some parameters of the impulse method is investigated, such as the repetition and spatial standard deviation, signal to noise ratio, etc. As the illustration of the impulse method application, the results for sound insulation measurements of several partitions in the laboratory, and of a wall with the ventilation duct in situ are presented.

***Keywords* –Acoustical measurements, Impulse method, Sound insulation, Standard deviation.**

I. INTRODUCTION

Sound insulation as one of the main acoustical descriptors of a building construction element can be measured in a specialized laboratory, but also *in situ* where the element is built in. Regarding the method, two measurement methods have been used for this measurement, and both of them are standardized [1], [2]. The classical method based on stationary broad-band random noise is usually applied for laboratory measurements. On the other hand, the recently developed impulse method can be used for both laboratory and field measurements of sound insulation [3].

Although this new impulse method is standardized and it has been applied relatively often in practice, there is hardly data in the literature presenting the results of this method [4]. Thus, the implementation of the impulse method for sound insulation measurement is investigated here. For that purpose, the measurements were carried out in both specialized laboratory of IMS Institute, and in situ. In order to process the measured responses, the software module is developed. The accuracy of sound insulation determination using room impulse responses could be affected by the influence of certain parameters, such as the excitation signal and the excitation level, signal to noise ratio, etc, some of which are analyzed here. The complete procedure for measurement of sound insulation in situ is illustrated by the insulation measurement of the wall with the ventilation duct at the Faculty of Electronic Engineering in Niš.

[1]Dejan Ćirić is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: dejan.ciric@elfak.ni.ac.rs

[2]Borislav B. Budisavljević is with IMS Institute, Bulevar Vojvode Mišića 43, 11000 Belgrade, Serbia, E-mail: bbb@sbb.rs

[3]Aleksandar Pantić is with Knauf Insulation, Industrijsko naselje Belo Polje, 17530 Surdulica, Serbia, E-mail: aleksandarpantic@knaufinsulation.com

II. IMPULSE METHOD FOR SOUND INSULATION MEASUREMENT

From the general signal theory it is known that a linear and time invariant system can be characterized by its impulse response. This is valid for a particular room, but also for the sound transition from one room to another. Thus, it is possible to measure the sound insulation between two rooms, that is the sound insulation of a particular partition area by measuring impulse responses in both source and receiving room, Fig. 1. The adequate processing of these responses can result in sound pressure levels of both rooms and corresponding sound insulation value.

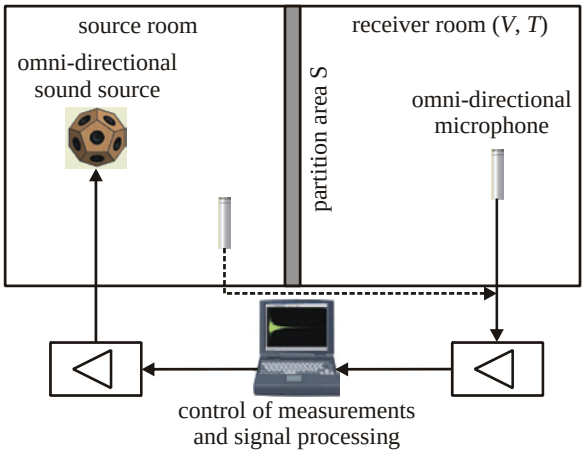


Fig. 1. Setup of measurement system for the sound insulation measurement by impulse method

The measurements of the room impulse responses can be done by different methods including widely used and standardized Maximum Length Sequence (MLS) technique and SineSweep technique [2].

The signal processing for sound insulation determination can be performed in time domain, but also in frequency domain. Here, only processing in time domain will be used and presented. This processing is based on the determination of sound pressure levels of source and receiver room from the measured impulse responses by

$$L_{s/r} = 10\log\left[\frac{W_0}{C_{ref}}\int_0^{\infty} h_{s/r}^2(\tau)d\tau\right] \quad (1)$$

where W_0 is a constant specifying the signal power per unit bandwidth of the excitation signal, and C_{ref} is arbitrarily selected reference value for the level calculation.

Now, since it is necessary to determine the spatially average sound pressure levels in both rooms based on measurements in several points in both rooms, the level difference is calculated by

$$D = L_s - L_r = 10\log\left[\frac{\frac{1}{m}\sum_{i=1}^{m} 10^{L_{s,i}/10}}{\frac{1}{n}\sum_{i=1}^{n} 10^{L_{r,i}/10}}\right] \quad (2)$$

The quantity that is usually used to represent the sound insulation is the sound reduction index that is calculated as

$$R = D + 10\log\left[\frac{ST}{0.16V}\right]. \quad (3)$$

where S is the area of the test element, and T is the reverberation time of the receiver room. The sound level difference or sound reduction index is usually determined for subset of third octave bands in certain frequency range. This number of third octave bands as well as number of microphone positions depend on the standard to be applied. For instance, the standard ISO 140-3 [1], defines at least frequency bands from 100 Hz to 5000 Hz, and five microphone positions in each room.

For the determination of a sound pressure level from the room impulse response, the important parameter is the range of the response in which the level is calculated according to Eq. (1), since it is not possible to use the range from 0 to infinity. Some recommendations are that this range should be from the response beginning to 20 or 15 dB down point from the peak of the squared impulse response, or alternatively it could be to the knee, where the main decay intersects the noise floor [2], [5], [6].

III. MEASUREMENT PROCEDURE

The measurements were carried out in the laboratories of IMS Institute in Belgrade that are certified for the sound insulation measurements by the classical method. However, sound insulation in this case was measured by application of the impulse method. The central component of the measurement system was the notebook computer used for the excitation signal reproduction, recording of the response and overall signal processing, as shown in Fig. 1. Besides, the measurement system consists of the amplifier, the omni-directional sound source, the omni-directional microphone and the microphone amplifier.

Beside the measurements in the specialized laboratory, they were also performed in situ, that is, the sound insulation of the partition wall between two laboratories at the Faculty of Electronic in Niš was also measured. Geometry of the source and receiver rooms together with the sound source and microphone positions are presented in Fig. 2. Important characteristic of the partition wall is that it contains the ventilation duct directly joining two laboratories from either side of the wall, that is, the source and receiver room. Thus, the sound insulation of this concrete wall is significantly reduced, which disturbs the working activities in these laboratories. The dimensions of the source and receiver rooms are 6×7.5×4 and 9×7.5×4 m^3, respectively. The mentioned ventilation duct has the dimension 0.4×0.7 m^2, and it is in the vicinity of the ceiling. The microphone positions are chosen in such a way that the distance to the closest wall is not smaller than 1.5 m, and from the sound source in the source is not smaller than 2 m. The height of the microphones is about 1.5 m, while the height of the sound source is about 1.8 m.

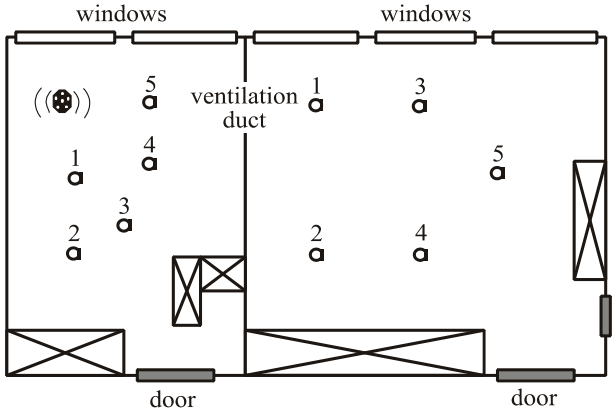


Fig. 2. Position of the sound source and microphone locations in the source and receiver room for measurement of sound insulation in situ

For both measurements, in the specialized laboratory and in situ, the Matlab software is developed for extraction of the room impulse responses and their complete processing for the purpose of calculation of the sound level difference and the analysis of the results. Room impulse responses were measured by application of SineSweep technique [7].

IV. RESULTS OF MEASUREMENTS IN LABORATORY

In order to illustrate the possibilities of the impulse method, the analyses of a dynamic range in the critical measurement of high sound insulation of porous concrete panels performed in the specialized laboratory of IMS Institute is done here. Since the sound level is determined from room impulse responses, then it is more appropriate to calculate Peak-to-Noise Ratio (PNR) or equivalently Dynamic Range (DR) of the decay curve representing the logarithm of squared impulse response filtered in third octave bands. DR represents a difference of the maximum level from the decay curve beginning and the knee where main decay intersects the noise floor.

Thus, PNR and DR values are determined from the impulse responses filtered in third octave bands and the results for the receiver room are presented in Fig. 3. The difference between these quantities is the consequence of non-stationary background noise level [8]. The presented values are almost in

the whole frequency range of interest greater than 20 dB. Thus, even in this extreme case of high sound insulation, the dynamic range of the impulse responses is high enough for reliable determination of the sound levels.

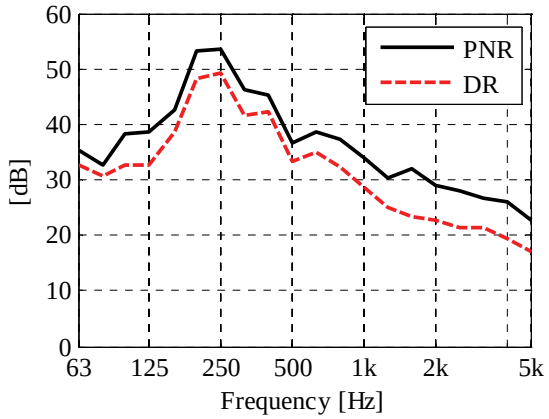


Fig. 3. PNR of the receiver room impulse response together with DR of the corresponding decay curves

In order to illustrate the gain enabled by the extraction of impulse responses, the level of the response to excitation signal in the receiving room and level of background noise in the same room are calculated and presented in Fig. 4. In great part of the frequency range of interest, their difference representing SNR has low values (below 5 or 6 dB) and only in the range around several hundred hertz it is above 10 dB.

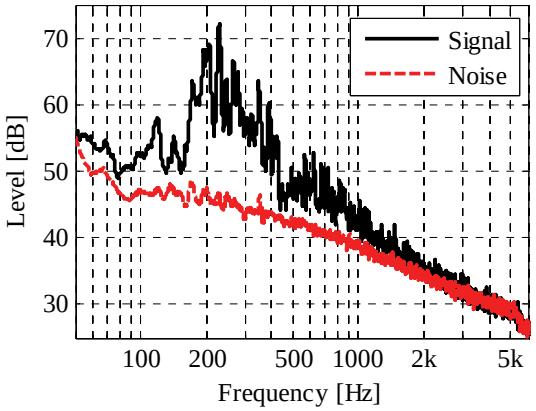


Fig. 4. The level of the response to excitation signal and level of background noise in the receiver room

The repeatability is tested on relatively small set of repeated measurements performed using the same excitation signal and level, but also using different excitation signal durations and levels. All the measurements were done in the same point in the source room, and in the same point in the receiver room of the IMS Institute laboratory. The results for standard deviation of the repetitions for nine measurements using three excitation sweep durations and three excitation levels are given in Fig. 5 together with the required values from ISO 3741 [9]. The presented values for the standard deviation are relatively small confirming the reliability of the measurements. These values are even smaller when the same excitation signal and level are used.

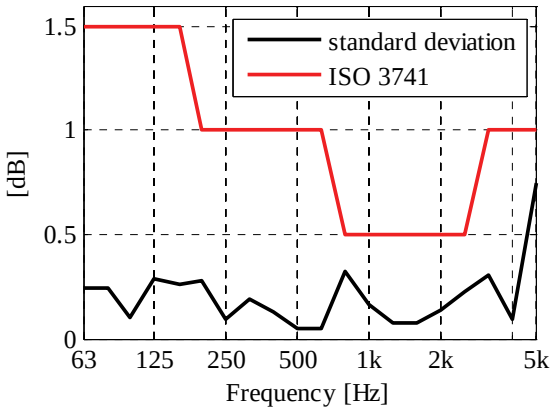


Fig. 5. Standard deviation of the repetition measurements for nine repetitions using three excitation signals and three excitation levels

The situation is somewhat different for the spatial standard deviation. The results for the standard deviation for the source room using five measurement points are presented in Fig. 6. These values are greater than the values for the repetition standard deviation. They are even greater than the required values defined in [9] at frequencies below 125 Hz.

The developed procedure for sound insulation measurement by impulse method was implemented for testing a number of partitioning structures within the project TR 21013 financed by Ministry of Science and Technological Development of Serbia. The results of sound insulation, that is, the level difference between the source and receiver room, for three characteristic examples – two porous concrete panels and the partition consisting of a window are given in Fig. 7. Thus, the porous concrete structures have somewhat different sound insulation values, but still significantly higher than the partition with the window in the whole frequency range. The results obtained by impulse method are confirmed by the classical measurement method that will be published in further publications of the authors.

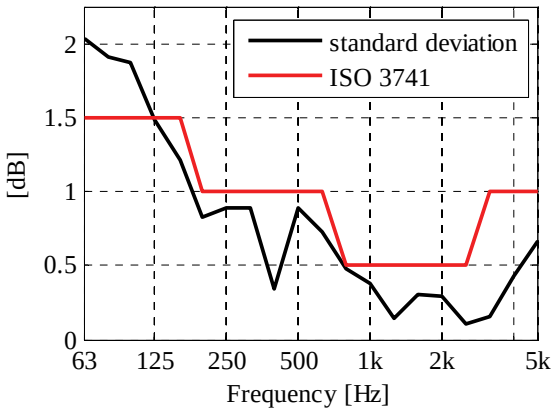


Fig. 6. Spatial standard deviation of the source room based on five measurement points

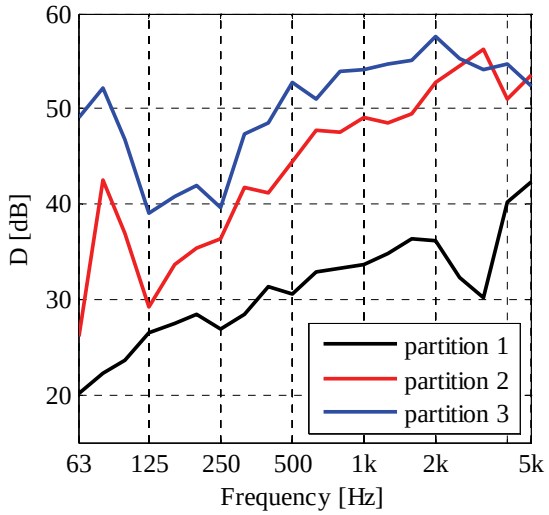


Fig. 7. Sound insulation (level difference) for three examples of partitioning structures measured in the laboratory: the window (partition 1) and porous concrete structures (partitions 2 and 3)

V. Results of Measurements in Situ

The sound insulation of the partitioning wall with the ventilation duct separating two laboratories was also measured by the impulse method representing measurement in situ. The sound levels in both source and receiver room determined from the measured and filtered impulse responses in third octave bands are given in Fig. 8(a). The difference of the spatially averaged levels in both rooms is presented in Fig. 8(b). The obtained values representing the sound insulation are rather low confirming already noted low insulation between these two laboratories.

VI. Conclusion

The procedure for implementation of the impulse method for sound insulation measurement is developed and presented here. Its application for measurements in the specialized laboratory and in situ is analyzed. The values for repetition and spatial standard deviation confirm that reliable results can be obtained by this method. Moreover, the gain in signal to noise ratio as a consequence of the room impulse response extraction enables measurements in extreme conditions such as measurements of high sound insulation values and measurements with high background noise. The presented examples of sound insulation measurements emphasize the possibilities of the impulse method and its flexibility especially for the measurements in situ.

Acknowledgement

Results presented in this paper are obtained within the project TR 21013 financed by Ministry of Science and Technological Development of Serbia.

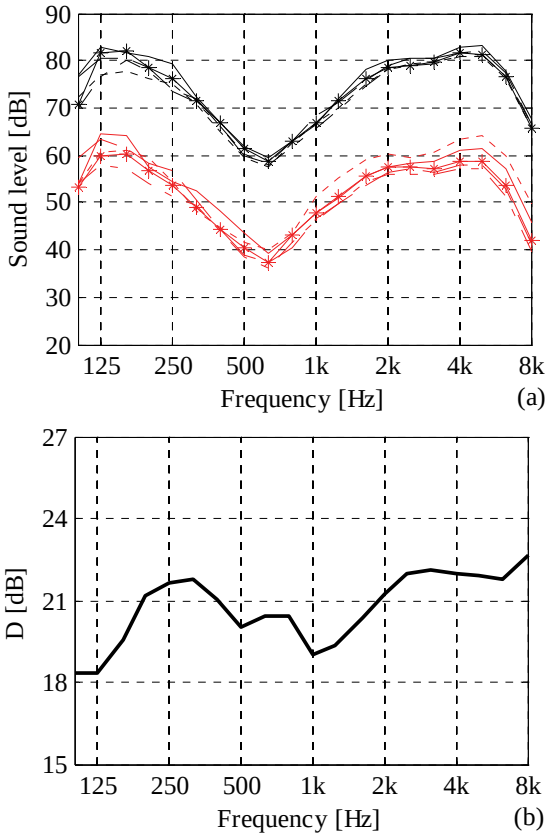


Fig. 8. Sound levels in the source and receiver room in situ measurements (a), and the difference of these levels (b)

References

[1] ISO 140: *Acoustics – Measurement of Sound Insulation in Buildings and of Building Elements – Part 3: Laboratory Measurements of Airborne Sound Insulation of Building Elements*, International Organization for Standardization, 1995.

[2] ISO DIS 18233: *Acoustics – Application of New Measurement Methods in Building Acoustics*, International Organization for Standardization, 2006.

[3] C. Hopkins, *Sound Insulation*, USA, Elsevier, 2007.

[4] R. Venegas, M. Nabuco and P. Massarani, "Sound Insulation Evaluation Using Transfer Function Measurements", Inter Noise 2005, Congress Proceedings, Rio de Janeiro, Brazil, 2005.

[5] M. Vorländer, M. Kob, "Practical Aspects of MLS Measurements in Building Acoustics", Applied Acoustics, vol. 52, pp. 239-258, 1997.

[6] D. Ćirić, M. Milošević, "Inaccuracies in Sound Pressure Level Determination from Room Impulse Response", Journal of the Acoustical Society of America, vol. 111, pp. 210-216, 2002.

[7] A. Farina, "Simultaneous Measurement of Impulse Response and Distortion with a Swept-sine Technique", 108th Convention of Audio Engineering Society, abstract in Journal of Audio Engineering Society, vol. 48, no. 4, p. 350, 2000, preprint 5093.

[8] D. Ćirić, M. Marković and B. Stojić, "Analysis of Background Noise Influence in SineSweep Measurements of Room Impulse Responses", proposed for ICEST 2010, to be held in Ohrid, FYR Macedonia, 2010.

[9] ISO 3741: *Acoustics – Determination of Sound Power Levels of Noise Sources – Precision Methods for Broad Band Sources in Reverberation Rooms*, International Organization for Standardization, 2003.

Analysis of Background Noise Influence in Impulse Response Measurement by SineSweep Technique

Dejan Ćirić[1], Miloš Marković[2], Branko Stojić[3]

Abstract – **Impulse response of a linear time invariant system represents one of its main characteristics. Unfortunately, there are some common problems in its measurement, especially in some cases, such as the measurement of a room impulse response in room acoustics. Independently of applied measurement technique, one of the most important problems is background noise.**

During the application of SineSweep technique for impulse response measurements, it is noted that the extracted impulse response contains "double horn" shape noise signal as a consequence of background noise. The influence of that noise is analyzed here by simulations and measurements. In addition, the analysis is extended to the cases where the system under test contains nonlinearity. Different impulse responses, that is, different systems are used as a test system including some filters and rooms. The results show that this specific shape of background noise in the extracted impulse response represents important negative characteristic that could affect further processing of the response. Besides, the shape of background noise in impulse response depends on present noise and the excitation signal parameters. Understanding of the background noise influence is very important for achieving high dynamic range representing an advantage of SineSweep technique.

Keywords – **Background noise, Impulse response, SineSweep.**

I. INTRODUCTION

One of the most significant characteristics of a linear time invariant system is the transfer function and its equivalent in time domain – impulse response (IR). In room acoustics, impulse response of a room represents a fundamental characteristic for determining a number of room acoustic quantities [1]. Thus, measurements of room impulse responses require a high level of accuracy. Various methods and techniques have been used for this measurement [1-3]. Signals applied for excitation in the measurement process can be categorized in different ways. The most common broad-band signals are Maximum Length Sequence (MLS), which is pseudo-random white noise, and swept sine (SineSweep) – sine signal whose frequency varies with time.

During the measurement process, different room conditions can have influence on the measurement quality. Some of them can cause such negative effects so that obtained impulse response becomes useless [4]. One of the practical problems present in measurements is background noise (BN). In most cases, this noise is white Gaussian noise, non-correlated with the excitation signal. The main consequence of presence of background noise in an impulse response (IR background noise or IR BN) is decreasing of signal-to-noise ratio (SNR) or peak-to-noise ratio (PNR). Different measurement techniques are more or less immune to background noise. Some of them require a low level of noise for reliable results (the impulse excitation method). On the other hand, using MLS technique, high level of SNR can be obtained in real conditions because of its strong immunity to all kinds of noise (white, impulsive or others) [3]. The usage of SineSweep measurement technique allows the exclusion of all or most of the harmonic distortion products, leaving only background noise as a limitation for the achievable SNR.

The influence of background noise in measurements of room impulse response by SineSweep technique is analyzed here by computer simulations. The results are confirmed by measurements.

II. BACKGROUND NOISE INFLUENCE

Any measurement method using the excitation signal of equal length, spectral distribution and total energy will lead to exactly the same amount of noise rejection if the entire period of the unwindowed impulse response is considered. The difference between the various measurement techniques lies merely in the way that the noise is distributed over the period of the recovered impulse response [2].

MLS technique is able to randomize the phase spectrum of any component of the output signal which is not correlated with the MLS input sequence [5]. Any disturbing signal will be phase randomized, and this will lead to a uniform distribution of the background noise along the deconvolved impulse response, Fig. 1. Such a uniform distribution of IR background noise appears in most techniques for IR measurements. This noise limits impulse response decay to the point where the noise level is established. In this way, the dynamic range of impulse response is also limited.

III. SINESWEEP TECHNIQUE

Excitation signal in SineSweep technique is a sine signal with linearly (linear swept sine) or exponentially (logarithmic swept sine) varying frequency.

[1]Dejan Ćirić, Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: dejan.ciric@elfak.ni.ac.rs

[2]Miloš Marković, Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: milos.markovic@elfak.rs

[3]Branko Stojić, Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: branko.stojic@elfak.rs

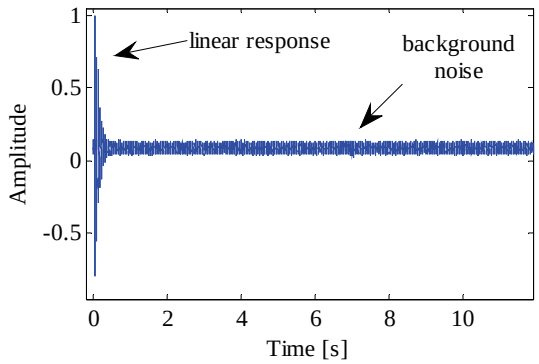


Fig.1. IR obtained by MLS technique

One of the main advantages of SineSweep technique is the possibility of obtaining a great dynamic range. The impulse response can be extracted either in time or in a frequency domain [6]. For linear deconvolution directly implemented in time domain, there is a need for proper inverse filter $f_{ss}(t)$ capable of transforming the excitation signal $x(t)$ into a delayed Dirac's delta function $\delta(t)$:

$$x(t) * f_{ss}(t) = \delta(t - t_d) \tag{1}$$

Then, the deconvolution of the system's impulse response $h(t)$ can be obtained simply convolving the measured output signal $y(t)$ with the inverse filter $f_{ss}(t)$:

$$h(t) = y(t) * f_{ss}(t) \tag{2}$$

The impulse response obtained in this way is twice as long as the excitation signal. Useful IR (linear response) starts at the moment delayed for the length of the excitation signal.

Another approach is to use a circular deconvolution in frequency domain. Impulse response can be obtained taking the IFFT of the transfer function:

$$h(t) = IFFT\left[\frac{FFT[y(t)]}{FFT[x(t)]}\right] \tag{3}$$

IV. Methods of Investigation

Background noise influence is analyzed through many simulations arranged in two groups. In the first one, Dirac's impulse is used as a pre-defined IR. This impulse is idealistic representation of impulse response. In the second group, instead of Dirac's impulse, simulated room IR is used. As an excitation signal, logarithmic swept sines of different durations are used. The duration of excitation signal is chosen to be equal to the duration of MLS signal for later comparison of SineSweep and MLS technique. Also, different background noises are generated. All of the BNs have Gaussian Probability Density Function (PDF), but their duration and level are different. All simulations follow the same method of investigation. Generated BN is added to the excitation sweep signal. That modified excitation signal is convolved with pre-defined impulse response. The result of this convolution is the output (response) signal with background noise. In the process of deconvolution, output signal is convolved with the inverse filter (generated on the base of excitation signal) yielding an impulse response with IR background noise present. This response is compared with starting (pre-defined) IR and the differences are observed. The influence of changing of BN parameters is considered, too.

In addition to the simulations, the background noise influence in SineSweep technique is analyzed in real conditions through the measurements of IR. Two types of measurements were performed: measurements of room IR carried out in one of the laboratories at the Faculty of Electronic Engineering in Niš and measurements of the band-pass filter IR.

V. Background Noise Influence in Simulations

A. Dirac's impulse as pre-defined IR

Dirac's impulse doesn't have all parts of real IR (e.g. room IR), but its shape is convenient for the analysis. If Dirac's impulse is used as a pre-defined impulse response, after adding background noise to the excitation signal, IR with background noise present is extracted only by the convolution of that signal plus noise (the modified excitation signal) with the inverse filter. For this group of simulations, the logarithmic sweep of duration equal to duration of MLS signal of order 16 is used as the excitation.

Without background noise, the extracted IR should be approximately equal to the starting Dirac's impulse delayed for the duration of the excitation signal. The results of analysis point out that when background noise is present a double horn shape signal (IR background noise) appears in the impulse response as a consequence of the background noise. The noise level in IR sets up gradually from the beginning of the response (region A), after reaching certain level, the noise level is increased but only slightly in relatively wide region (region B), and after reaching its maximum it is decreased also in relatively wide region (region C), Fig. 2.

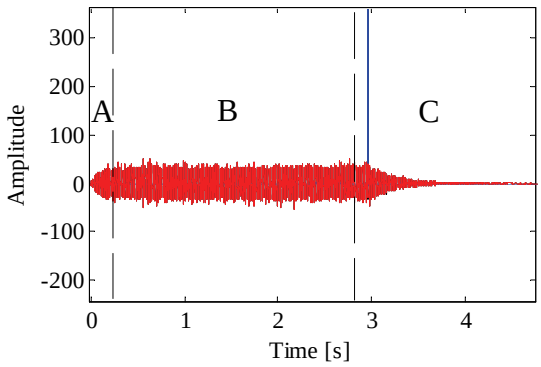


Fig.2. Three regions in double horn shape of IR background noise (red curve representing only the noise in IR without linear response) together with the linear response (blue curve representing zoomed linear IR)

The mentioned changing of the IR background noise level extends in both parts of the overall response, before and after the linear response. The limits of the noise regions (A, B and C) shown in Fig. 2 are only provisional, since it is difficult to make a clear distinction between the regions. These limits are also dependent on the length of the excitation sweep. Thus, the limit between the regions B and C (black dashed vertical line) could be in the first part of the response (before the linear impulse response) or it could be in the useful part of the linear impulse response.

The increase of the background noise amplitude (A_{BN}) leads to an increase of IR background noise amplitude (in the impulse response). The level of IR background noise (Mean Square Value (MSV) IR BN) is proportional to the level of background noise added to the excitation (MSV BN). Values of those levels for different background noise amplitudes (A_{BN}) are given in Table I. The ratio of MSV IR BN to the MSV BN is approximately $1.75 \cdot 10^3$ for all noise amplitudes.

TABLE I
LEVELS OF BACKGROUND NOISE

	A_{BN} [rel. units]				
	1	2	4	8	16
MSV BN*10^3	5.46	21.7	87.1	348	1395
MSV IR BN*10^6	9.58	38	152	612	2450

B. Room impulse response as pre-defined IR

Instead of Dirac's impulse response, this group of simulations uses room impulse response obtained by simulation algorithm [7]. This response is convolved with the excitation sweep yielding the output signal without background noise. IR without background noise is determined after the described deconvolution with the excitation (Eq. 2). This procedure is repeated when the background noise is added to the excitation sweep signal yielding as a result IR with background noise present. The logarithmic sweep signals of duration equal to durations of MLS signals of order 11 to 16 are used as the excitation.

The main characteristics of the IR background noise observed in the first group of simulations are present in this case, too. The IR background noise starts at the beginning of the IR, it is located in both parts of the overall response (before and after linear response) and the increase of noise level in the excitation leads to the increase of level of the IR background noise.

Impulse responses are then filtered in octave and one third octave bands, and IR background noise is analyzed. The background noise in the response is still non-stationary, but its shape is not completely the same as in broad band impulse response. Thus, at the beginning and at the end of the response there could be the region of the stationary background noise (region D). In the rest of the response, there is a region of more or less stationary background noise (region E), but of higher level than at the beginning and at the end of the response, Fig. 3. Between these two regions of two levels of IR background noise, the noise relatively suddenly increases and decreases its level (region F).

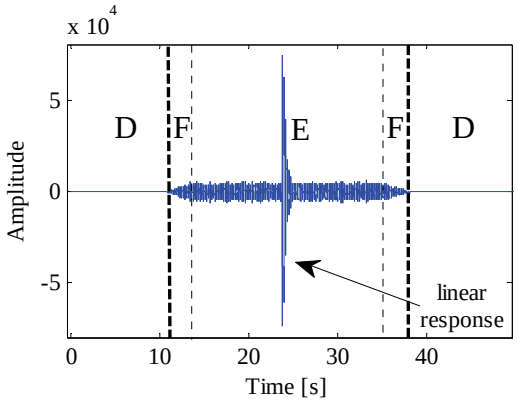


Fig.3. IR obtained by SineSweep technique and filtered in octave band at 800 Hz

Higher level region of IR background noise has approximately the same duration independently of the filter bandwidth (octave or one third octave). The length of this region depends only on the excitation signal duration. However, the filter central frequency, that is, the filter cut-off frequency determines the starting point of the higher level region: higher frequencies (wider filter pass-bands) move the starting point of higher level region towards the beginning of IR.

As a consequence of the regions with different noise levels, the noise level as well as PNR determination depends on the range in which these quantities are determined. Thus, one approach could be to determine the noise level in the range of higher noise level. Alternatively, the noise level could be determined in wider range, from the cross section of the linear response and noise to the end of the response. Of course, the noise level in the second case would be lower, that is, obtained PNR of IR would be higher.

The existence of non-stationary level of background noise implies also the problem of multi decay shape of the decay curve (integrated impulse decay). A typical example is given in Fig. 4, where the decay determined from broadband IR is presented together with the decay determined from filtered IR.

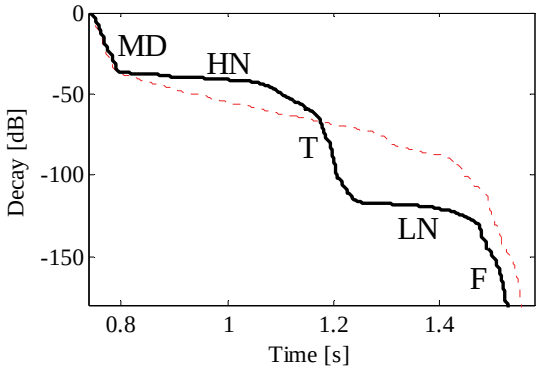


Fig.4. Decay curve obtained by backward integration of the broadband IR (thick black line) and IR filtered in octave band at 800 Hz (thin dashed red line)

Thus, instead of three parts, the decay curve consists of 5 parts. The first one is main decay (MD). The second part with smaller decay (HN) represents the region of higher noise level. The third part represents the first rapid fall of (T), and this is the region of transition between higher and lower noise level in the impulse response (the width of this region could be somewhat different in different impulse response, that is, in responses filtered in different octave or third octave bands). The fourth region represents the second region with smaller decay (LN), and this region corresponds to the last part of the response with lower background noise level, while the last part is the region of rapid fall of, because of finite upper limit of backward integration (F).

VI. Background Noise Influence in Measurements

A. Measurements of room IR

The influence of background noise in SineSweep technique is also analyzed in the measurements of room IR. Excitation sweep signals of durations equal to durations of MLS signals of order 15 to 19 are used. Extracted IRs confirm the expected results. The background noise in IR starts at the beginning of the IR, it is located in both parts of the overall response and the increase of noise level in the excitation leads to the increase of level of IR background noise. Besides, the shape of background noise in the filtered response is not completely the same as in broad band impulse response, as already noted in the results of simulations.

When the nonlinearity is present during the measurement of a room IR, the distortion products appear in addition to IR background noise, Fig. 5. Depending on the levels (amplitudes) of the products and noise, these products could be buried in the noise, that is, masked by IR background noise.

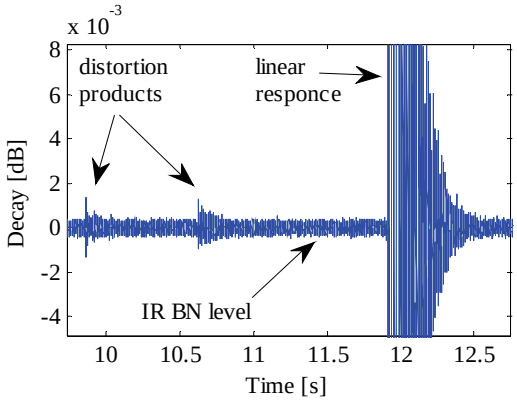


Fig.5. Measured IR: the distortion products are masked by background noise energy (zoomed IR)

B. IR measurements of pass-band filters

Measurements of IRs of two different filters, low- pass and high-pass filters are also performed.

Measured filters IRs indicate similar results as the simulations with filtered IR. Background noise in measured IRs is non-stationary, and it consists of two different levels, lower and higher. This could be explained by the fact that the IRs of these filters are not completely broad-band, but they contain only the part of the frequency range (the pass bands of the filters). Start moment of the higher level region is determined by the filter passband, that is, by its cut-off frequency.

VII. Conclusion

The presence of background noise in measurement of an IR by SineSweep technique results in IR background noise with double horn shape if the IR is broadband. This noise extends in both parts of the overall response (before and after the linear response). Level of the IR background noise is in proportion with level of noise present during the measurement. High level of background noise in IR decreases SNR and PNR of impulse response.

Filtering of broadband IR can change the IR background noise shape and its position in the response. The central frequency (cut-off frequency) of used passband filter (octave or third octave) determines the starting moment of higher level noise region. The existence of different levels of background noise in filtered IR implies the problem of multi decay shape of the decay curve. Thus, the decay curve obtained by backward integration of the filtered IR consists of five instead of three parts.

Acknowledgement

Results presented in this paper are obtained in the project no. 21013 financed by Ministry of Science and Technological Development of Serbia.

References

[1] L. Otshudi, J. P. Guilhot and J. L. Charles, “Overview of Techniques for Measuring Impulse Response in Room Acoustics”, *Proc. Institute of Acoustics*, **10**, pp. 407-414, 1988.

[2] S. Müller and P. Massarani, “Transfer-function Measurement with Sweeps”, *Journal of Audio Engineering Society*, **49**, pp. 443-471, 2001.

[3] G. B. Stan, J. J. Embrechts and D. Archambeau, “Comparison of Different Impulse Response Measurement Techniques”, *Journal of Audio Engineering Society*, **50**, pp. 249-262, 2002.

[4] D. Ćirić and M. Milošević, “Transient Noise Influence in MLS Measurements of Room Impulse Response”, *Acta Acustica un. With Acustica*, **91**, pp. 110-120, 2005.

[5] D.D. Rife and J. Vanderkooy, “Transfer-Function Measurement with Maximum-Length Sequences”, *J. Audio Eng. Soc*., **37**, pp. 419-443, 1989.

[6] A. Farina, “Simultaneous measurement of impulse response and distortion with a swept-sine technique”, Pre-print of the *108th Convention Audio Engineering Society*, Paris, Abstract in *Journal of Audio Engineering Society*, **48**, pp. 350, 2000.

[7] A. Pantić, D. Ćirić, “Modeling of sound energy decay in room”, *Proceedings of LII Conference of ETRAN*, Palić, Serbia, 2008, paper AK2.5 (in Serbian).

Web-based Application of the Text-to-Speech Synthesis System "Speak Macedonian"

Branislav Gerazov, Toni Janevski and Zoran Ivanovski[1]

***Abstract* – This paper describes an online presentation environment for the text-to-speech (TTS) system "Speak Macedonian". The project included the generation of a standalone MATLAB executable of the TTS system and a website environment that would allow people to use it. The website environment was realized with a combination of an HTML website and a PHP script. The PHP script realizes the interface between the user and the system and also provides data logging so that every activation of the application can be logged for analysis.**

***Keywords* – online presentation, MATLAB, PHP, HTML, Internet**

I. INTRODUCTION

The artificial synthesis of the human voice has inspired people's work for centuries. Speech synthesis has truly come a long way, from the XVIII century mechanical contraptions of von Kempelen, the 1930's mechanic VODER, the analog electronic synthesizers of the 1950's, to today's digital computer based systems. Development of speech synthesis in the digital era has been marked by three distinct paradigms: formant, articulatory and concatenative synthesis. Formant and articulatory synthesis are based on parametric modeling of the speech waveform and the speech production process, respectively. They served as the basis of TTS systems of the 1970's up until the early 1990's, [1]. Because of the difficult and expensive process of developing the high quality models used in these TTS systems, they were developed only for a handful of languages.

With the increase of computing power and memory, TTS systems of the 1990's moved towards the relatively simpler paradigm of concatenative speech synthesis. Concatenative synthesis is based on straight-forward generation of the required output speech through the concatenation of small prerecorded speech segments, also called units. The simplicity behind this approach, spurred the creation of new TTS systems for various languages around the world. In the Slavic speaking part of the Balkans various systems have appeared, most prominent of which is the "AlphaNum" unit-selection TTS system for Serbian and later Croatian, [2]. Other systems include "SpeechLab" for Bulgarian and "Govorec" and "Proteus" for Slovenian.

Several attempts have been made to synthesize speech in Macedonian. First attempts of building such a system date back to 1997, [3]. Later trials emulated Macedonian using the Croatian diphone inventory under the MBROLA framework, [4], with a similar approach by the AlfaNum team, that used the system's Serbian diphone inventory for the job, [5]. Today, two high quality TTS systems for Macedonian have been developed: TTS-MK, which has been finalized in 2009 [6], [7], and "Speak Macedonian", our system, which is in its final stages of development, [8].

This article presents an online presentation environment for our TTS system. It is designed to bring our system closer to the public in Macedonia and the world. This is to serve both for promotional purposes but also for gaining user insight on various aspects of the synthesized speech quality. The environment is not designed for testing purposes. Final tests of the system will be carried out in a controlled environment in the facilities of our faculty, as work on the system reaches completion.

The core of the addressed problem is enabling a TTS system realized completely in the MATLAB programming language to execute on a server environment without a running MATLAB installation, and to build a web script able to pass arguments to and from the MATLAB application and the user. The presented solution was realized using MATLAB's built-in compiler and a combination of HTML and PHP. In the following sections we outline the designed solutions.

II. THE TTS SYTEM "SPEAK MACEDONIAN"

The TTS system "Speak Macedonian" as presented in [8], is based on the paradigm of concatenative speech synthesis. It uses a mixed unit inventory with segment lengths of one and two voices. This offers quality speech output with an optimized size inventory. The system includes automatic pitch, duration and amplitude modeling of the segments providing for artificial prosody generation. It also includes a voice transformation module presented in detail in [9], that allows transformation of the default male voice to female, childlike and elderly, as well as various in-between modifications. The TTS system's graphical user interface (GUI) is given in Fig. 1. It features a text field for text input and two pushbuttons – one for speech generation and one for repetition. It also includes controls for the voice transformation module.

[1]Branislav Gerazov, Toni Janevski and Zoran Ivanovski all are with the Faculty of Electrical Engineering and Information Technologies, Karpos II B.B., 1000 Bitola, Macedonia, E-mail: {gerazov , tonij , mars} @feit.ukim.edu.mk

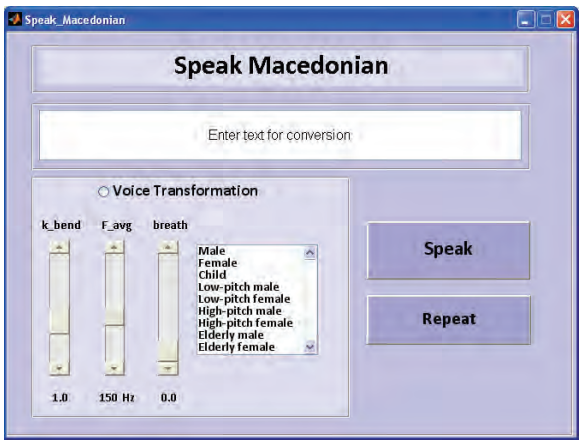


Fig. 1. The graphical user interface of the TTS system "Speak Macedonian"

An example of synthesized speech output is given in Fig. 2. The figure shows the time-domain waveform, spectrogram and the intonation contour of the synthesized word "makedonski" (eng. "Macedonian").

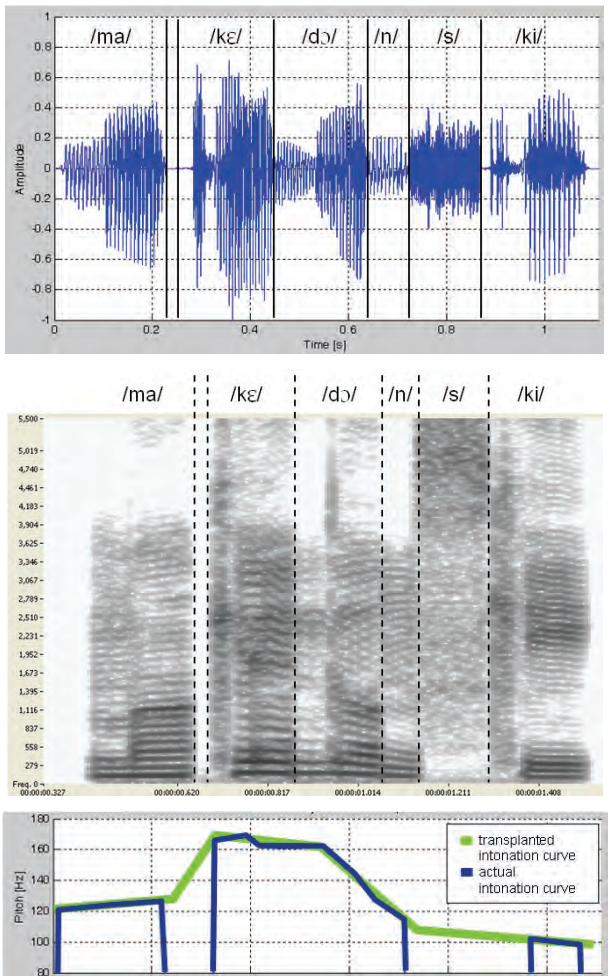


Fig. 2. Time-domain waveform (top), spectrogram (middle) and the intonation contour (bellow) of the synthesized word "makedonski"

III. STANDALONE MATLAB EXECUTABLE

For the purposes of relocating the TTS system to a server platform it was necessary to make it work without the need of a running MATLAB installation. This was carried out by generating a standalone executable file from the TTS system's MATLAB code. The executable file was generated using the built-in MATLAB Compiler, [10]. The principal purpose of the compiler is to offer distribution of MATLAB applications and libraries to users that lack a MATLAB installation. It can be used to compile m-files and other MATLAB code to standalone C/C++ applications for UNIX, Windows and Macintosh platforms as well as C/C++ libraries such as Windows' Dynamically Linked Libraries (DLL's).

Applications generated by the MATLAB Compiler consist of two parts: a binary executable file and an archive file. The executable file contains the project's main function, and is created using wrapper files that nest the m-code and provide an interface to it. The archive file or the Component Technology File (CTF), contains the user created MATLAB functions and the data files used by the application.

The MATLAB compiler can be easily operated through the Deployment Tool GUI shown in Fig. 3. The compilation process generates C/C++ source code and the EXE and CTF files that are needed for the distribution of the project. The two files can additionally be packaged by the Deployment Tool in a self-extracting archive file.

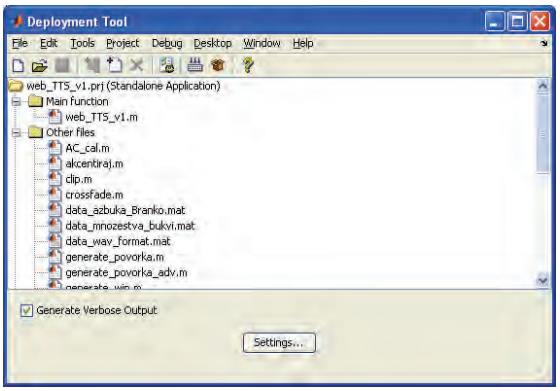


Fig. 3. MATLAB Compiler's Deployment Tool GUI compiling the TTS system

To use the compiled application the executable and the CTF archive (or the package) need only to be copied (or extracted) to the destination platform. If the destination platform lacks MATLAB, then it is also necessary to install the MATLAB Component Runtime (MCR) on it. The MCR contains a data base of all the MATLAB built-in libraries that allow m-code execution.

When the EXE file is first started it will extract the CTF archive into a new folder. Although the filenames of the user-written m-code are left intact, the code within these files is protected using the Advanced Encryption Standard (AES) with keys encrypted with 1024-bit RSA keys. This denies access to the project's code by third-party users of the distribution package. The binary executable and the mat-data structures are not encrypted.

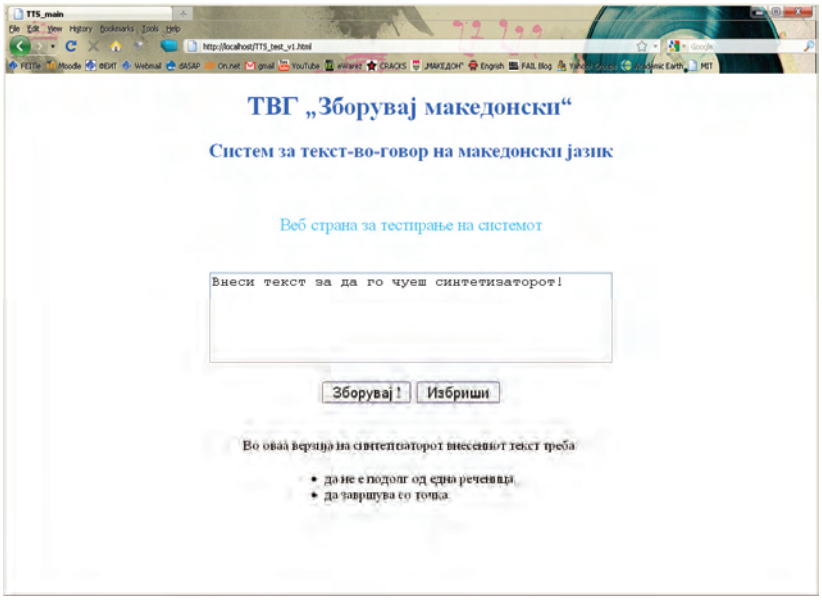


Fig. 4. Layout of the web page for online presentation of the TTS system “Speak Macedonian“

For the purposes of interfacing the TTS system through a website it was necessary to redesign its interaction method. Namely, the role of the GUI was to be taken entirely by the website. This was carried out with the use of two auxiliary files – one text file and one audio file. When started, the executable loads the text TXT file that contains the text input by the user and starts the TTS synthesis process. After the synthesized speech is output the executable writes the audio data in a WAV audio file. The creation of the text file with user input text and the transfer of the audio file back to the user is to be done by the website interface.

IV. Website interface

The website interface that users can use to access our TTS system was built using a combination of PHP and HTML. The layout of the finished website is shown in Fig. 4. It consists of a text area for user input and two pushbuttons. The submit button “Зборувај!” (eng. “talk”), calls the PHP service script that generates a text file with the user input, starts the MATLAB executable and sends the generated audio file to the user. The pushbutton “Избриши” (eng. “clear”) resets the text area to null.

The project has so far been tested on a local-host machine using a WampServer installation. Wampserver (Windows Apache MySQL PHP server) is a free internet development environment for Windows, [11]. It can be used to locally develop web sites and applications based on Apache, MySQL and PHP, before they are put online.

A. HTML code

The HTML code was written in accordance to the XHTML 1.0 Transitional standard. All styling and formatting was done using the Cascading Style Sheets (CSS) formatting language. A key part of the webpage is the form declaration. The action attribute in the form declaration points to the PHP service script that executes the TTS system. The form is declared to accept user input in UTF-8 i.e. Unicode formatting allowing for the use of Cyrillic’s. The method of sending user input is chosen to be POST, as it does not limit the length of the input text. The textarea element which creates the text input field, sets the onfocus attribute so that the default text (“Внеси текст за ...”) is cleared when the user clicks on it. Further clicks on the text field won’t clear it allowing the user easy modification of the input in successive trials of the TTS system.

B. PHP script

In this section we will present the key points of the PHP script that uses the user’s input from the HTML form to call the MATLAB application and send its output to the user. The PHP script executes in the steps shown in Fig. 5.

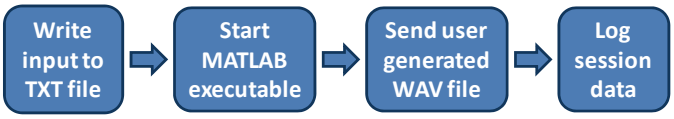


Fig. 5. Block diagram of PHP service script

First, the script takes the input text from the HTML form and writes it to an auxiliary TXT file. Then, it activates the MATLAB executable, which processes the text and generates an audio WAV file containing the synthesized speech. Next the PHP script sends this WAV file to the user in the form of an attachment, through the use of HTTP headers. At the end, the script logs the data including the current time, user’s IP address, input text and success of transfer, in a log file.

The transfer of the WAV file was realized through the sending of HTTP headers for Cache-Control that disabled cashing of the sent file along the links of the server-host connection, and HTTP headers that declared the length of the file, MIME file type (WAV audio), and the method that the

file is to be sent with (as an attachment). The sending of these headers opens a file download window in the user's browser window, Fig. 6. Next the PHP script reads the WAV file to the server's output buffer and flushes the buffer to start the file transfer. At the end, the script checks whether the transfer was successful by checking the status of the connection.

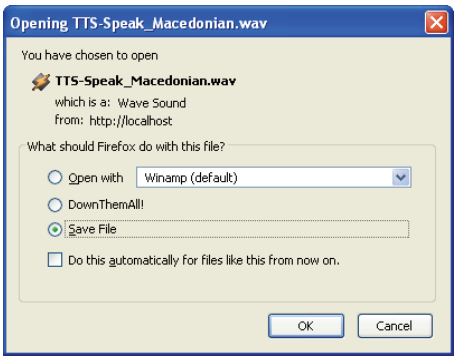


Fig. 6. Sending the generated synthesized speech output to the user by the PHP service script as seen by the user

Two more practical key points of the PHP script are also worth mentioning:

• set_time_limit( ) is used to prolong the maximum time execution limit of the PHP script because the MATLAB executable can run longer than the default 30 sec, especially when started for the first time (due to the extraction of the CTF archive) or when synthesizing long text inputs.

• When writing the text to the TXT file it was necessary to first write a header in the file that declared it as a Unicode TXT file "\xEF\xBB\xBF".

V. Results and Future Work

The project has so far been tested on a local server installation. It has proven to work for various user inputs, and it is the authors' opinion that it is ready to be put online. Before this can be done, it is necessary to finalize the TTS system in respect to finishing the intonation generation module and rounding up the unit inventory.

Our plan is to upload and integrate the project on the Digital Image Processing Team (DIPteam) website, [12], which is already active at our Department. This will be followed with promotion of the system in the Macedonian Internet community, which will hopefully lead to presentation of the system to the masses.

VI. Conclusion

The main purpose of the project was the online presentation of the TTS system "Speak Macedonian" developed at our department. The project included the generation of a standalone MATLAB executable of the TTS system and a website environment that would allow people to use it. The website environment was realized with a combination of an HTML website and a PHP script. The PHP script realizes the interface between the user and the system and also provides data logging so that every activation of the application can be logged for analysis. The project was tested on a local server installation and is expected to be put online after finalization of our TTS system. New projects based on other digital signal processing systems built in MATLAB will follow.

References

[1] D. O'Shaughnessy, "Modern Methods of Speech Synthesis", IEEE Circuits and Systems Magazine, vol. 7, No 3, 3rd Quarter 2007, pp. 6 – 23.

[2] M. Sečujski, R. Obradović, D. Pekar, Lj. Jovanov and V. Delić, "AlfaNum System for Speech Synthesis in Serbian Language", In Proc. of the 5th Conf. Text, Speech and Dialogue, Brno, 2002.

[3] L. Josifovski, D. Mihajlov and D. Gorgevik, "Speech Synthesizer Based on Time Domain Syllable Concatenation", SPECOM '97 Cluj-Napoca, Oct. 27-30, 1997.

[4] M. Zrmanovska, *Sinteza na makedonskiot govor vrz baza na tekst so prozodija*, Master Thesis, Faculty of Electrical Engineering and Information Technologies, Skopje, Macedonia, 2005.

[5] V. Delić, M. Sečujski, D. Pekar, N. Jakovljević and D. Mishković, "A Review of AlfaNum Speech Technologies for Serbian, Croatian and Macedonian", IS-LTC 06, Ljubljana, Slovenia, 9. - 10 october, 2006.

[6] S. Chungurski, I. Kraljevski, D. Mihajlov and S. Arsenovski, "Concatenative Speech Synthesizers and Speech Corpus for Macedonian Language", 30th International Conference ITI Cavtat/Dubrovnik, Croatia, Jun. 23-26, 2008.

[7] S. Čungurski, *Govorni sintetizatori na makedonski jazik*, PhD Dissertation, Faculty of Electrical Engineering and Information Technologies, Skopje, Macedonia, 2009.

[8] B. Gerazov, G. Shutinoski and G. Arsov, "A Novel Quasi-Diphone Inventory Approach to Text-To-Speech Synthesis", MELECON '08, Ajaccio, France, May 5-7, 2008.

[9] B. Gerazov, S. Bogdanova, and Z. Ivanovski, "Linear predictive based voice transformation module for Macedonian TTS", TELFOR 2009, Belgrade, Serbia, Nov 24 – 26, 2009.

[10] *MATLAB Compiler User's Guide,* The MathWorks, 1995–2010.

[11] http://www.wampserver.com/en/

[12] http://dipteam.feit.ukim.edu.mk

DCT-domain Scaling Algorithm with Application in Stereo Image Compression

Antoaneta A. Popova[1] and Alexander A. Krupev[2]

***Abstract* – Our goal is to obtain a downsized version of an image in terms of its 8x8 block-DCT coefficients without passing through the spatial domain. DCT-domain scaling allows for fast, on-the-fly, resizing in DCT-based picture formats (like the widely used JPG) without the need for decompression, bilinear resizing and recompression. The ability of the human vision to compensate for resolution difference in the stereo image pair allows for smaller file sizes with preserved stereo perception.**

***Keywords* – Stereo images, resize, compression, DCT-domain, JPEG.**

I. INTRODUCTION

In this work will be discussed some algorithms for resizing in the DCT-domain (Discrete Cosine Transformation) and their application for stereo image compression.

A known fact is that picture resolution directly influences the bandwidth requirements for its transmission, so a straightforward approach is to downsize it for purposes of image compression. Stereo pairs consist of two similar images intended for each eye in order to achieve depth perception. Human vision is able to compensate the lack of higher frequency information in one of the images if it is present in the other [9], [7]. This allows us to downsize one of the images before transmission and upsize it to its initial resolution again before stereoscopic visualization. Most visual materials in Internet and visual information systems, including stereo image pairs, are stored in a compressed form, DCT-based JPG format [10] being most common. So if resizing is performed directly in the DCT frequency domain instead of the spatial image domain, we don't have to perform inversed DCT (iDCT) transform, spatial resizing and DCT transform again. Without recompression, the process is much computationally simplified and allows for faster, on-the-fly resizing according to the user and/or bandwidth requirements.

When considering stereo image compression, it is important to measure how well the depth perception is preserved in the reconstructed images. The evaluation of the reconstructed compressed images is still an open problem. We offer objective measure (PSNR), but it doesn't take into account the peculiarities of the human visual system [8]. Subjective experiments are performed also.

[1]Antoaneta A. Popova is with the Faculty of Communications, Technical University - Sofia, Kliment Ohridski 8, 1000 Sofia, Bulgaria, e-mail: antoaneta.popova@tu-sofia.bg

[2]Alexander A. Krupev is with the Faculty of Communications, Technical University - Sofia, Kliment Ohridski 8, 1000 Sofia, Bulgaria, e-mail: asfalot@abv.bg

Works on DCT-resize have been published by Dugad and Ahuja [1], Mukherjee and Mitra [2], Merhav and Bhaskaran [3], Yim and Isnardi [4] and others. They all use implicit iDCT – DCT conversion. A scheme of Dugad and Ahuja's method for 1D signal is shown below (Fig. 1).

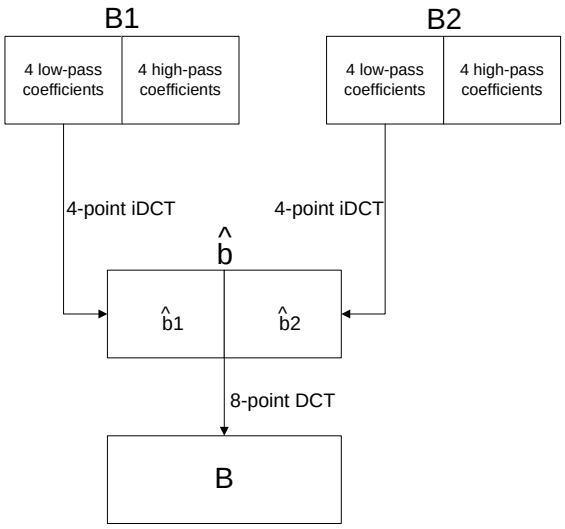


Fig. 1. 1D downsizing using Dugad and Ahuja's method

We see that iDCT is performed on 4 samples instead of 8 and low-pass filtering in the spatial domain is omitted thanks to the property of iDCT to smooth the edges i.e. we receive a low-passed and down-sampled version of the original 8-point block [5] with much lower computational complexity.
Similarly the up-sampling is done by padding the DCT block with zeroes.

Our main goal is to suggest and implement an algorithm for DCT domain resize based on this principle and apply it on stereo pairs for compression purposes. The quality of the stereoscopic visualization before and after compression will be evaluated also.

II. DCT-DOMAIN RESIZING SCHEME

The developed scheme works in the following steps:

1. The stereo pair is split into two images left (L) and right (R).
2. The left image blocks $L_{i,j}$ are sent to the decoder without processing.

3. $l_{i,j}$ are obtained after iDCT transform in the decoder.

4. The right image DCT coefficient blocks $R_{i,j}$ are truncated to a smaller size, determined by the user. The new smaller nxn (n≤8, n≥1) blocks $\tilde{R}_{i,j}$ can have either standard square shape or triangle shape, proposed by the authors (Fig. 2) for energy compaction reasons. Zig-zag scanning is performed on the remaining coefficients. Block size and/or shape information (up to 4 bits) is sent to the decoder once for the whole image.

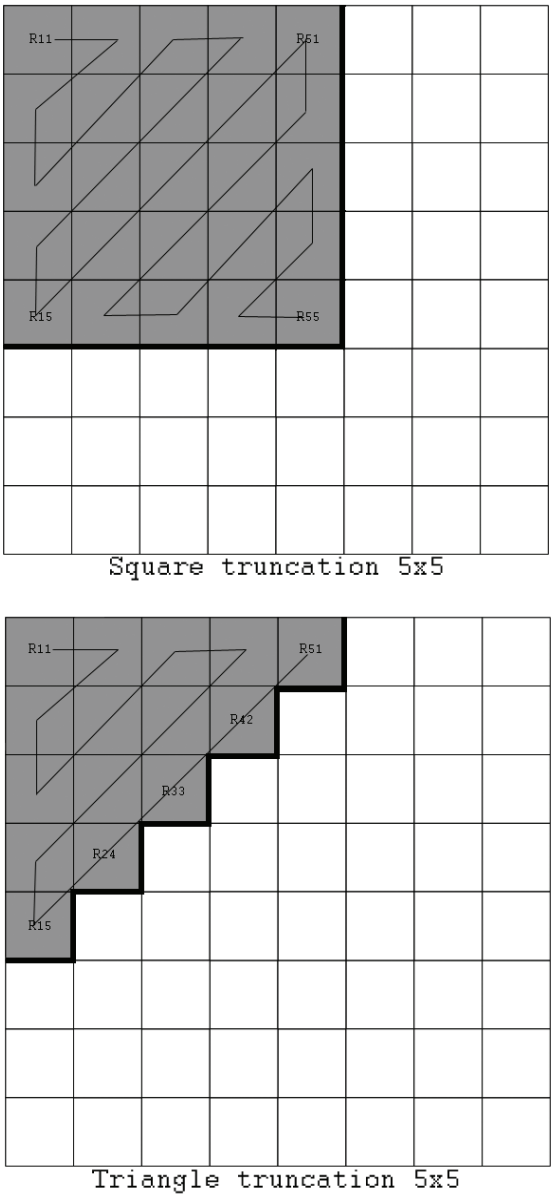


Fig. 2 DCT block square and triangle truncation with subsequent zig-zag scanning for n=5

5. The $\tilde{R}_{i,j}$ blocks are restored in the decoder after zig-zag scanning using the block size information. The truncated part is padded with zeroes – thus $\hat{R}_{i,j}$ blocks are obtained.

6. iDCT transform is applied on the restored to full size $\hat{R}_{i,j}$ DCT blocks and the reconstructed image blocks $\hat{r}_{i,j}$ to be displayed are obtained.

DCT and iDCT are implemented in the following form:

$$S(v,u) = \frac{C(v)}{2}\frac{C(u)}{2}\sum_{y=0}^{7}\sum_{x=0}^{7} s(y,x)\cos[(2x+1)u\pi/16]$$
$$\cos[(2y+1)v\pi/16]$$
$$s(x,y) = \sum_{v=0}^{7}\sum_{v=0}^{7}\frac{C(v)}{2}\frac{C(u)}{2}S(v,u)\cos[(2x+1)u\pi/16]$$
$$\cos[(2y+1)v\pi/16]$$
$$C(u) = 1/\sqrt{2} \quad for \quad u = 0 \qquad C(v) = 1/\sqrt{2} \quad for \quad u = 0 \qquad (1)$$
$$C(u) = 1 \quad for \quad u > 0 \qquad C(v) = 1 \quad for \quad u > 0$$

$l_{i,j}$ and the reconstructed image blocks $\hat{r}_{i,j}$ can be displayed and/or joined and saved by the user. The scheme is shown on Fig. 3.

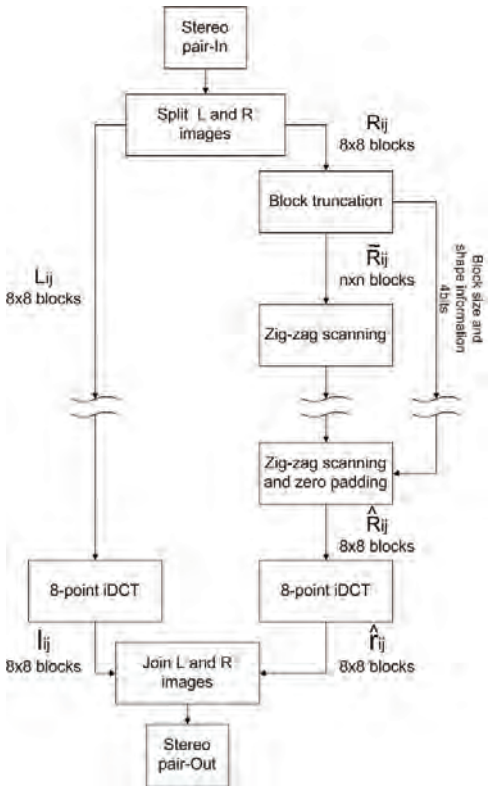


Fig. 3 Stereo pair DCT-domain resizing scheme

III. EXPERIMENTAL RESULTS

The scheme from Fig. 3 is implemented in Matlab 7.0.3 working environment. A PSNR evaluation module is added, comparing the processed image blocks $\hat{r}_{i,j}$ to the original $r_{i,j}$ right image according to Eq. 2:

$$PSNR = 10\log_{10}\left(\frac{MAX_I^2}{MSE}\right) = 20\log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right)[dB] \quad (2)$$

$$MAX_I = 255$$

Here MSE is:

$$MSE = \frac{1}{pq}\sum_{i=0}^{p-1}\sum_{j=0}^{q-1}[I(i,j) - \hat{I}(i,j)]^2 \quad (3)$$

p, q are the dimensions of the image, I and $\hat{I}$ - pixel intensity values before/after processing, ranging from 0 to 255.

A monochrome stereo pair of a juggler (resolution 784x600) has been processed (Fig. 4). 8 versions for square and triangle truncation have been obtained. Their PSNR values are shown in Table 1. The graphical representation is shown on Fig. 6 - PSNR as function of the DCT coefficients.

The resulting stereo pair for n=1 (1 DCT coefficient used - maximum compression, lowest quality) is shown on Fig. 5

Fig 4 Stereo pair of a juggler

Generally triangle truncation yields better results, up to 2.5dB in this case. But as we mentioned, PSNR doesn't take into account the subjective visual perception.

For the purposes of subjective evaluation the processed stereo pairs are visualized using the anaglyph method, based on color separation [6], and the shutter glasses method (we used 120 Hz ViewSonic VX2268wm LCD screen with NVidia 3D Vision stereoscopic glasses).

Fig. 5 Processed stereo pair from Fig. 4.

TABLE I

PSNR VALUES FOR THE JUGGLER IMAGE

DCT block size	**8x8**	**7x7**	**6x6**	**5x5**
PSNR (square trunc.)	∞	48.6737	42.2547	37.4448
PSNR (triangle trunc.)	44.4462	40.0701	36.6419	33.6821
DCT block size	**4x4**	**3x3**	**2x2**	**1x1**
PSNR (square trunc.)	33.3659	29.8392	26.5579	23.0081
PSNR (triangle trunc.)	30.9855	28.4413	25.9459	23.0081

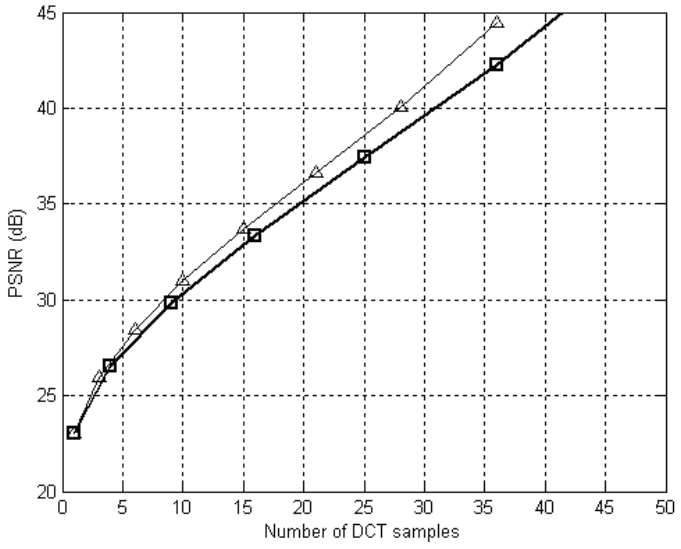


Fig. 6 The dependence between PSNR and the number of DCT samples. Thin line represents the proposed triangular truncation. Thick line – square truncation.

For n=1 stereo perception is lost and the image quality is very bad. For n=2 stereo perception is acceptable, image quality is still bad (noticeable pixelization). For n=3 and

square truncation stereo perception is excellent, but image appears somewhat blurry. Triangle truncation gives noticeable artifacts and some pixelization, the stereo perception is good though. For n=4 stereo perception is excellent as well as image quality. Triangle truncation yields a bit blurrier image, but it uses only 10 coefficients as opposed to 16 for square truncation – the compression achieved is much better. For 5x5 DCT blocks (triangle or square) image quality is almost indistinguishable from the original stereo pair (and at n=8 images are identical and $PSNR \to \infty$).

Two of the visualized anaglyphs are shown below on Fig. 6

Fig. 6 Anaglyphs created from the stereo pairs from Fig. 4 (left-best quality) and Fig. 5 (right-best compression, worst quality)

IV. Conclusion

The DCT-scaling scheme can be applied in conjunction with most other schemes for stereo image compression like joint pyramid encoding, block based disparity map encoding, pioneering block-based predictive encoding and others. Optimal results are achieved for 4x4 and 5x5 truncation. Reducing block size further hampers image quality first and even stereo perception (for 2x2 block size and lower).

Acknowledgement

This work was supported by National Ministry of Education and Science of Bulgaria under contract BY-I-302/2007 "Audio-Video information and communication system for active surveillance cooperating with a Mobile Security Robots".

References

[1] R.Dugad and N. Ahuja, "A fast scheme for image size change in the compressed domain," IEEE Transactions on Circuits and systems for Video Technology, vol. 11, no. 4, pp. 461--474, 2001.

[2] J. Mukherjee and S. Mitra, "Image resizing in the compressed domain using subband DCT" IEEE Transactions on Circuits and systems for Video Technology, vol. 12, No. 7, July, pp.620-627, 2002

[3] N. Merhav and V. Bhaskaran, "Fast algorithms for DCT-domain image down-sampling and for inverse motion compensation," IEEE Transactions on Circuits and systems for Video Technology, vol. 7, no. 6, pp. 468--476, 1997.

[4] C. Yim and M. Isnardi, "An Efficient Method for DCT-Domain Image Resizing with Mixed Field/Frame-Mode Macroblocks", IEEE Transactions On Circuits And Systems For Video Technology, Vol. 9, No. 5, August 1999.

[5] K. N. Ngan, "Experiments on two-dimensional decimation in time and orthogonal transform domains", Signal Processing, vol. 11, pp. 249-263, 1986.

[6] A. Krupev and A. Popova, "Flickering Reduction Algorithm in Anaglyph Stereoscopic Images", ICEST'08 Proceedings of Papers, Vol. 1, pp.125-128, Nis, Serbia, 2008.

[7] M. Bax and A. Vitus, "Stereo Image Compression", http://www.stanford.edu/~mbax/ee392c/report.html, Stanford, Dec. 1997.

[8] I. Dinstein, M.G. Kim, A. Henik, and J. Tzelgov, "Compression of stereo images using subsampling transform coding," Optical Engineering, vol. 30, no. 9, pp. 1359-1364, Sept. 1991.

[9] L. Kaufman, *Sight and Mind: An Introduction to Visual Perception"*, Oxford University Press, 1974.

[10] W. Pennebaker and J. Mitchell, "*JPEG still image data compression standard (3rd ed.)"*, Springer. p. 291, 1993.

SESSION CSIT I

Computer Systems and Internet Technologies I

A CDMA and PAM Signaling Interconnect Architecture

Goran S. Jovanović[1], Tatjana R. Nikolić[2] and Mile K. Stojčev[3]

***Abstract* – The system performance of complex CMOS VLSI ICs are limited nowadays by its interconnect bandwidth in both on-chip and off-chip communications. In order to improve the overall system performance without drastically increasing communication resources and complexity, communication protocol, and signaling technology that exploits parallelism at on- and off-chip system level, resulting in higher concurrency on a shared bus or direct link, a merged CDMA and PAM interconnect is proposed in this paper. The CDMA logic uses different PN codes to separate the information carried on different bus lines over shared bus, while PAM approach is based on multilevel signaling, so their combination offers an interesting potential solution. The transceiver circuit has been developed using Xilinx FPGA technology for realization of the CDMA coder/decoder, and 0.35 μm CMOS four layers process technology for implementation of the 4_PAM coder/decoder block. Simulation results show that the effective communication bandwidth is 1.1 Gbps, with an average power consumption of 25 mW per single channel at 200 MHz operating frequency and power supply of 3.3 V.**

***Keywords* – Signaling Interconnect, CDMA&PAM shared bus, SoC**

I. INTRODUCTION

The digital system performance consists of two parts: computation performance and communication performance. With rapid development in CMOS technology scaling, the computation performance of a chip has been increased drastically. As computation performance goes up, the required communication throughput needs also increase with the same rate. However, until now, the communication bandwidth has been scaling much more slowly. Therefore, the communication between multiple on- or off-chip semiconductor intellectual property, IP, blocks and electronic modules is becoming a dominant cost, performance, and power factor in contemporary electronic systems [1-2]. To achieve a high data transfer rate, the bus bandwidth must be increased. Increasing the bus bandwidth, however, increases the pin count and enlarges the chip area. It also leads to complicated routing between different IP blocks within the VLSI IC and among modules on the same printed circuit board.

The concept of transferring multiple bits over each symbol through modulation techniques has been proposed to efficiently solve the aforementioned problems [2-3]. Code Division Multiple Access, CDMA, and Pulse-Amplitude Modulation, PAM, technique that incorporate multilevel amplitudes rather than binary signals, with order to increase data rate, is proposed in this paper. The basic idea is as follows: If two CDMA coded consecutive bits in the sequence are grouped and converted to one of four levels, 4-PAM signaling, then each level is twice as long as a bit period, demanding only half the bandwidth required for transmission of the binary system.

II. TAXONOMY OF ON-CHIP COMMUNICATION

On-chip communication architectures, see Fig. 1, can be divided into the following three main classes:

a) Point-to-point (P2P) interconnects - pairs of processing units communicate directly over dedicated physically wired connections. Because of its simplicity, this architecture has been widely adopted in many applications. P2P communication architecture can provide the utmost in communication performance at the expense of dedicated channels among all the communicating IP pairs. However, these architectures suffer from lack of scalability in terms of high complexity, cost, and design effort. Custom interconnect (see Fig. 1), sometimes referred as ad-hoc interconnect, is simply connecting processing elements by wires when there is a necessity. On the other hand, uniform interconnect often has well defined interconnect topology, which can be precisely specified by equations or graphs [4], [5].

b) Bus architectures - long wires are grouped together to form a single physical communication channel, which is shared among different logical channels. The bus consists of a set of shared parallel wires to which various components are connected. An arbitration mechanism is used to control sharing of the bus. Only one component on the bus can have control over the shared wires at any given time in order to perform data transfer. A hierarchical shared bus (see Fig. 1) defines segmented bus architecture. Bus segments are connected via a bridge, which may buffer data. An alternative approach is to use a split-bus. This refers to a set of custom-design segmented buses. Bus segments can be interconnected in an ad-hoc manner, or based on a systematic approach [4], [5], [6].

[1]Goran S. Jovanović is with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedova 14, 18000 Niš, Serbia, Email: goran.jovanovic@elfak.ni.ac.rs

[2]Tatjana R. Nikolić is with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedova 14, 18000 Niš, Serbia, Email: tatjana.nikolic@elfak.ni.ac.rs

[3]Mile K. Stojčev is with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedova 14, 18000 Niš, Serbia, Email: mile.stojcev@elfak.ni.ac.rs

c) Network-on-Chip (NoC) - is an architecture inspired by data communication networks, such as LANs and WANs. Inter-processor communication is achieved by sending message packets between IP blocks using an on-chip packet-switched network [3], [7], [9].

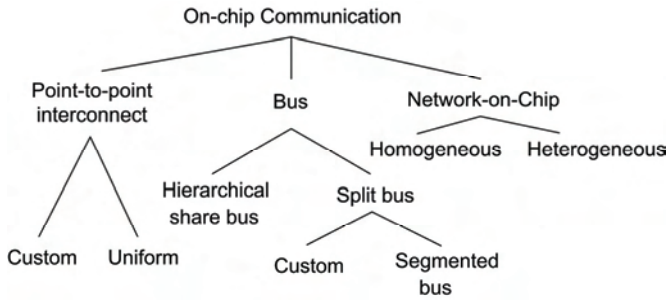


Fig. 1. Taxonomy of on-chip communication architectures

III. WIRES AND INTERCONNECTS

As interconnects have become a key limiting factor in VLSI design in deep submicron regime, chip design must first consider the constraints of interconnects in each design stage. A direct way to increase the information carrying capacity of an interconnect is to simply make the interconnect wider or faster. However, wider buses require more interconnect resources and packaging pins, and the extra drivers use more silicon real estate, dissipate more power and create more noise; faster speed leads to higher power dissipation, noise and radiation can be worse. All these issues must be considered in early decisions of bus width and speed [1], [2].

Differential signaling offers the highest signal integrity in which the common-mode noise introduced by return paths is rejected by the differential receiver. It forms the foundation of many high-performance signaling standards, including LVDS and ECL. The price paid for this is more complex drivers and receivers plus twice the amount of wires. Besides, mixed standards are used. For example, in many systems the clock is transmitted fully differential to minimize clock skew, but the signals are not differential, such signals are termed single-ended [1], [9].

Although there are different high-level abstract interconnect schemes, such as point-to-point interconnect, bus architecture, switched network, etc, the underlying physical global wire structure will be the same, namely the parallel wire structure. For point-to-point connections, each link is realized by a certain number of parallel wires; for bus architecture, the bus itself is a number of wires in parallel and the same applies to a link connecting two switches in a network-on-chip. Therefore it is of great importance to study the interconnect in general, the parallel wire structure in particular, and model them accurately. The following discussion serves this purpose.

IV. SIGNALING OVER PARALLEL WIRES

The most common method of reducing the delay over long interconnects is to insert repeaters (inverters) at appropriate points. However, delay is not the only concern associated with the interconnect. Another major issue is the bandwidth supported by the interconnect under certain constraints, such as limited area, limited power consumption, and limited freedom in choosing repeater insertion strategy. In the sequel we study how delay and bandwidth are related and derive a suitable bandwidth under different constraints.

Ideally future interconnect systems must encompass the following important features:

- ultra high data rates, usually > 100 Gbps,
- concurrent multi input-output service for simultaneous and bidirectional communications on a shared transmission medium,
- real-time re-configurability in connectivity and bandwidth for optimized channel efficiency and fault-tolerance, and
- the fabrication of interconnect systems must be compatible with the current SoC and SiP (System in Package) technologies for low-cost system production [1], [9].

The concept of transferring multiple bits over each wire has been proposed to solve the problem of high data transfer rate. This concept can be classified into two types. One is so-called bidirectional I/O interface [10], and the other is accomplished through modulation techniques [1-2]. In some applications, data and clock channels are combined as single channel to reduce the wire count. It necessitates the clock/data recovery circuits to extract the clock from the incoming data stream [2]. On the other hand, by reconsidering the representations of binary data, several kinds of modulation techniques were addressed, including the pulse-width modulation, PWM [11], a combinational of PWM and PAM signaling technology [5], code division multiple access, CDMA [12], frequency division multiple access, FDMA [9], and CDMA/TDMA (Time Division Multiple Access) [13].

In this paper a 1.1 Gbps interface circuit intended for data transfer over single interconnect channel using a combination of CDMA and PAM scheme is presented. In order to reduce the number of wires over the bus and increase the bus bandwidth the binary data are first CDMA coded and after that 4-PAM which incorporates multilevel amplitudes rather than binary signals is implemented.

A. Signaling issues and comparison

To achieve higher computational performance at the system level, instead of just increasing processor clock speed, parallelism in computation is widely used for chip multiprocessing, CMP, and symmetric multiprocessor, SMP, system design. These systems achieve computational concurrencies by using scheduling techniques such as multitasking and multithreading. However, the on- and off-chip interconnect communication architectures are still mostly based on conventional time-division multiplexing (TDM)-based communication protocols (typically known as time-division multiple access (TDMA) or time interleaving) that do not allow real communication parallelism. So the key motivation of this paper is to explore how to exploit parallelism and concurrency in communication without increasing resources complexity. This paper suggests how the

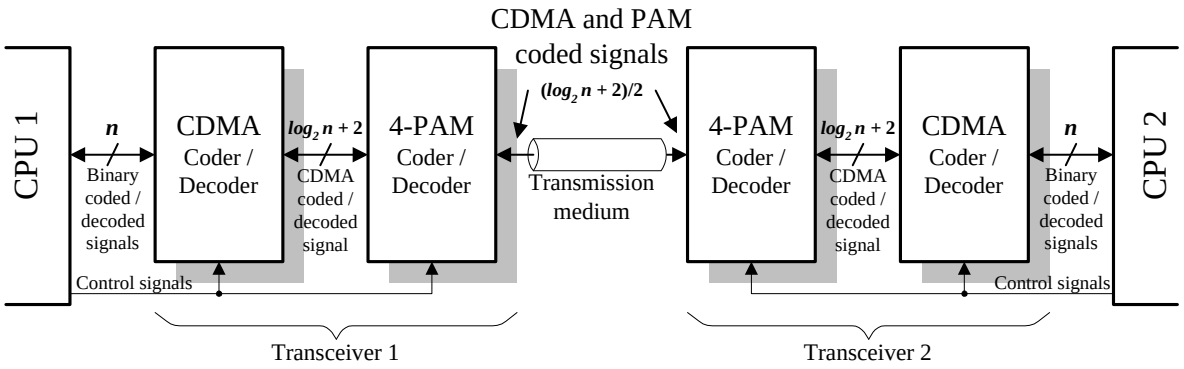


Fig. 2. Block diagram of the CDMA and PAM signaling interconnect structure

standard I/O interfaces should be modified for higher system performance with less cost. Using our results, designers of systems with multiple heterogeneous semiconductor-based processing devices will be able to take advantage of parallelism at both the computation and communication levels.

Current- or voltage-mode 2-PAM binary signaling has been used most widely in applications for high-speed on- and off-chip communications due to its simplicity, high noise immunity, and reliability. 4-PAM multilevel signaling has also been introduced, since its lower baud rate can double the eye width and also reduce frequency-dependent channel losses without increasing bus speed. However, for a given power, 4-PAM is more susceptible to channel interference noise such as crosstalk, since the multiple signal levels are more closely spaced than 2-PAM, resulting in higher BEER. Also the effective eye width is further reduced due to transitions between signal levels that are not adjacent.

V. MERGED CDMA AND PAM INTERCONNECT ARCHITECTURE

A block diagram of the merged CDMA and PAM signaling interconnect architecture, called CPIA, intended for data transfer over transmission medium, such as for example SoC and multiprocessor shared bus, is pictured in Fig. 2. Both building blocks, the CDMA– and 4_PAM–coder/decoder are constituents of a single transceiver. During data transfer from CPU_1 to CPU_2 the CDMA-coder/decoder of a Transceiver_1 is driven with n bus binary coded signals. At its output it generates $\log_2 n + 2$ binary coded signals. The 4_PAM coder converts two-level input signals to four-level output signals and send them through a transmission medium over $(\log_2 n + 2)/2$ lines. At the other end of a transmission line the decoder logic of a Transceiver_2 performs an opposite action.

In the sequel we will briefly describe the hardware structure of the CDMA-coder/decoder and 4_PAM-coder/decoder blocks.

B. CDMA-coder/decoder

The encoding scheme is illustrated in Fig. 3(b). Data from different bus lines drive the encoder. Each data bit is spread into S bit encoded data generated by XOR operations with a unique S-bit spreading code. Each bit of the S-bit encoded data generated by XOR operations is called a data chip. Then, the data chips which come from different data bus lines are added together arithmetically according to their bit positions in the S-bit sequences. After the add operations, we will get S sum values of S bit encoded data. Binary equivalents of the sum values are transferred over $\log_2 n + 2$ - bus lines to the 4_PAM coder/decoder block.

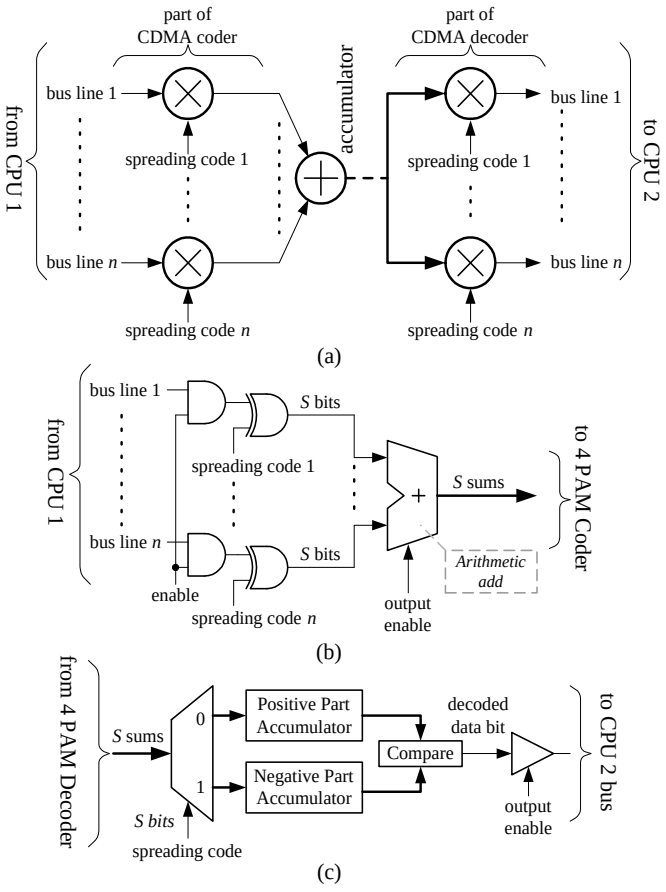


Fig. 3. CDMA coder/decoder: (a) CDMA technique principle; (b) Digital CDMA encoding scheme; (c) Digital CDMA decoding scheme

The decoding scheme is sketched in Fig. 3(c). The decoder accumulates the received sum values into two separate parts, a positive part and a negative part, according to the bit value of the spreading code used for decoding. After accumulating the sum values, the original data bit can be decoded by comparing the values between the two accumulators. For more details concerning the hardware structure of the CDMA coder/decoder see Reference [14].

C. 4 PAM coder/decoder

The encoding scheme of the 4_PAM coder is pictured in Fig. 4(b). It consists of three building blocks, 2-to-4 Decoder, Buffer-Stage, and Output-Stage. The 2-to-4 Decoder converts 2-bit CDMA coded signals, D_i and D_{i+1}, into four binary signals P_0, P_1, P_2 and P_3. At a given instant only one of the signals from P_0 up to P_3 is active. The Buffer-Stage is used for driving the Output_Stage. A control signal EN enables (disables) the Output_Stage, i.e. when EN=0 at a pin Out we have high-impedance. The Output-Stage is composed of four switches, referred as S_0, S_1, S_2 and S_3. M=8 points to the fact that a corresponding switch is realized with 8 transistors that operate in parallel. In order to generate a multi-level signal at the pin Out we have that $V_{D3} > V_{D2} > V_{D1} > 0$ V. Switches S_3 and S_0 are realized as a single level–, while S_2 and S_1 as double level–switches. The voltage level at the pin Out is defined according to the following Table I.

TABLE I
VOLTAGE LEVEL AT PIN OUT

switch				Out
S_3	S_2	S_1	S_0	
off	off	off	off	High-impedance
off	off	off	on	0 V
off	off	on	off	1.1 V
off	on	off	off	2.2 V
on	off	off	off	3.3 V

The decoding scheme of the 4_PAM decoder is given in Fig. 4 (c). The comparator block and thermometer coder logic are its basic building blocks. Principle of the thermometer coder logic is described according to the truth Table II.

TABLE II
THERMOMETER TRUTH TABLE

inputs			outputs	
o3	o2	o1	out2	ouit1
0	0	0	0	0
0	0	1	0	1
0	1	1	1	0
1	1	1	1	1

The Comparator block is composed of three identical comparator units C_1, C_2 and C_3.They compare the input signal, which feed their non-inverting inputs, with referent voltages derived from the resistor network. In our case we have V_1=0.5 V, V_2= 1.65 V, and V_3= 2.75 V. All comparator units are active during the time interval when the clock pulse CLK is positive.

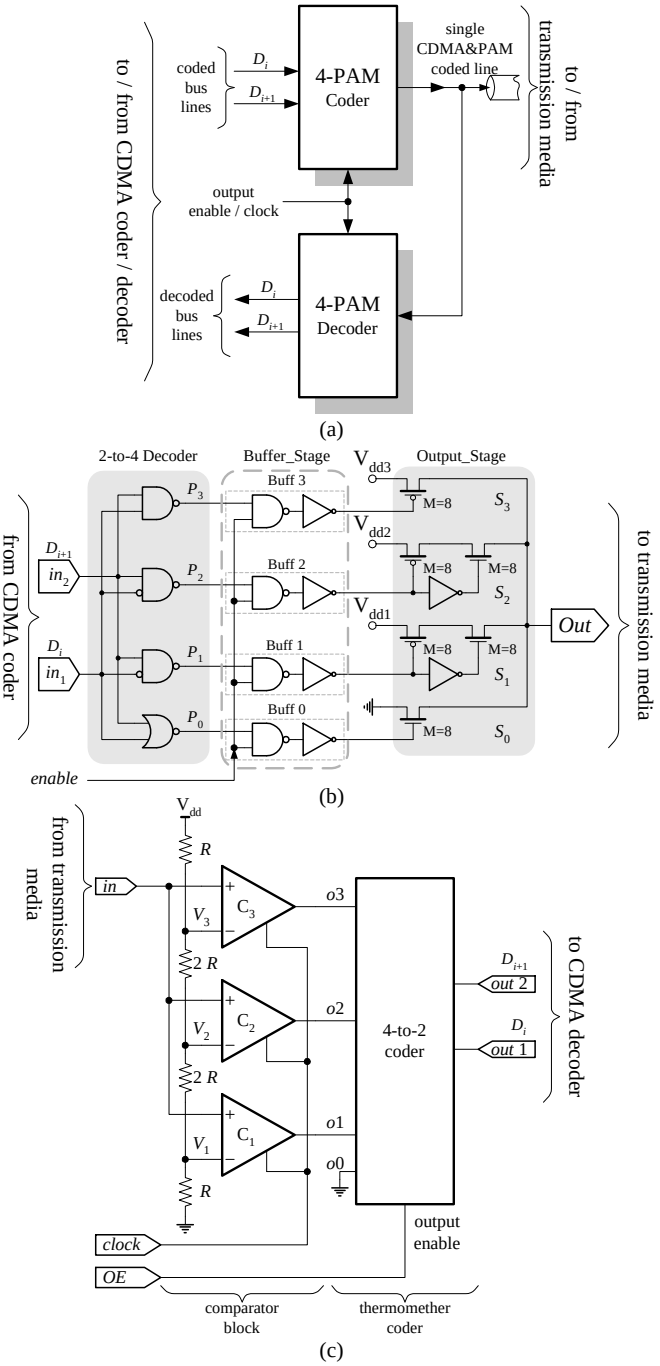


Fig. 4. PAM coder/decoder; (a) PAM coder/decoder structure; (b) 4 PAM coder; (c) 4 PAM decoder

D. Comparator block architecture

A detailed hardware structure of the Comparator block is given in Fig. 5. It consists of two constituents, Linear-Block, LB, and Digital-Block, DB.

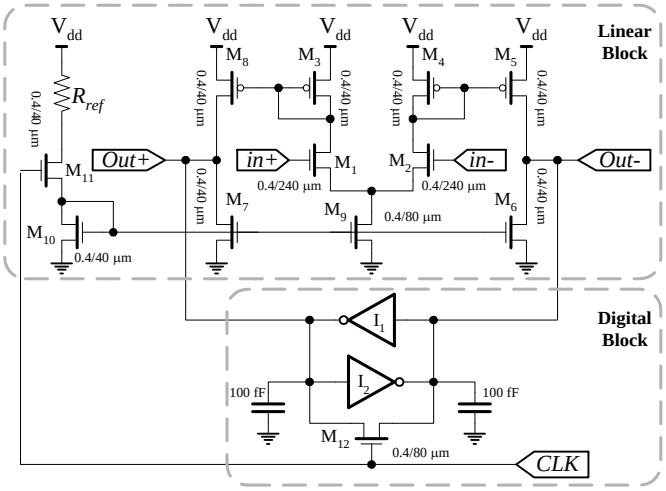


Fig. 5. Hardware structure of the Comparator block

The LB is composed of a differential stage M_1 and M_2, a current sink M_9, two pairs of current mirrors M_3–M_8 and M_4–M_5, two pairs of output stages M_5–M_6 and M_7–M_8, and a bias circuit M_{10}–M_{11}. The transistor M_{11} acts as a switch. M_{11} is in on-state when *CLK*=1, during which time period it provides condition for correct operation of the LB.

The DB logic consists of two weak-inverters I_1 and I_2. When *CLK*=1, M_{12} is on, the inverters I_1 and I_2 enter in linear mode of operation. During this period capacitors C_1 and C_2 are charged to values present at the LBs' outputs Out+ and Out-, respectively. When *CLK*=0, M_{12} is off, the LB switches to inactive mode of operation, I_1 and I_2 remain in active state and maintain the charges of C_1 and C_2 unchangeable.

VI. Experimental results

In order to estimate the performance attainable with the multilevel signaling method the design solution of a merged CDMA and PAM interconnect architecture has been implemented using: a) Xilinx FPGA technology for realization of the CDMA coder/decoder block; and b) conventional 0.35 µm length 2 polysilicon 4 metal layers with maximum voltage range of 3.5 V CMOS technology [15] for realization of the 4_PAM coder/decoder block. Starting from the fact that the proposed architecture consists of two different building blocks connected in cascade, a completely digital CDMA coder/decoder and a mixed-signal 4_PAM coder/decoder, various optimization techniques were used for their performance evaluation.

As a first, the CDMA coder/decoder constituent was realized as wrapper interface logic located between the CPU and 4_PAM coder/decoder block. It was described at RTL level using VHDL. For synthesis, routing and mapping a Xilinx development CAD tool WebPack 9.2i was used [16]. Design verification was performed using testbenches intended for parallel excitation of all data lines. The wrapper was implemented on FPGA devices from Spartan 2, Spartan 3, Virtex 4, Virtex 5 and Virtex E series. A pair of CDMA coder/decoder block seems to occupy appropriate space, in average 2000 equivalent gates, considering CPU cost of about 30 000 gates, what is less than 8% of hardware overhead per CPU. Let note that the communication bandwidth of the CDMA coder/decoder is much higher with respect to that of the 4_PAM coder/decoder circuit. Bearing this in mind these performance metrics will be omitted here. More details concerning performance of the CDMA coder/decoder logic are given in Reference [14].

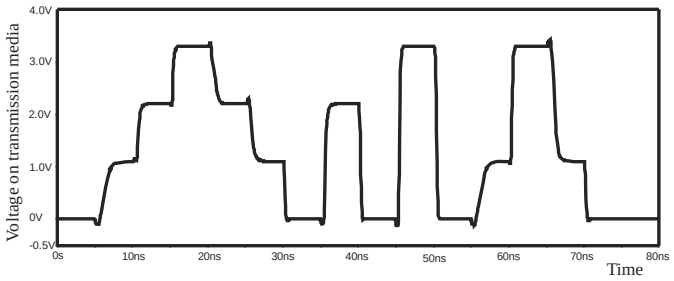


Fig. 6. Voltage on transmission media

The design of a 4_PAM coder/decoder circuit for mixed-signals was verified using HSpice simulation. Model level 47 that correspond to 0.35 µm CMOS technology was used. During simulation several different supply V_{DD} voltages (1.1 V, 2.2 V and 3.3 V) needed for correct operation of the Output-Stage were used. The supply voltage for all other logic circuits was fixed to 3.3 V. Fig. 6 deals with HSpice simulation which relates to dynamics of multi-level signals at the output of the Output-Stage. All types of signal transitions are presented.

TABLE III

SIGNAL TRANSITIONS AT THE NEAR END OF A TRANSMISSION LINE

Transition	Capacitive load *C* = 1 pF			
	raising edge [ps]	raising edge delay [ps]	falling edge [ps]	falling edge delay [ps]
0 ⇔ 1	619	960	189	183
0 ⇔ 2	190	552	234	275
0 ⇔ 3	170	454	212	408
1 ⇔ 2	210	704	347	655
1 ⇔ 3	223	409	344	829
2 ⇔ 3	196	361	549	900
	Capacitive load *C* = 3 pF			
0 ⇔ 1	939	1175	219	216
0 ⇔ 2	308	615	252	317
0 ⇔ 3	223	504	257	446
1 ⇔ 2	337	751	533	796
1 ⇔ 3	251	445	539	967
2 ⇔ 3	244	394	718	956
	Capacitive load *C* = 5 pF			
0 ⇔ 1	1280	1300	255	240
0 ⇔ 2	430	670	290	350
0 ⇔ 3	270	540	305	480
1 ⇔ 2	460	800	710	880
1 ⇔ 3	280	470	740	1090
2 ⇔ 3	290	420	890	1000

TABLE IV
DELAYS INVOLVED BY THE RECEIVER

Transition	raising edge [ps]	falling edge [ps]
0 ⇔ 1	446	0
0 ⇔ 2	0	445
0 ⇔ 3	437	192
1 ⇔ 2	0	193
1 ⇔ 3	444	203
2 ⇔ 3	439	448

The obtained results, presented in Table III for three different capacitive load of a transmission line (C=1 pF, 3pF, and 5 pF), relate to: i) pulse rising (falling) time (from 10% up to 90% pulse amplitude); and ii) pulse delay of a rising (falling) edge (50% to 50% pulse amplitude). By analyzing the results given in Table III we can conclude that maximum delays for rising edges are obtained for 0↔1 signal transitions, while for falling edges for transitions 1↔3 and 2↔3. All these case relate to *C*=5 pF capacitive load of transmission line. Maximum delay for falling edge is obtained for transitions 1↔3 and 2↔3. According to Table III we conclude that largest delays are obtained for 0↔1, 1↔3 and 2↔3 signal level transitions.

Delays involved by the receiver for all types of signal transitions are given in Table IV. In Table V the power consumption of a transmitter-receiver pair (single channel) for 200 MHz operating frequency and V_{DD}= 3.3 V are presented.

TABLE V
POWER CONSUMPTION OF A TRANSMITTER-RECEIVER

Capacitive load C_L [pF]	Power consumption [mW]
1	25.51
3	25.92
5	26.37

According to the results given in Table III and Table IV we can conclude that the maximal operating frequency of a transceiver is 571 MHz, what corresponds to 1.75 ns signal delay (including delay of a transmitter, receiver and transmission line loaded with C=5 pF). This implies that a communication bandwidth of 1.1 Gbps per single channel can be achieved.

VII. CONCLUSION

The goal of the on- and off-chip interconnect designs is to obtain the optimum combination of bandwidth, latency, number of wires and cost. Traditional inter-chip and intra-chip communications are based solely on TDMA approach. To overcome the limitations of the TDMA techniques a number of new interconnect schemes, mainly based on CDMA, PAM, PWM and FDMA have been investigated recently in order to increase the data rate, concurrency, and reduce latency and power consumption. In this paper a combined CDMA and PAM signaling interconnect architecture is proposed. Use of this method allows the required number of bus lines, n, for a given bandwidth to be reduced to $(\log_2 n + 2)$. The CDMA logic has been implemented in Xilinx FPGA family while the PAM interface in 0.35 μm CMOS technology. Simulation results show that the proposal can operate at 1 Gbps per single channel (wire) over transmission line loaded with C = 5pF. The scheme is generally applicable to any IP block – to – IP block link within the shared SoC bus or chip-to-chip link that can be implemented in a point – to – point fashion.

REFERENCES

[1] J. Kim, I. Verbauwhede, M-C Frank Chang, "Design of an Interconnect Architecture and Signaling Technology for Parallelism in Communication", IEEE Transactions on Very Large Scale Integration (VLSI) Systems, Vol. 15, No. 8, pp.881-893, August 2007.

[2] S. Pasrisha, N. Dutt, *On-Chip Communication Architectures: System-on-Chip Interconnect*, Morgan Kaufmann, Boston, 2008.

[3] D. Foley, M. Flynn, "A Low-Power 8-PAM Serial Transceiver in 0.5 μm Digital CMOS", IEEE Journal of Solid-State Circuits, Vol. 37, No. 3, pp.310-316, March 2002.

[4] TST. Mak, P. Sedcole, P. Y. K. Cheung, W. Luk, "On-FPGA communication architectures and design factors", In: Proc. International Conference on Field Programmable Logic and Applications, Vol 1, pp. 1-8, Madrid, 2006.

[5] M. Mitic, M. Stojcev, "An overview of on-chip buses", Facta Universitatis, Series: Electronics and Energetics, Vol.19, No.3, pp.405-428, November 2006.

[6] S. Dey, K. Lahiri, A. Raghunathan, "Design of communication architectures for high-performance and energy-efficient systems-on-chips", In: Jerraya AA, Wolf W, editors. Multiprocessor Systems-on-Chips, Morgan Kaufman, pp. 187-222, San Francisco, 2004.

[7] L. Benini, G. De Micheli, *Networks on Chips*, Elsevier, Boston, 2006.

[8] A. Jantsch, H. Tenhunen, Introduction. In: Jantsch A, Tenhunen H, editors. *Networks on Chip*, Kluwer Academic Publishers, pp. 1-24, Boston, 2003.

[9] M-c Frank Chang, I. Verbauwhede, C. Chien, Z. Xu, J. Kim, J. Ko, Q. Gu, B-C Lai, "Advanced RF/Based Interconnect Schemes for Inter- and Intra-ULSI Communications", IEEE Transactions on Electron Devices, Vol. 52, No. 7, pp.1271-1285, July 2005.

[10] J.Y. Sim, Y.S. Sohn, S.C. Heo, H.J. Park, S.I. Cho, "A 1-Gb/s Bidirectional I/O Buffer Using the Current Mode Schemes", IEEE Journal of Solid-State Circuits, Vol. 34, No. 4, pp.1271-1285, April, 1999.

[11] W-H. Chen, G-K. Debug, J-W. Chen, S-I. Liu, "A CMOS 400/Mb/s Serial Link for AS-Memory Systems Using a PWM Scheme", IEEE Journal of Solid-State Circuits, Vol. 36, No. 10, pp. 1498-1505, October 2001.

[12] X. Wang, T. Ahonen, J. Nurmi, "Applying CDMA Technique to Network-on-Chip", IEEE Transactions on Very Large Scale Integration (VLSI) Systems, Vol. 15, No. 10, pp.1091-1100, October 2007.

[13] B. C. Lai, P. Schaumont, I. Verbauwhede, CT-bus, "A Heterogeneous CDMA/TDMA Bus for Future SoC", In: Proc. 38th Annu. Asilomar Conference on Signals, Systems, and Computers, Vol. 2, pp. 1868-72, 2004.

[14] T. Nikolić, M. Stojčev, G. Đorđević, "CDMA Bus-Based On-Chip Interconnect Infrastructure", Microelectronics Reliability, Vol. 49, No.4, pp.448-459, April 2009.

[15] 0.35 μm CMOS SPICE Parameters, available at https://www.eecs.berkeley.edu/~boser/courses/software/spice/CMOS35/index.html, February 2010.

[16] ISE 9.2i Quick Start Tutorial Xilinx ISE 9.2i, Software Manuals Helps, http://www.xilinx.com/support/sw_manuals/xilinx92/download

Application Control on Quality of Service Resource for Multimedia Sessions

Evelina Pencheva[1] and Ivaylo Atanasov[2]

***Abstract* - The paper investigates the capabilities for open access to quality of service (QoS) management in convergent networks. By analysis on policy and charging control functions in Internet Protocol Multimedia Subsystem (IMS), requirements for third party application control on QoS are identified. The application view on authorized QoS resources for SIP session is modeled. The behavioral equivalence of the authorized QoS resources model and SIP session state model is proved formally.**

***Keywords* - open access, policy and charging control, labeled transition systems, behavioral equivalence**

I. INTRODUCTION

Internet Protocol Multimedia Subsystem (IMS) is defined as access independent IP-based service control architecture, aimed to provide all kinds of multimedia. The requirement for IMS in the conjunction with the underlying IP connectivity access network and transport network (IP-CAN) is to provide quality of service. Quality of service (QoS) is used to differentiate multimedia offering from traditional Internet services which in most cases do not provide QoS.

Via the IMS, the user equipment negotiates its capabilities and QoS parameters during session setup or modification, using Session Initiation Protocol (SIP). The capability to authorize and control the usage of QoS resources, intended for IMS media and based on SIP signaling parameters, is called IP policy control. The IP policy control can increase utilization of bearer resource by controlling parameters in the access and transport networks. To ensure coherent charging between IP-CAN and IMS, the IP policy control is harmonized with charging control, and the overall interaction between the IMS and the IP-CAN network is called Policy and Charging Control (PCC). The PCC encompasses the following high level functions for IP-CANs (e.g. GPRS, WLAN, Fixed broadband, etc.): flow based charging, including charging control and online credit control, and policy control e.g. gating control, QoS control, user plane event reporting and network initiated IP-CAN bearer establishment. Delegating the QoS resource authorization in wireless networks to the IMS layer provides a number of benefits such as session and service continuity.

One of the design principles of IMS is separation of applications from generic service control functionality and provision of open interfaces. The open access to network functions allows third party applications to invoke communication functions. Instead of using control protocols, the applications access network resources via application programming interfaces (APIs).

The necessity of open access to QoS control is substantiated in [1]. The authors of [2] present a protocol-based approach to QoS management, where the collected information about the network is provided to external applications in a structurally standardized format. In [3], the authors suggest a solution in which the QoS management functions are delegated to a node monitoring the signaling interchange by the user equipment and the network.

The aim of the paper is to present an approach to modeling the third party application view on QoS resources authorized for SIP session. Based on the analysis of PCC architecture, we identify the required functionality for open access to QoS management. We prove formally that the application view on the QoS resources state conforms to the SIP session establishment with authorization of QoS resources.

The paper is structured as follows. In Section II, we summarize the requirements for open access to QoS management based on the PCC framework. Section III describes some of the functions for third party control on QoS resources and their mapping onto SIP control protocol. The application view on QoS resources authorized for multimedia session is modeled in Section IV. In Section V, we prove the behavioral equivalence of the suggested model of QoS resources as seen by application and the SIP session state machine. Finally we conclude the paper.

II. REQUIREMENTS FOR OPEN ACCESS TO PCC

The PCC includes mechanisms for controlling the bearer traffic by using IP policies.

During session establishment and modification in IMS, the user equipment negotiates a set of media characteristics. If an operator applies policy control, the relevant session description information is used to form IP QoS authorization data. The third party application can be involved in QoS authorization by requesting specific QoS parameters to be applied, modified or removed.

During resource reservation the requested QoS parameters are compared with the authorized QoS parameters and PCC rules for opening the gate for incoming and outgoing traffic are generated. Opening the gate command could be sent as a response to an initial authorization request from the media gateway or it can be sent as a standalone decision. If a standalone decision is used, then operator can ensure that user-plane resources are not used before the IMS session is finally accepted. In this case, user will lose all announcements which are to be delivered before completing the session, as the media

[1]Evelina N. Pencheva is with the Faculty of Telecommunications, TU-Sofia, 7 Kl. Ohridski blvd, 1000 Sofia, Bulgaria, E-mail: enp@tu-sofia.bg

[2]Ivaylo I. Atanasov is with the Faculty of Telecommunications, TU-Sofia, 7 Kl. Ohridski blvd, 1000 Sofia, Bulgaria, E-mail: iia@tu-sofia.bg

gateway will drop all incoming user-plane IP packets. The responsibility for opening the gate can be transferred to the third party application. The revoke function is used to force the release of previously authorized bearer resources in the IP-CAN. The third party application can use the function to prevent bearer misuse after session termination.

Any QoS events are reported and it is possible to track status of the IMS signaling and user plane bearers which the user equipment currently uses. To receive notifications about QoS events the third party application needs to manage its subscriptions for notifications. By using information about bearer and signaling path status, the application can improve service execution. The QoS-related event types include the following:

- loss of bearer that can result in QoS degradation (if all the bearers are lost, the application can request QoS resource release);
- recovery or establishment of a new bearer;
- IP-CAN change which may result in applying specific charging;
- out of credit indicates that the user credit limit is reached;
- normal connection termination;
- successful session establishment.

III. Mapping of QoS management functions onto SIP protocol

APIs for open access to QoS management can be defined following the identified requirements. The application server that provides APIs for application-managed QoS has to support control protocols also. The interface between the application server and IMS entity responsible for session management is SIP-based. SIP session information (including QoS parameters) is described by means of Session Description Protocol (SDP) and is transferred in the SIP message body. The initial request is sent as SIP INVITE message. Session modification during establishment is requested by UPDATE message, and re-INVITE message is used for modification of established session. Information about SIP session is transferred by INFO message. The initial filter criteria for application triggering are stored as a part of user profile and are provided on user registration. Preliminary condition for QoS related notifications is an agreement between network operator and service provider for notification authorization. The third party application subscribes to QoS events and the application server requests QoS monitoring and reporting. Table 1 summarizes the mapping of application-managed QoS functions onto SIP messages.

Table 1 Application-managed QoS Functions and Mapping onto SIP Messages

Function	*Description*	*SIP message*
notifyQoSEvent	notifies the application about occurrence of QoS events	INFO
reserveSpecificQoS	authorizes and makes reservation for QoS resources	INVITE, UPDATE
qosApproval	requests opening the gate for user traffic	INFO
modifySpecificQoS	modifies temporary established QoS	re-INVITE
removeSpecificQoS	removes the temporary established QoS	re-INVITE
releaseQoSResources	releases QoS resources	BYE
getQoSInfo	retrieves information related to the temporary QoS features	-
deassignQoSResources	releases the relationship between the application and the QoS resources and associated objects	-

IV. Application view on quality of service resources

The application view on the states of QoSResources object has to be synchronized with the SIP session states. From the application point of view, the QoSResources object is in one of the following states, as shown in Fig.9. In Null state, the QoSResources object is created but no QoS resources are authorized for the session. In QoSAuthorized state, the rights for establishment of sessions with specific QoS features are verified and QoS resources are allocated. Any QoS modifications requested by the application require re-authorization of QoS resources. Reporting of QoS events might result in temporary QoS degradation or release. For example, if all media flows are deactivated but the BYE message is not received, the application can request session release. Notifications about access network change can affect charging, if access specific services are provided. In ApplicationQoSReleased state, the application has requested to release the QoSResources object, the media gateway collects the possible QoS information. In case the application has not requested additional QoS related information, the QoSResources object is destroyed immediately. In NetworkQoSReleased state, the authorized QoS resources are released and the media gateway can return QoS information to the application if requested. In this state, the application can release the QoS resources object by calling the deassignQoS-Resources method. Note that the application has to release the object itself as good practice requires that when an object was created on behalf of a certain entity, the entity is also responsible for destroying it when the object is no longer needed.

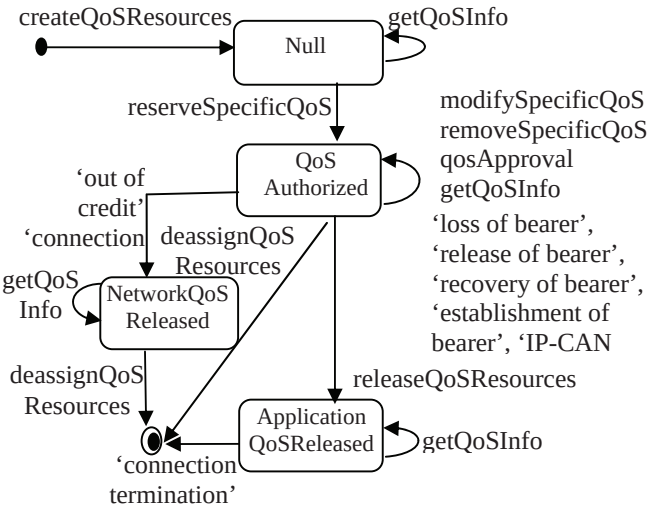


Fig.1 Application view on QoSResources object

V. CONFORMANCE OF APPLICATION VIEW ON QOS RESOURCES TO SIP SESSION STATE

The application server that supports APIs for QoS management toward SIP-based network has to maintain two mutually synchronized state machines – one for the application view on the states of QoS resources and another for the SIP session states. Both state machines have to expose behavioral equivalence, i.e. their observable behavior should be equivalent.

To prove behavioral equivalence between state machines formally, the notion of *labeled transition systems* is used [4].

Definition 1: A *Label Transition System* (LTS) is a quadruple $(S, Act, \rightarrow, s_0)$, where S is countable set of states, Act is a countable set of elementary actions, $\rightarrow \subseteq S \times Act \times S$ is a set of transitions, and $s_0 \in S$ is the set of initial states.

A labeled transition system $(S, Act, \rightarrow, s_0)$ is called *finitely branching*, if for any state $s \in S$, set $\{(a, s') \mid a \in Act, (s, a, s') \in \rightarrow\}$ is finite. The labeled transition systems we consider in the paper are only labeled transition systems that are finitely branching.

We use the following notations:

- $s \xrightarrow{a} s'$ stands for the transition (s, a, s');
- $s \xrightarrow{a}$ means that $\exists\, s' : s \xrightarrow{a} s'$;
- $s \overset{\mu}{\Longrightarrow} s_n$, where $\mu = a_1, a_2, ..., a_n : \exists\, s_1, s_2, \ldots, s_n$, such that $s \xrightarrow{a1} s_1 \ldots \xrightarrow{an} s_n$;
- $s \overset{\mu}{\Longrightarrow}$ means that $\exists\, s'$, such as $s \overset{\mu}{\Longrightarrow} s'$;
- $\overset{\hat{\mu}}{\Longrightarrow}$ means $\Longrightarrow$ if $\mu \equiv \tau$ or $\overset{\mu}{\Longrightarrow}$ otherwise.

Definition 2: Two labeled transition systems $T = (S, A, \rightarrow, s_0)$ and $T' = (S', A, \rightarrow', s_0')$ are *weakly bisimilar* if there is a binary relation $U \subseteq S \times S'$ such as if $s_1 \, U \, t_1 : s_1 \subseteq S$ and $t_1 \subseteq S'$ then $\forall a \in Act$:

- $s_1 \overset{a}{\Longrightarrow} s_2$ implies $\exists\, t_2 : t_1 \overset{\hat{a}}{\Longrightarrow}' t_2$ and $s_2 \, U \, t_2$;
- $t_1 \overset{a}{\Longrightarrow}' t_2$ implies $\exists\, s_2 : s_1 \overset{\hat{a}}{\Longrightarrow} s_2$ and $s_2 \, U \, t_2$.

We present the SIP session state machine and the QoS resources state model as seen by third party application with labeled transition systems which are intentionally simplified.

With $T_{SIPQ} = (S_{SIPQ}, Act_{SIPQ}, \rightarrow_{SIPQ}, s_0)$ is denoted a labeled transition system which represents a simplified SIP session state model where

- S_{SIPQ} = { $Idle_{SIP}$, SessionProgress, WaitPRack, PRacked, WaitUpdate, Updated, Wait180, Alerting, WaitAck, Established, $Wait200_{INFO}$, $Wait200_{INVITE}$, $Wait200_{BYE}$ };
- Act_{SIPQ} = {INVITE, BYE, 183, PRACK, 200_{PRACK}, UPDATE, INFO, 180, 200_{INVITE}, $200_{INFO,}$ 200_{UPDATE}, 200_{BYE}, ACK};
- $\rightarrow_{SIPQ}$ = {$Idle_{SIP}$ INVITE SessionProgress, SessionProgress 183 WaitPRack, WaitPRack PRACK PRacked, PRacked 200_{PRACK} WaitUpdate, WaitUpdate UPDATE Updated, Updated 200_{UPDATE} Wait180, Wait180 180 Alerting, Alerting 200_{INVITE} WaitAck, WaitAck ACK Established, Established INFO $Wait200_{INFO}$, $Wait200_{INFO}$ 200_{INFO} Established, Established INVITE $Wait200_{INVITE}$, $Wait200_{INVITE}$ 200_{INVITE} Established, Established BYE $Wait200_{BYE}$, $Wait200_{BYE}$ 200_{BYE} $Idle_{SIP}$};
- s_0 = {$Idle_{SIP}$}.

In the application view on QoS resources, there are no authorized QoS resources for the SIP session in the Null, NetworkQoSReleased, and ApplicationQoSReleased states. Hence, from session control point of view, these states share a common abstraction and can be labeled by the same label.

With $T_{RM} = (S_{RM}, Act_{RM}, \rightarrow_{RM}, s_0')$ is denoted a labeled transition system representing the application view on the states of QoS resources authorized for a SIP session where:

- S_{RM} = { Null, QoSAuthorized};
- Act_{RM} = {reserveSpecificQoS, modifySpecificQoS, removeSpecificQoS, qosApproval, releaseQoSResources, reportNotificaton};
- $\rightarrow_{RM}$ = {Null reserveSpecificQoS QoSAuthorized, QoSAuthorized modifySpecificQoS QoSAuthorized, QoSAuthorized removeSpecificQoS QoSAuthorized, QoSAuthorized qosApproval QoSAuthorized, QoSAuthorized reportNotification QoSAuthorized, QoSAuthorized reportNotification Null, QoSAuthorized releaseQoS Null};
- $s_{0'}$ = {Null}.

The transitions triggered by reportNotification() method report QoS event enumerated in Table 1 and shown as informal text in Fig.1. The getQoSInfo method does not affect the session state and QoS resource state, so it is excluded when proving behavior equivalence.

Proposition: The labeled transition systems T_{RM} and T_{SIPQ} are weakly bisimilar.

Proof: To prove the bisimulation relation between two labeled transition systems, it has to be proved that there is a bisimulation relation between their states. With U it is denoted a relation between the states of T_{RM} and T_{SIPQ} where U = {(Null, $Idle_{SIPQ}$), (QoSAuthorised, Established)}. Table 2 presents the bisimulation relation between the states of T_{RM} and T_{SIPQ}.

The mapping between the methods of application-managed QoS interfaces and SIP messages in Table 1 shows the action similarity. Based on the bisimulation relation between the states of T_{RM} and T_{SIPQ} it can be stated that both systems expose equivalent behavior.

VI. Conclusion

The open network interfaces for third party software developers create opportunities for provisioning of new customer-oriented value-added services. The required functionality for open access to QoS management might be derived from the functional architecture of policy and charging control in IP-based multimedia networks.

The application server that provides APIs for QoS management has to maintain the state machines both representing the authorized QoS resource for the session as seen by application and SIP session state. The open access to QoS control provides more flexibility in resource management as far as the QoS provisioning is one of the main requirements to convergent networks.

TABLE 2 BISIMULATION RELATION BETWEEN QOS RESOURCES STATES AND SIP SESSION STATES

Transitions in T_{RM}	*Transitions in T_{SIPQ}*
Null reserveSpecificQoS QoSAuthorized	$Idle_{SIPQ}$ INVITE SessionProgress SessionProgress 183 WaitPRack, WaitPRack PRACK PRacked, PRacked 200_{PRACK} WaitUpdate, WaitUpdate UPDATE Updated, Updated 200_{UPDATE} Wait180, Wait180 180 Alerting, Alerting 200_{INVITE} WaitAck, WaitAck ACK Established
QoSAuthorized modifySpecificQoS QoSAuthorized	Established INVITE $Wait200_{INVITE}$, $Wait200_{INVITE}$ 200_{INVITE} Established
QoSAuthorized	Established INVITE
removeSpecificQoS QoSAuthorized	$Wait200_{INVITE}$, $Wait200_{INVITE}$ 200_{INVITE} Established
QoSAuthorized commitSpecificQoS QoSAuthorized	Established INFO $Wait200_{INFO}$ $Wait200_{INFO}$ 200_{INFO} Established
QoSAuthorized reportNotifications QoSAuthorized	Established INFO $Wait200_{INFO}$, $Wait200_{INFO}$ 200_{INFO} Established
QoSAuthorized reportNotifications Null	Established INFO $Wait200_{INFO}$, $Wait200_{INFO}$ 200_{INFO} $Idle_{SIPQ}$
QoSAuthorized releaseQoS Null	Established BYE Wait200 Wait200 200_{BYE} $Idle_{SIPQ}$

Acknowledgement

The work is conducted under the grant of the Project DO 02-135/2008 funded by Bulgarian Ministry of Education and Science.

References

[1] Elkotob, M., (2008), Autonomic Resource Management in IEEE 802.11 Open Access networks, Dissertation, Lules University of Technology, Sweden, Retrieved from http://epubl.ltu.se/1402-1757/2008/38/LTU-LIC-0838-SE.pdf, 2009

[2] Katchabaw, M., Lutfiyya, H., & Bauer, M., (2005) Usage based service differentiation for end-to-end quality of service management. *Computer Communications* vol 28, issue 18, November 2005, pp. 2146-2159

[3] Santoni, D., & Katchabaw, M., (2007), Resource Matching in a Peer-to-Peer Computational Framework. *Proc. of the International Conference on Internet Computing*, Las Vegas, Nevada, USA, 2007, pp.89-95

[4] Panangaden P., (2009), Notes on Labeled Transition Systems and Bisimulation, http://www.cs.mcgill.ca/~prakash/Courses/comp330/Notes/lts09.pdf.

Implementing Algorithms for the Binary (0-1) Knapsack Problem

Dušan B. Gajić[1]

***Abstract* – The knapsack problem is one of the most important and, at the same time, most fascinating NP-complete problems in the field of combinatorial optimization. From a practical point of view, it is interesting to estimate which knapsack problem instances can be efficiently handled, in terms of both qualitative and quantitative parameters. Another problem is to determine how much decisions made on implementation issues, e.g., application of code tuning techniques and compiler optimizations, affect actual performance of algorithms for the knapsack problem. This paper presents the results and conclusions drawn from the computational experiments done with C/C++ implementations of both classical and recent algorithms for the binary (0-1) knapsack problem. These conclusions can be useful as guidelines in practical applications. Further, analyzing these results could be helpful in creating directions for future research.**

***Keywords* – binary (0-1) knapsack problem, computational experiments, C/C++ algorithm implementations, code tuning techniques, compiler optimizations**

I. INTRODUCTION

When discussing algorithm performance in computer science, we are accustomed to dealing with terms that inhabit the realm of mathematical abstraction, like the asymptotic notation and the worst-case running time. But what is the real-world performance of algorithms that are actually implemented and executed on a computer? What types of problem instances can be practically solved, measured both quantitatively and qualitatively? Which decisions can a programmer make in order to speed up a computational process guided by an implemented algorithm? Thinking about these questions, stated solely in the context of the binary knapsack problem, motivated the research reported in this paper. In order to see the actual experimental results from an appropriate standpoint, a reader should be first introduced to the landscape of the binary knapsack problem.

The classical binary or 0-1 knapsack is an NP-hard [12] maximization problem defined by the following setting: given a list of n items, characterized by their respective values v_i ($i = 1, 2,\ldots, n$) and weights w_i ($i = 1, 2,\ldots, n$), find a subset of these items that can be packed in a knapsack, with maximum carrying weight of W, whose total value is a maximum. Without any loss of generality, we can assume that all of the weights and values of items are positive integers. The Integer Linear Programming (ILP) model of the 0-1 knapsack is one of the simplest and it is created by introducing the Boolean decision variables x_i, with $x_i = 1$ if the item i is selected to be included, and $x_i = 0$ otherwise. Formally, the ILP model of the binary knapsack can be formulated as follows:

$$\text{maximize} \sum_{i=1}^{n} v_i x_i \tag{1}$$

$$\text{subject to} \sum_{i=1}^{n} v_i x_i \le W \tag{2}$$

$$x_i \in \{0,1\}, i \in \{1,\ldots, n\}$$

The sum in Eq. (1) is called *the objective function*, and the inequality in Eq. (2) is *the constraint*.

The 0-1 knapsack problem is actually a member of a large family of knapsack problems, presented in detail in [12], and, more recently, in [6], [8]. Knapsack problems arise naturally in many real-world applications, e.g., scheduling problems, capital budgeting, resource allocation with financial constraints, cargo loading, and cutting stock. In applied mathematics, especially operations research and cryptography, knapsack problems were studied and used in numerous contexts (e.g., Merkle-Hellman knapsack cryptosystem).

The binary knapsack problem can be solved by means of exact and approximation algorithms. These algorithms have different performances, the study of which, in the light of various implementation issues, is reported in this paper. The chosen representative of the family of the exact algorithms for the 0-1 knapsack is the state of the art combo algorithm, described in [10]. An approximation algorithm chosen for implementation and testing is the fully polynomial time approximation scheme, described in [16]. Its selection was motivated by two different reasons. First, it offers a representative behavior of the approximation algorithms with errors arbitrarily close to zero at an acceptable implementation cost for testing purposes. Second, it is a computationally and memory intensive algorithm, and thus it is suitable as the benchmark for testing different programming implementations.

The data acquired in the presented research can be useful at least for the following purposes:

i) Creating recommendations for selection and implementation of the algorithms for solving 0-1 knapsack instances that arise in practical applications.
ii) Developing guidelines for possible future research through understanding the limitations and bottlenecks of the current algorithms.

[1]Dušan B. Gajić is with the Computational Intelligence and Information Technology Laboratory, Department of Computer Science, Faculty of Electronic Engineering, A. Medvedeva 14, 18000 Niš, Serbia, E-mail: dusan.gajic@elfak.rs

II. Exact Algorithms for the 0-1 Knapsack Problem

First exact algorithms for the 0-1 knapsack problem were developed in the 1950s, using the dynamic programming paradigm developed by Bellman [4].

Kolesar in 1967 was the first one to construct a branch-and-bound algorithm for the 0-1 knapsack problem (for the best algorithm created through this approach see [12]). Branch-and-bound algorithms can effectively deal only with small and easy problem instances [11].

Balas and Zemel [3] proposed to randomly set the optimal values of some of the decision variables, and then focus the enumeration on the most promising ones. The subset of the items built in this way is called the core. The core problem can then be solved by either heuristics or branch-and-bound algorithms. This approach proved to be helpful in solving larger problem instances.

Martello, Pisinger and Toth [10] proposed a hybrid technique of combining dynamic programming with tight bounds. This algorithm, called the combo algorithm, has the theoretical worst-case time complexity of $O(nW)$, but it has proven in practice to be the state of the art in solving 0-1 knapsack problem instances to optimality [8], [10], [11]. As a consequence of this fact, an implementation of the combo algorithm in C by Pisinger, available from [7], was chosen as the one of the subjects of the tests that were conducted in the presented research.

III. Approximation Algorithms for the 0-1 Knapsack Problem

An approach based on settling for an approximate solution to the 0-1 knapsack problem offers reasonable algorithm running times, even for the problem instances with exponentially-growing values and weights.

A reasonable way to measure the distance between an approximate and an optimal solution is the relative performance ratio ρ which bounds the value of maximum ratio between an approximate and an optimal solution. For maximization problems, relative difference or error ε is defined as $1-\rho$. Thus we refer to $(1-\varepsilon)$-approximation algorithms when the performance ratio of an approximate solution to an optimal one is larger than or equal to $1-\varepsilon$. An algorithm is an ε-approximation scheme if it is a $(1-\varepsilon)$-approximation algorithm for every input value of $\varepsilon > 0$. Further, an algorithm is said to be a polynomial time approximation scheme, abbreviated PTAS, if, for each fixed $\varepsilon > 0$, its running time is polynomial in the input size n. If the previous definition is expanded in a way to require that the running time of an algorithm is polynomial both in n and in $1/\varepsilon$, we get a fully polynomial time approximation scheme, abbreviated FPTAS.

The binary knapsack problem is a member of a "fortunate" class of NP-complete problems in terms of approximation, i.e., polynomial time approximations with errors arbitrarily close to zero exist for it. The first FPTAS for the knapsack problem was presented in 1975 [5]. A FPTAS by Keller and Pferschy [8] has recently improved on all of the previous schemes, but only under the assumption that $n \geq 1/\varepsilon$ and at a high implementation cost, due to the fact that the algorithm is fairly complicated (as stated by the authors themselves in [8]). Based on this fact, as well as the other reasons already stated in Section 1, a FPTAS described by Vazirani [16] was selected for implementation.

Basically, all of the fully polynomial time approximation schemes for the 0-1 knapsack problem follow the same approach of scaling values of items in order to reduce the number of different total sums of values, and then apply dynamic programming by values to a scaled instance. For the implemented algorithm, this scaling of the values of items is done in the following way:

$$v_i^* = \left\lfloor \frac{v_i}{v_{\max}} \cdot \frac{n}{\varepsilon} \right\rfloor \tag{3}$$

where $\varepsilon > 0$ and $v_{\max} = \max v_i$, $i \in \{1,...,n\}$

To sum up, the design of almost all known approximation schemes is based on the idea of trading accuracy for time. The given problem instance is transformed into a less "precise" one, depending on the concrete value for ε, which is then solved exactly by means of dynamic programming. The question whether PTAS or FPTAS is the best scheme we can hope for when dealing with an NP-hard problem is difficult and it has no clear and straightforward answer [16].

IV. Programming Implementations and Computational Experiments

In order to conduct the experiments, a test generator for different types of input instances had to be first built. The C++ test environment using this generator was constructed in a generic way, so that it can be reused. It allows observations of behavior of a wide range of algorithm implementations for different problem sizes, instance types, and data ranges. Problem instance sizes used in testing span from 50 to 10 000 items, and data range R takes value from the following set: $\{10^3, 10^5, 10^7\}$. Generated instance types are in accordance with test data sets presented in [8], and they are as follows:

a) Uncorrelated: the weights w_i and the values v_i are distributed uniformly random in $[1, R]$.

b) Weakly correlated: the weights w_i are distributed uniformly random in $[1, R]$, and the values v_i are set in $[w_i - R/10, w_i + R/10]$ and $v_i \geq 1$.

c) Strongly correlated: the weights w_i are distributed uniformly random in $[1, R]$, and the values v_i are set to $v_i = w_i + R/10$.

d) Subset-sum: the weights w_i are distributed uniformly random in $[1, R]$, and the values v_i are set to $v_i = w_i$.

For each of the instance types, data ranges, and problem sizes, 100 input instances were generated. The knapsack capacity was set to a value of: $W = 0.5 \sum_{i=1}^{n} w_i$. All of the experiments were done using a PC with *Intel Core i7 920* quad-core processor and 4 GBs of RAM, running a 64-bit *Kubuntu 9.10* operating system with *Linux kernel 2.6.31.19*. The C/C++ source code was compiled using the *GNU Compiler Collection* (*gcc*), version 4.4.1. The CPU times,

appearing in all of the tables in this section, correspond only to the execution of the algorithms themselves and do not include time used for test generation and I/O operations.

The combo algorithm was tested first. The test results for the computationally easy instances are presented in Table I, and for the computationally hard ones in Table II. An in-depth explanation of why the uncorrelated and the weakly correlated instances are easier for computation than the strongly correlated and the subset-sum instances can be found in [8]. It is important to state here that the problem instances that are most often met in practice correspond to the weakly correlated type. The test results clearly show that the combo algorithm has an outstanding performance for almost all of the instances. But, it also has its limitations, as we can see from Table II. It cannot deal with instances that are both computationally hard and large in range ($R \geq 10^7$). This was expected as a consequence of the NP-hardness of the 0-1 knapsack problem.

TABLE I
COMBO ALGORITHM, EASY INSTANCES, CPU TIMES [MILLISECONDS], n - INPUT SIZE, R - DATA RANGE

n \ R	*uncorrelated*			*weakly correlated*		
	10^3	10^5	10^7	10^3	10^5	10^7
50	0.01	0.01	0.01	0.03	0.03	0.04
100	0.02	0.02	0.02	0.04	0.06	0.06
500	0.05	0.06	0.06	0.10	0.22	0.25
1000	0.07	0.09	0.08	0.13	0.46	0.47
5000	0.27	0.35	0.37	0.30	2.26	2.56
10000	0.38	0.72	0.76	0.73	4.42	6.02

TABLE II
COMBO ALGORITHM, HARD INSTANCES, CPU TIMES [MILLISECONDS], n - INPUT SIZE, R - DATA RANGE, – STANDS FOR ALGORITHM TERMINATED WITH NO RESULT

n \ R	*strongly correlated*			*subset-sum*		
	10^3	10^5	10^7	10^3	10^5	10^7
50	0.15	3.59	-	0.09	7.83	-
100	0.26	5.15	-	0.09	2.61	-
500	0.42	2.80	-	0.08	1.58	-
1000	0.50	2.85	-	0.06	1.65	-
5000	1.13	3.21	-	0.11	1.81	-
10000	1.82	3.71	-	0.19	1.98	-

The second in line for testing was the FPTAS. The results of the tests done with the fully optimized C++ implementation operating on the easy and the hard instances are given in Tables III and IV, respectively. The tests clearly show that satisfying for an approximate solution makes the job much easier, when dealing with large data. The FPTAS has stable performance for all of the instance types, while the running time grows fast with the size of the input. The identified bottleneck of the algorithm is the memory consumption in the dynamic programming phase, which limits application of this concrete implementation to instances with $n < 1000$. More advanced and complicated schemes, e.g., the one described in [8], share the same general type of limitation, but improve on the upper limit value for n.

TABLE III
FPTAS, EASY INSTANCES, CPU TIMES [MILLISECONDS], $\varepsilon = 0.1$, n - INPUT SIZE, R - DATA RANGE, – STANDS FOR ALGORITHM TERMINATED (OUT OF MEMORY)

n \ R	*uncorrelated*			*weakly correlated*		
	10^3	10^5	10^7	10^3	10^5	10^7
50	2.12	2.33	2.44	2.27	2.20	2.30
100	16.99	18.84	18.42	17.97	17.86	17.75
500	2.2×10^3	2.3×10^3	2.1×10^3	2.2×10^3	2.1×10^3	2.0×10^3
1000	-	-	-	-	-	-

TABLE IV
FPTAS, HARD INSTANCES, CPU TIMES [MILLISECONDS], $\varepsilon = 0.1$, n - INPUT SIZE, R - DATA RANGE, – STANDS FOR ALGORITHM TERMINATED (OUT OF MEMORY)

n \ R	*strongly correlated*			*subset-sum*		
	10^3	10^5	10^7	10^3	10^5	10^7
50	2.76	2.77	2.79	2.37	2.43	2.49
100	20.32	20.62	20.66	18.87	18.50	18.74
500	2.5×10^3	2.5×10^3	2.3×10^3	2.2×10^3	2.2×10^3	2.1×10^3
1000	-	-	-	-	-	-

The second batch of experiments was conducted in order to acquire information on how much decisions concerning implementation issues affect actual performance of algorithms for the 0-1 knapsack problem. More precisely, the following effects on performance were measured and analyzed:

- application of templates from the C++ STL
- application of code tuning techniques
- application of compiler optimizations

The Standard Template Library, abbreviated STL, is the first library of generic algorithms and data structures for C++, created by Stepanov and Lee [15]. The STL was developed with four main ideas in mind [15]: generic programming, abstractness without loss of efficiency, the Von Neumann computation model, and value semantics. In the light of the presented research, it sounded interesting to see how well does the idea of enabling high level abstraction without loss of efficiency hold in the context of an algorithm implementation for the binary knapsack.

Code tuning is a single name for a set of techniques with a common goal of improving performance of an implemented algorithm. Code tuning techniques are mostly motivated by the Pareto principle, and they are described in detail in [13], [14]. The techniques applied in the experiments reported here include: minimization of array references and work inside loops, loop fusion, strength reduction, elimination of system routines and common sub-expressions, etc.

Modern compilers offer various levels of compile-time optimizations [2]. Some authors suggest that these optimizations are superior to any of the code tuning techniques [13] and that a programmer should just write clear code and leave optimizations to compilers. Therefore, it was of matter of interest to this research, to test these claims in practice, in the context of the 0-1 knapsack problem.

Table V presents the results of the tests conducted with C++ implementations of the FPTAS. The FPTAS was chosen as the benchmark because it features intense numerical computations and has a high demand on memory. The tests were done using weakly correlated 0-1 knapsack instances, where $R = 10^5$, $\varepsilon = 0.1$, which best correspond to problems most frequently met in practice. Compiler optimizations used were invoked with the flag *–O3* for *gcc* compiler, and they include options like the function in-lining, loop unswitching, tree vectorization, etc. Compiler optimization techniques are described in detail in [2]. The columns in Table V represent the following algorithm implementations:

A. code using the STL, without code tuning and compiler optimizations
B. code not using the STL, without code tuning and compiler optimizations
C. code not using the STL and compiler optimizations, with code tuning
D. code using the STL and compiler optimizations, without code tuning
E. code not using the STL, with code tuning and compiler optimizations

TABLE V
CODE IMPROVEMENT, CPU TIMES [MILLISECONDS], *n* - INPUT SIZE

n	*A*	*B*	*C*	*D*	*E*	*time saved A-E*
50	20.56	10.86	8.31	2.43	2.20	89%
100	146.86	80.02	67.23	18.62	17.86	88%
500	17924.01	9616.97	8110.82	2293.84	2088.03	88%

The most important conclusions that can be drawn from the presented data are as follows:

i) Application of the STL classes without the compiler optimizations leads to the significant loss of performance (compare columns *A* and *B* in Table V). Modern C++ compilers are tuned to minimize any abstraction penalty arising from heavy use of the STL. Therefore, the STL classes do offer higher abstraction without loss of efficiency, but only when coupled with compile-time optimizations.
ii) Code tuning techniques offer some speedup, but their effects are almost completely canceled when compiler optimizations are enabled. This happens because most of the tuning techniques are actually on the list of optimizations that modern compilers can offer. Therefore, code tuning proved to be a time-consuming method that provides only a small speedup in the case of the compile-time optimized C++ FPTAS implementation.
iii) Application of the compiler optimizations alone leads to the time savings of 87-88%, so it can be stated that they are the single most important improvement factor in the case of the C++ FPTAS.

We can conclude from the experimental data that an increase in performance of almost an order of magnitude can be achieved through optimization solely at the implementation level, at least for the tested algorithm.

V. CONCLUSION

The first set of the computational experiments with the programming implementations of the algorithms for the 0-1 knapsack problem identified some of the principal limitations and bottlenecks of the two analyzed algorithms, one exact and one approximate. This information can be helpful in estimating which problem instances can be efficiently handled in today's practical applications. Further research in algorithms that combine tight bounds with dynamic programming, like the combo algorithm, looks promising.

The second part of the research dealt with the implementation issues and their effects on the performance of the 0-1 knapsack FPTAS. The recorded results are in line with the statement from Knuth [9]: "Premature optimization is the root of all evil". Compile-time optimizations proved to be by far the most important factor in the improvement of the C++ FPTAS implementation. Code tuning techniques, applied in the code construction phase, offer some speedup, but often at the price of lowering code readability. Further, modern compilers are better at optimizing straight than intricate code, and intricacies are frequently created through tuning. So, another sound advice might also come in the following words [1]: "Programs must be written for people to read, and only incidentally for machines to execute".

REFERENCES

[1] H. Abelson, G. Sussman, *Structure and Interpretation of Computer Programs*, MIT Press, 2nd edition, 1996.
[2] A. Aho, M. Lam, R. Sethi, J. Ullman, *Compilers: Principles, Techniques, and Tools*, Addison Wesley, 2nd edition, 2006.
[3] E. Balas, E. Zemel, "An algorithm for large zero-one knapsack problems", Operations Research, no. 28, pp. 1130-1154, 1980.
[4] R. E. Bellman, *Dynamic Programming*, Princeton University Press, Princeton, NJ, 1957.
[5] O. Ibarra, C. Kim, "Fast approximation algorithms for the knapsack and sum of subset problems", Journal of the ACM, no. 22, pp. 463-468, 1975.
[6] D. Gajić, *Algorithms for NP-complete Problems*, Master's thesis, Faculty of Electronic Engineering, Niš, 2009.
[7] *http://www.diku.dk/~pisinger/codes.html*
[8] H. Kellerer, U. Pferschy, D. Pisinger, *Knapsack Problems*, 1st edition, Springer, 2004.
[9] D. Knuth, „Structured Programming with *go to* Statements", Computing Surveys, vol. 6, no. 4, 1974.
[10] S. Martello, D. Pisinger, P. Toth, "Dynamic Programming and Strong Bounds for the 0-1 Knapsack Problem", Management Science, vol. 45, no. 3, pp. 414-424, 1999.
[11] S. Martello, D. Pisinger, P. Toth, "New trends in exact algorithms for the 0-1 knapsack problem", Technical Report 97/10, DIKU, University of Copenhagen, 1997.
[12] S. Martello, P. Toth, *Knapsack Problems: Algorithms and Computer Implementations*, John Wiley & Sons, 1990.
[13] S. McConnell, *Code Complete: A Practical Handbook of Software Construction*, Microsoft Press, 2nd edition, 2004.
[14] M. Scott, *Programming Language Pragmatics*, Morgan Kaufmann, 3rd edition, 2009.
[15] A. Stepanov, M. Lee, "The Standard Template Library", HP Laboratories Technical Report 95-11(R.1), 1995.
[16] V. Vazirani, *Approximation Algorithms*, Springer, 1st edition, 2001.

Volume Modeling and Seeding Parameters Calculation for Hail Suppression Purposes

Marko Kovačević[1], Vladan Mihajlović[1], Igor Antolović[1], Bratislav Predić[1], Dejan Rančić[1], Zoran Babić[2]

Abstract – **Different techniques are used for volume modeling, based on advantages they provide in the particular domain. The most popular models are surface and volume based ones. Choosing the most appropriate model for cloud representation is essential for hail suppression purposes. The voxel based model is chosen for modeling of clouds. The choice is made to satisfy the request for precise and efficient calculation of rocket launching parameters for seeding the hail-threatening cloud. The developed algorithm is based on computation of number of voxels in the seeding area, affected by the part of rocket trajectory during reagent seeding. The presented solution is designed to replace the existing algorithm for rocket launching parameters calculation implemented in Hail Suppression Information System (HASIS 3D). The new solution uses 3D mode for seeding area extraction and 3D criteria for rocket launching parameters calculation, thus making it much more precise than the previous one.**

Keywords – **Volume modeling, Voxel, Hail, HASIS.**

I. INTRODUCTION

The importance of information systems in meteorology domain cannot be overemphasized [1]. In this paper, we focus on specific type of information systems involved in weather modification domain – information systems for hail suppression. Main purpose of these systems is to predict the occurrence of hail hazard in the near future and manage hail suppression activities. HASIS (Hail Suppression Information System) is an information system for hail suppression, developed at the Faculty of Electronic Engineering in Niš, to satisfy requirements of Republic Hidrometeorological Service of Serbia (RHMSS) [2]. HASIS is a complete system that uses methodology based on cloud seeding using rockets and stationary radars for monitoring weather conditions. HASIS obtains radar data of cloud reflectivity in polar coordinate system, with center at radar location. To efficiently use this data, the conversion to Cartesian coordinate system is performed. The result is the 3D matrix of cloud reflectivity data that can be visualized and analyzed in 2D or 3D view [3]. HASIS provides measurement of hail-threatening cloud parameters, calculation of cloud seeding area, selection of rocket launching stations (RLSs) and calculation of rocket launching ballistic parameters [4].

Previous version of HASIS system was only able to analyze atmosphere situation using 2D radar data products. Consequently, 2D mode of cloud seeding area extraction produces seeding area with equal horizontal 2D cross sections on all altitudes. This approximation of the seeding area can be considerably different from the real 3D area that should be seeded. This deficiency is resolved in the new version of HASIS system, HASIS 3D, which is able to process data in 3D space. The new quality provides possibility for improving the system for seeding parameters calculation. The system for seeding parameter calculation, implemented in the previous version of HASIS system, is conformed to the methodology of cloud seeding area extraction in 2D mode, and cannot exploit the advantages of new methodology. In this paper, we present a new approach for seeding parameters calculation that exploits all the advantages of new methodology of cloud seeding area extraction in 3D mode. The voxel based model is chosen for modeling of hail-threatening clouds, mainly because it is the most suitable for representation of input radar data set and efficient examination of spatial relationship between rocket trajectories and seeding area. The new algorithm for seeding parameters calculation is designed to efficiently use voxel based model and rocket trajectories in 3D space. The new system achieves higher level of accuracy of calculated parameters then the old one, while retaining the satisfactory processing performance.

The paper is organized in six sections. The section 2 describes methodology for automatic selection of RLSs and calculation of ballistic parameters for hail suppression rocket launching that is supported by HASIS system. The section 3 gives a brief theoretical insight in volume modeling problem, essential for proper selection of cloud modeling technique. Section 4 presents the selected cloud modeling technique and the reasons for given selection. Section 5 describes the new system for automatic selection of RLSs and calculation of ballistic parameters for rocket launching. The ending section summarizes the advantages of the proposed system and specifies the interests for further research.

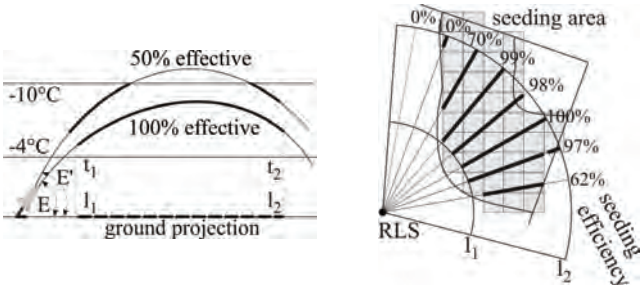
II. EXISTING METHODOLOGY

As previously mentioned, cloud analysis can be performed in 2D or 3D view of cloud reflectivity data [3]. Based on meteorological knowledge and experience, radar operator can determine whether the observed cloud is hail-threatening. If the cloud is classified as hail-threatening, operator proceeds to

[1]Marko Kovačević, Vladan Mihajlović, Igor Antolović, Bratislav Predić, Dejan Rančić is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: markko.marce@gmail.com, wlada@elfak.ni.ac.rs, antol@elfak.ni.ac.rs, bpredic@elfak.ni.ac.rs, ranca@elfak.ni.ac.rs

[6]Zoran Babić is with the Republic Hydrometeorological Service of Serbia, Kneza Višeslava 66, 11000 Beograd, Serbia, E-mail: zbabic2003@yahoo.com

the 2D view, where he analyzes horizontal cloud cross sections on desired altitudes and specifies the type of the cloud (one of six types). Then, HASIS system automatically calculates horizontal cross section of the area to be seeded. The third dimension of the seeding area is calculated knowing the fact that the aim of cloud seeding is to inject the chemical reagent as much as possible in the volume limited by the -4°C and -10°C isotherm (Fig. 1a). It should be noticed that the described methodology [4] calculates the seeding area that has equal horizontal cross sections on all altitudes. This clearly doesn't correspond to the actual cloud area that should be seeded.



(a) **(b)**

Fig. 1. a – elevation calculation, b – azimuth calculation

After seeding area calculation, automatic selection of RLSs and calculation of rocket launching parameters is performed. This must be completed very quickly, because the whole process, from detecting the hail-threatening cloud to rocket launching, is limited to 2 minutes. This is the reason the calculation is performed in two steps [4]. Rocket launching parameters are: azimuth, elevation and type of rocket. Type of rocket uniquely determines seeding time delay (the time between launching and start of seeding) and seeding duration.

First step – elevation calculation: For every type of rocket from every RLS calculate the elevation with maximum seeding efficiency, ignoring the cloud existence (Fig. 1a). The seeding efficiency criteria for given elevation represents the amount of seeding material injected in the volume between the -4°C and -10°C isotherm. For the chosen elevation the horizontal projection of the seeded track is calculated (l_1 and l_2) and memorized for further usage by the azimuth calculation step. Elevation calculation is usually done only once a day with the arrival of a new synoptic report with the new values for -4°C and -10°C level altitudes.

Second step – azimuth calculation: For every type of rocket from every RLS (with previously calculated elevation) calculate the azimuth with maximum seeding efficiency (Fig. 1b). This is a 2D problem. Previously determined horizontal cross section of the seeding area is represented by a collection of radar cells shown as gray rectangles in Fig. 1b. The rocket seeding trajectory is represented by a straight line from l_1 to l_2 away from the RLS, for particular azimuth.

Given approach is suitable for the existing methodology of seeding area extraction, performed in 2D mode. It cannot exploit the advantages of new methodology of seeding area extraction, performed in 3D mode. In this paper, we present a new approach for automatic selection of RLSs and calculation of rocket launching parameters that resolves the emphasized drawbacks and increases accuracy of the result, while retaining nearly the same processing efficiency.

III. VOLUME MODELING

In computer graphics, the term volume modeling [5] is used to denote a set of techniques for creating a logical entity (model) of 3D object, using a computer. Model is a 3D object description, made using strictly defined language or data structure. Model represents some of the 3D object characteristics, important for the specific appliance. Appliances of models are diverse: visualization, various types of analysis, simulations etc. Selection of the volume modeling technique depends on the domain of appliance and posed requirements (memory and processing requirements for visualization, for various calculations etc.). There are three main types of models: *wireframe*, *surface based* and *volume based* model. In this paper we will focus on surface and volume models, suitable for modeling complex 3D objects.

Surface based model [6] represents a 3D object as a set of vertices, boundary edges and boundary surfaces. The volume of a 3D object is modeled by defining a boundary that separates 3D object from the rest of the world. Boundary of complex 3D objects can be approximated by *polygonal mesh* or by *surface patches*. Polygonal mesh is the most commonly used form, as any object can be approximated by polygonal mesh, to arbitrary accuracy (Fig. 2). However, as the desired accuracy increases, i.e. level of details (LOD), so does the number of polygons in the mesh, and also processing time and memory requirements. On the other hand, small number of polygons in the mesh results in a poor approximation of the boundary, and limited use in domains where high level of precision is requested. Surface patch is a part of some surface, surrounded by 3 or 4 curves (Bezier curves, B-splines etc.). Compared to polygonal approximation, approximation by surface patches is more economical, as far less surface patches then polygons is needed for achieving the desired accuracy.

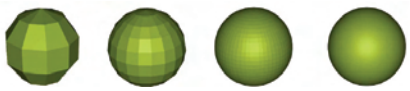


Fig. 2. Polygonal approximation of a sphere, with different LODs

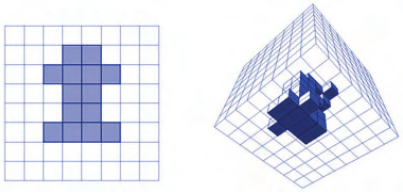


Fig. 3. Comparison between pixel formation in 2D space (left) and voxel formation in 3D space (right)

Volume based model [6], [7] contains information about vertices of a 3D object, its boundary edges, boundary surfaces and enclosed volume. Due to the large amount of information about a 3D object that this type of model contains, it is widely used for visualization, various calculations, simulations etc. State of the art volume based model is *voxel based model*.

Volume that contains a 3D object being modeled, is broken down to small volumetric elements of equal size – *voxels* (volumetric pixels). Voxel is the smallest volumetric element on the regular rectangular grid in 3D space. Voxels in 3D space are analogue to pixels in 2D space (Fig. 3). Every voxel is associated with a numerical value that represents some measurable feature of 3D space (color, density, type of material, hardness etc.). These values are stored in 3D matrix, called *3D volume buffer*, that represents a portion of space being modeled.

Differences between surface and volume graphics capabilities are the result of fundamentally different mediums for representation (i.e. modeling) and manipulation of 3D scenes. For this purpose surface graphics uses a list of geometry objects, while volume graphics employs a 3D volume buffer. Table I shows the comparison between surface and volume graphics, given for most important capabilities regarding the problem discussed in this paper [7].

Volume graphics is beginning to supersede surface graphics for modeling, handling and visualizing volumes as well as for modeling and rendering synthetic scenes composed of surfaces. This is particularly the case when it comes to volume modeling and visualization of some specific complex 3D objects (clouds, smoke, fire etc.).

TABLE I
COMPARISON BETWEEN SURFACE AND VOLUME GRAPHICS

Capability	Surface graphics	Volume graphics
1. Memory and processing requirement	Variable - depends on scene and object complexity	Large but constant
2. Rendering of interior and amorphous phenomena	No, surface only	Yes, rendering of inner structures as well as surfaces
3. Adequacy for sampled data and intermixing with geometric data	Partially and indirectly (fitting followed by surface rendering)	Directly (supports representation and direct rendering)
4. Measurements (e.g., distance, area, volume)	Analytical, but often complex	Discrete approximation, but simple

IV. SELECTION OF CLOUD VOLUME MODELING TECHNIQUE

For volume modeling of hail-threatening cloud (that contains calculated seeding area) the best choice is **voxel based modeling technique**. There are several important reasons that support this choice.

Voxel based model is extremely suitable for modeling of sampled data. In HASIS system, data acquisition is done by data sampling, i.e. radar scanning of environment. To efficiently use the obtained data, the conversion to Cartesian coordinate system is performed. The result is 3D matrix of cloud reflectivity data. This matrix is transformed, using new methodology of seeding area extraction. The result is a 3D matrix equivalent to the 3D volume buffer for voxel based model of scanned portion of space (voxels that reside in the seeding area are marked with 1, and voxels outside the seeding area with 0). Use of voxel based model for calculation of rocket launching parameters eliminates the need for additional transformation of data, as this 3D matrix is an input for the new algorithm for automatic selection of RLSs and calculation of rocket launching parameters. This would not be the case if surface based model was used.

Because of an amorphous nature of the cloud structure, geometric definition of boundaries of the extracted seeding area can be very complex. As a consequence, all calculations that include seeding area, done on surface based model, are potentially time consuming. On the other hand, voxel based model offers the possibilities for very simple calculation of different measurements (distance, volume etc.). This is an essential feature, because, as mentioned before, the calculation of launching parameters must be done very quickly. Precision of calculations is not a bottle neck, due to the fine resolution of 3D volume buffer (in the typical case, the 3D volume not larger than 20km x 20km x 2km is represented using 3D volume buffer of size 400 x 400 x 40).

V. CALCULATION OF LAUNCHING PARAMETERS

Input parameters of the new system for automatic selection of RLSs and calculation of ballistic rocket launching parameters are: seeding area, location of RLSs and hail suppression rocket parameters.

Input seeding area parameters are: 3D volume buffer, referent point of scanned volume and voxel size. 3D volume buffer determines the location of the seeding area inside the scanned volume (voxels that reside in the seeding area are marked with 1, and voxels outside the seeding area with 0). It is important to point out that all clippings of the seeding area, due to the position of -4°C and -10°C isotherm and forbidden areas marked by the flight control, are done before creating 3D volume buffer. Position of the scanned volume in 3D space is fully determined via coordinates of referent point and voxel size. Based on this information it is possible to determine whether specific point in 3D space belongs to the modeled volume or not. In the case it does, it is possible to calculate the indexes of the voxel that contains particular point.

RLSs (hail suppression stations) are defined with their world coordinates, stored in the database of HASIS system. New system for launching parameters calculation takes into account that some launching directions on particular RLSs are forbidden, that different rocket launchers have different rotation steps and that every RLS does not necessarily have all types of hail suppression rockets.

Hail suppression rockets are classified by their type that determines seeding time delay (the time between the rocket launching and the beginning of seeding) and seeding duration. Rocket launching parameters that the new system should determine are: rocket type, elevation and azimuth. These parameters uniquely define rocket trajectory. The database of HASIS system contains definition of rocket trajectory for every rocket type and every significant elevation. Rocket trajectories are defined using parameter functions x=x(t) and

z=z(t) that represent the correlation between distance (x-coordinate and z-coordinate, respectively) and time. Given parameter functions stand for azimuth equal 0°, in which case the y-coordinate is constant. When azimuth differs from 0°, x-coordinate and y-coordinate are computed using simple trigonometry. So, rocket trajectory is uniquely defined by parameter functions x=x(t) and z=z(t) and azimuth. Using this information, rocket trajectory coordinates can be calculated in arbitrary moment of time

Based on the previous discussion, we conclude that index of voxel affected by the rocket trajectory can be calculated for arbitrary moment of time. In the new system for launching parameters calculation, the efficiency of a rocket trajectory is measured by the *number of voxels* of the seeding area affected by the part of rocket trajectory during reagent seeding. By inspecting all possible rocket trajectories from a specific RLS, launching parameters for that RLS are determined. Pseudo code of this algorithm is shown in Fig. 4.

```
Read information about RLS from database.
foreach permitted rocket trajectory RT of RLS
begin
    Read information about rocket trajectory from database.
    foreach permitted azimuth A of RLS
    begin
        Move along the rocket trajectory, from the beginning of
        seeding to the first voxel of modeled volume.
        Continue moving along the rocket trajectory while the
        rocket trajectory is in the modeled volume and seeding is
        in progress; count voxels of the seeding area affected by
        the rocket trajectory.
        If calculated voxel count is the biggest calculated so far,
        memorize parameters of the rocket trajectory (including
        azimuth).
    end foreach.
end foreach.
Write memorized parameters of the rocket trajectory in the
database.
```

Fig. 4. Launching parameters calculation for the specific RLS

As the pseudo code shows, all discussed restrictions are considered. In order to determine which voxels are affected by the rocket trajectory, time sampling of the rocket trajectory is needed. Choosing appropriate time step is very important. Short time steps can lead to very bad processing performance. On the other hand, long time steps can cause many voxels to be skipped. Implemented algorithm uses *variable time step*. Longer time step is used while the rocket trajectory has not yet reached the modeled volume. Near and in the modeled volume time step is chosen so the distance between two consecutive points on the rocket trajectory is less than tenth of the length of the shortest voxel edge. This way, probability of skipping a voxel is very small.

Described algorithm is repeated for all RLSs. RLS is declared inactive if no rocket trajectory from this station is able to seed the required minimal number of voxels. Algorithm uses spatial data coherence to achieve higher efficiency [8]. Although using 3D seeding area, processing performance of the presented method is at the same level as processing performance of the existing method. In typical cases, the time needed for launching parameter calculation is less then *5 seconds*.

VI. Conclusion

Unlike the old system, the new system for automatic selection of RLSs and calculation of ballistic parameters for hail suppression rocket launching is able to exploit all the advantages of new methodology for seeding area extraction in 3D mode. In the new system the real volume of seeding area is used, which makes calculated parameters for hail suppression rocket launching much closer to the optimal ones. Due to the chosen volume modeling technique (voxel based modeling technique) and an adequate algorithm for launching parameters calculation, efficiency of the process is retained at the same level, despite much higher calculation complexity. Integration of the new method in the hail suppression system HASIS 3D is in progress.

New approach provides possibilities for future research and improvement. Instead by the number of voxels of the seeding area affected by the rocket trajectory, the efficiency of the rocket path can be measured by the amount of time rocket spends in the seeding area, during reagent seeding. In proposed method, every rocket launching station is inspected separately. The next step will be to consider spatial relationship between trajectories of rockets launched from numerous stations. In this case, efficiency of the configuration of rocket trajectories would be an issue.

References

[1] I. Mathieson, S. Dance, L. Padgham, M. Gorman, M. Winikoff, “An open meteorological alerting system: issues and solutions”, Proceedings of the 27th conference on Australasian computer science - CRPIT '04, 2004, vol. 26, pp. 351 - 358.

[2] D. Rančić, M. Smiljanić, S. Đorđević-Kajan, A. Kostić, P. Eferica, P. Vuković, Z. Vučinić, “Radar Data Processing for Cloud Seeding in Hail Suppression Information System”, Proceedings of RADME 98 - Theoretical, Experimental and Operational Aspects of Radarmeteorology, Rome, Italy, 1997, pp. 137-149.

[3] V. Mihajlović, S. Djordjević-Kajan, D. Rančić, B. Predić, I. Antolović, P. Eferica, Z. Babić, “Architecture of HASIS-3D System Designed for Hail Suppression Purposes”, Proceedings of the XLII ICEST Conference, Ohrid, Macedonia, June 2007, pp.403-406.

[4] D. Rančić, M. Smiljanić, S. Đorđević-Kajan, A. Kostić, P. Vuković, P. Eferica, Z. Vučinić, “Software Support for Cloud Seeding in Hail Suppression Information System”, YUINFO 99, Kopaonik, Serbia, 1999.

[5] J. Royer, “3D Modeling and Visualization”, 9th International CODATA Conference Data Visualization - Earth and Geo Science, Berlin, Germany, 7-10 November 2004.

[6] J. D. Foley, A. Van Dam, S. K. Feiner, J. F. Hughes, *Computer Graphics Principles And Practice*, Second Edition in C, Addison-Wesley Publishing Company, 1996.

[7] A. Kaufman, D. Cohen, D. Yagel, “Volume Graphics”, IEEE Computer, Vol. 26, No. 7, July 1993, pp. 51-64

[8] M. Jones, R. Satherley, “Voxelisation: Modelling for Volume Graphics”, Volume Graphics, Springer, 2000.

Elitism Based Evolutionary Algorithm for Discrete Optimization Problems

Vassil Guliashki[1], Chavdar Korsemov[2] and Hristo Toshev[3]

***Abstract* – In this paper different elitism strategies are considered. They are combined in an evolutionary algorithm, designed to solve discrete optimization problems. The computational complexity of the proposed algorithm is investigated and evaluated. Conclusions about the use of elitism are drawn.**

***Keywords* – Evolutionary algorithms, Elitism, Discrete optimization.**

I. INTRODUCTION

In this paper is considered integer programming problem, which can be stated in the following form:

$$\textbf{Min } F(x) \quad (1)$$

$$\textbf{subject to: } g_i(x) \le 0; \quad i = 1,\ldots,m; \quad (2)$$

$$l_j \le x_j \le u_j; \quad j = 1,\ldots,n; \quad (3)$$

$$x \in \mathbf{Z}^n, \quad (4)$$

where x is an n-dimensional vector of variables x_j, $j = 1,\ldots,n$; which accept discrete values only. By l_j and u_j are denoted the bounds (lower and upper) of x_j, and $F(x)$ is the multimodal objective function. There is no necessary $F(x)$ to posses accessible to calculation derivatives in an explicit analytical form. The functions $g_i(x)$, $i = 1,\ldots,m$; are convex nonlinear functions and m is the number of nonlinear constraints **(2)**

The integer programming problems belong to the class of NP-hard optimization problems (see [2, 6]). There does not exist an exact algorithm, which can solve these problems in time, depending polynomially on the problem input data length or on the problem size. For this reason many efficient approximate evolutionary algorithms and metaheuristic methods have been created to find out the global optimum of such complex optimization problems. The objective function **(1)** with integer variables is usually multimodal function. The approximate algorithms may not be able to find out the exact global optimal solution. A possible strategy to solve such complex optimization problems is to combine the qualities of a directional type method with the good features of a population based (evolutionary) algorithm. The directional type steps may accelerate the convergence in regular regions of the search space, while the evolutionary algorithms are able to escape the trap of local optima, exploring the whole feasible domain. For precisely locating the optimal solution(s) some kind of local search procedure may be included in the optimization algorithm.

The population based algorithms (see [5]) handle a population of individuals and make them evolve according to some rules that are clearly specified for each algorithm. At each iteration periods of *self-adaptation* (intensification of the search process in some region of the search space) alternate with periods of *co-operation* (information collective gathered during the search process is used to direct further the search). The periods of self-adaptation correspond to execution of mutation, improvement or local search procedure, and the periods of co-operation are connected with the selection, crossover, trace updating or generation of trial points, i.e. with some (explicit or implicit) sharing among the individuals useful information gathered during the search. The population based algorithms are organized according the following general scheme:

```
Generate an initial population of individuals;
While no stopping condition is met do
    Co-operation,
    Self-adaptation,
EndWhile.
```

Figure 1: General scheme of a population based algorithm

Many metaheuristic algorithms correspond to this framework and can be described within it.

One major difficulty with the creation of global search algorithms is to overcome the premature convergence towards local optima. To obtain the global optimum of the problem at hand a diversification phases of the search process are necessary, so that new areas of feasible domain, that are remained still unexplored, can be investigated. This means that after each intensification phase and locating a local optimum the search procedure has to steer the individuals away from the just explored region and from the found local optimum, as well as away from all other known local optima and their corresponding regions. During the diversification phase the individuals are modified independently but with unexpected results in the sense that they are not necessarily improved. A famous example for diversification is the tabu list strategy in the tabu search algorithms, where some characteristics of solutions or movements (steps in given directions) are stored as forbidden (tabu) for certain number of iterations. In this manner are avoided the cycling and the trap of local optimality. In general the metaheuristic algorithms

[1]Vassil Guliashki is with the Institute of Information Technologies, "Acad. G. Bonchev" str. Bl. 29A, 1113 Sofia, Bulgaria, E-mail: vggul@yahoo.com

[2]Chavdar Korsemov is with the Institute of Information Technologies, "Acad. G. Bonchev" str. Bl. 29A, 1113 Sofia, Bulgaria, E-mail: chkorsemov@iinf.bas.bg

[3]Hristo Toshev is with the Institute of Information Technologies, "Acad. G. Bonchev" str. Bl. 29A, 1113 Sofia, Bulgaria, E-mail: toshev@iinf.bas.bg

alternate phases of intensification with diversification phases. To obtain guaranteed the global optimal solution these algorithms should perform a search process, exploring the whole feasible domain, i.e. they should perform systematically diversification after each intensification period.

Another major difficulty with the creation of global search algorithms is to obtain good convergence speed, because they could converge very slowly. This could be achieved by means of some kind hybridization. To create successful and efficient global search methods the researchers very often make combinations of two or more metaheuristics in hybrid methods. Such techniques have been proposed in [1, 3, 4, 7]. For example tabu search can be coupled with directional search and scatter search approaches. Another important way to accelerate the performance of an evolutionary algorithm is to use the features of the best individuals obtained during the search process and the historically good information they have accumulated, combining them to generate new offspring individuals for the next population, i.e. to use some kind of elitism.

In this paper we propose an evolutionary algorithm for solving the problem (1)-(4), which includes different elitism strategies for selection the parent individuals, generating the next population in the search process.

II. ELITISM STRATEGIES

A. Scatter search elitism strategy

Such kind strategy has been proposed for the first time by F. Glover (see [3]). Let we have a population P containing p individuals or solutions in the feasible domain for the considered optimization problem (1-4). The initial population P can be randomly generated around the Tchebicheff center x^{tch} of the feasible domain, determined by the constraints (2)-(3). Let the feasible domain be denoted by X. The Tchebicheff center $x_{tch} \in X$ is the point located at the maximal Euclidean distance from the constraint surfaces. We assume that the Tchebicheff center is obtained by means of a method for solving the relaxed convex problem with continuous variables. Then x^{tch} is rounded to the nearest integer point ix^{tch}.

The objective function values should be calculated for each individual in the population P. Using this information the most attractive elements of P can be selected and used to generate new elements that replace previous members of P. Elite elements (those with especially good objective function values) become part of a special historical collection that is used at the final stage of the search to contribute to the generation of final solution. The following sets are referenced and updated at each iteration:

T – a set of trial solutions;

S – a set of s selected current generators;

H – a set of elite historical generators, consisting of h best (with highest evaluation) solutions generated historically throughout the search. On the first iteration H will consist of the h best elements of S.

The scatter search strategy consists of the following steps:

Step 1. Set $i = 0$ and $itlim = itl$, where itl is a positive integer. Generate initial set of trial solutions T(i).

Step 2. Select the best s solutions from T(i) and create S(i).

Create set of elite historical generators H(i), including the h best elements of S(i) in H(i).

Step 3. Let x and y are two solutions, $x \in$ S and $y \in$ S. For each pair $(x, y)_j$ consider the line through x and y, given by the expression

$$l(t) = x + t(y–x), \qquad \textbf{(5)}$$

where t is scalar weight.

The solutions $l(–1/3)$, $l(1/3)$, $l(2/3)$, $l(4/3)$ are generated. Each of them is rounded of to its nearest integer feasible solution and are obtained four new solutions: $xj^{(1)}$, $xj^{(2)}$, $xj^{(3)}$, $xj^{(4)}$;

Calculate for each generated offspring solution its objective function value.

Step 4. Update the set H(i) as follows: If some generated offspring solution has better objective function value than some solution in H(i) replace the corresponding solution in H(i) by this solution.

Step 5. Set $i = i$+1. If $i > itlim$, then STOP (the procedure terminates), else use the trial solutions T(i-1) to generate new trial solutions T(i) and go to Step 2.

B. Neighbor S_i *sets – based elitism strategy*

The main idea of this strategy is to select a given percent of the best solutions included in j different neighbor sets S_i, i=1,…j; (i.e. in j different neighbor areas of the search space) and to create a new population PS sharing their accumulated information. Then each pair solutions (x, y), where $x \in PS$ and $y \in PS$, is used as a parent pair to generate four new solutions analog to Step 3. in the Scatter search strategy. The corresponding procedure is described by the following steps:

Step 1. Let s be the cardinality of each set Si. Select $\lfloor s/j \rfloor$ best solutions from each of neighbor sets S_i, i=1,…j; (here $\lfloor v \rfloor$ denotes the integer part of v) and create the new population PS. Create set of elite historical generators H* ≡ PS.

Step 2. Let x and y are two solutions, $x \in PS$ and $y \in PS$. For each parent pair $(x, y)_p$ consider the line through x and y, given by the expression **(5)**: $l(t) = x + t(y–x)$, where t is scalar weight.

The solutions $l(–1/3)$, $l(1/3)$, $l(2/3)$, $l(4/3)$ are generated. Each of them is rounded of to its nearest integer feasible solution and are obtained four new solutions: $xp^{(1)}$, $xp^{(2)}$, $xp^{(3)}$, $xp^{(4)}$;

Calculate for each generated solution its objective function value.

Step 3. Update the set H* as follows: If some generated offspring solution has better objective function value than some solution in H* replace the corresponding solution in H* by this solution. STOP (the procedure terminates).

III. THE PROPOSED ELITISM BASED EVOLUTIONARY ALGORITHM EEA

The idea of the proposed EEA algorithm is the following:

Let the feasible domain be denoted by X. The Tchebicheff center $x_{tch} \in X$ is the point located at the maximal Euclidean distance from the constraint surfaces. We assume that the Tchebicheff center is obtained by means of a method for solving convex problems with continuous variables. Then x_{tch} is rounded to the nearest integer point i_{tch}. A regular simplex with n+1 vertices is generated around the solution i_{tch}. There are (n+1) combinations of n vertices, correspondingly for each facet of the simplex. The feasible domain X is divided in (n+1) sub-regions as follows: The rays starting at the weight center of all simplex vertices cs and passing through the vertices $v(j)$, j=1,…,n; belonging to each facet determine $K^{(i)}$ cones, i = 1,…, n+1; in the feasible domain. Similar division of the feasible region is proposed in [4]. The iterative procedure generates at each iteration k a number i of trial points sets $T(i)_k$ for each facet of the simplex i=1,…,n+1; Each set T(i) includes obtained by rounding off (to the nearest integer solutions) of points along the rays r_j starting at cs and passing trough v(j), as well as along the ray r_{ci}, starting at cs and passing through the weight center cv_i of the current facet. By means of each T(i) a set Si of s selected current generators is created and the described scatter search strategy is performed on it. All Si sets at the k-th iteration are used that to perform the Si sets – based elitism strategy. Than the sets of elite historical generators H_k and H^*_k are compared and in case H_k contains some better solutions than a solution in H^*_k, the better solution replaces the worse. At the end of all iterations the procedure finishes by applying the scatter search strategy on the final H*.

The steps of EEA algorithm are presented below:

Step 1. Find the Tchebicheff center x_{tch} and round off it to the nearest feasible integer solution i_{tch}.

Step 2. Generate a regular simplex with n+1 vertices, using i_{tch} as one vertex. Generate the other vertices on the base of the elementary geometry in the following manner:

$$v^{(i)}_j = \begin{cases} i_{tch\,j} + \varphi_1 & \text{if } j \neq i \\ i_{tch\,j} + \varphi_2 & \text{if } j = i \end{cases} \qquad \textbf{(6)}$$

$$\varphi_1 = \alpha.\left[\frac{\sqrt{(n+1)} + n - 1}{n\sqrt{2}} \right] \qquad \textbf{(7)}$$

$$\varphi_2 = \alpha.\left[\frac{\sqrt{(n+1)} - 1}{n\sqrt{2}} \right] \qquad \textbf{(8)}$$

Let i_{tch} be denoted as $v^{(0)}$. Round off each $v^{(j)}$, j = 1,…,n; to its nearest integer point.

Step 3. Calculate the weight center of the simplex:

$$cs = \frac{\sum_{j=0}^{n} v^{(j)}}{n+1} \qquad \textbf{(9)}$$

Round off cs to its nearest integer point. Set i = 1.

Step 4. For i = 1,…, n+1; explore the cone $K^{(i)}$ as follows:

Create a set of trial points $T^{(i)}$, including in it the simplex vertices of the current simplex facet and their weight center cv, where

$$cv = \frac{\sum_{v^{(i)} \in K^{(i)}} v^{(i)}}{n} \qquad \textbf{(10)}$$

Round off cv to its nearest integer point.

Calculate the objective function values of the trial solutions $x \in T^{(i)}$. Create the set Si of the best s members $x \in T^{(i)}$.

Step 5. Perform the described Scatter search strategy on the set Si. Update the set of elite historical generators Ho.

Step 6. Perform the described Si –based elitism strategy. Update the set of elite historical generators Ho*.

Let: $xold^{(j)} = v^{(j)}$, j = 1,…, n; and $xold^{(0)} = cv$.

Calculate $p^{(j)} = \gamma(xold^{(j)} - cs)$, j = 1,…, n; and $p^{(0)} = \gamma(cv - cs)$, where γ is a scale multiplier, tuned according to the concrete problem.

Step 7.

ITERATION

Let the iteration has the current number ik. Calculate the points $xnew^{(j)} = xold^{(j)} + ik.p^{(j)}$, j= 0,…, n; Round off each $xnew^{(j)}$ to its nearest integer point. In case there is a violated constraint from the system **(2)-(3)** reduce the corresponding $p^{(j)}$ as follows:

- If a constraint of type $\boldsymbol{x_k + a = 0}$ is violated, where $\boldsymbol{a}$ can have positive or negative value, then the corresponding component p_k of $p^{(j)}$, is used to reduce $p^{(j)}$:

$$p^{(j)} = \left| \frac{a - p_k}{p_k} \right| p^{(j)} \qquad \textbf{(11)}$$

- If a constraint of type $\boldsymbol{g_i(x) \leq 0}$ is violated then $p^{(j)} = 0.8\ p^{(j)}$. If it is necessary repeat this reduction until the rounded off integer $xnew^{(j)}$ becomes feasible.

- If there are more than one constraints, violated by $p^{(j)}$, then chose the most reduced vector $p^{(j)}$, so that the rounded off integer $xnew^{(j)}$ becomes feasible.

Use the generated solutions $xnew^{(j)}$ as trial solutions. For each facet include the corresponding trial solutions in the set $T^{(i)}_{ik}$. Calculate also ik new trial points on each segment seg_i between cv and $v(i)$, where each segment seg_i is divided to (ik+1) parts of equal length. The new generated ik solutions on each segment seg_i are rounded off to the corresponding nearest integer feasible point and are included in $T^{(i)}_{ik}$. Evaluate the objective function values of the trial solutions $x \in T^{(i)}_{ik}$.

Perform the described Scatter search strategy on the set Si. Update the set of elite historical generators H_{ik}.

Perform the described Si –based elitism strategy. Update the set of elite historical generators H_{ik} *.

ENDofITERATION

Repeat the *ITERATION* until it is not possible to generate any new feasible points by means of $p^{(j)}$, j=0,…, n;

Step 8. Compare all obtained H_{ik} and H^*_{ik}. In case H_{ik} contains some better solutions than a solution in H^*_{ik}, the better solution replaces the worse. Create a final population H* of the solutions in all H^*_{ik}. Perform the described Scatter search strategy on the set H*. Find the best generated solution x*. STOP.

The computational complexity of EEA algorithm is considered and evaluated as follows:

Theorem: The EEA algorithm has a polynomial computational complexity $O(n^4)$.

Proof: There are n+1 different cones as described in Step 4. On each facet are generated n+1 trial solutions and additional $ik.n$ trial solutions, where ik is the iteration number. There are $s(s-1)/2$ combinations of two solutions belonging to Si, hence $s(s-1)/2$ segments will be considered. The scatter search strategy evaluates 4 solutions for each segment, i.e. $2s(s-1)$ new solutions. The same number of new solutions is generated also in the Si – based elitism strategy. The scatter search strategy is applied n+1 times at each iteration and the Si – based elitism strategy – only once (see Step 7.). Hence at each iteration are evaluated no more than $2(n+2)s(s-1)$ solutions, where s depends linear of n (see Step 4). There are performed ik iterations. Depending on the feasible domain we could expect that $ik \sim n$. Then the total number of evaluated solutions during the performance of EEA algorithm is proportional to $(n\text{-}1)(n+2)n^2$. Hence the Theorem is proved. ■

IV. Conclusions

The presented new evolutionary method EEA has the following good features and advantages:

- During the exploration of each sub-region the EEA algorithm systematically diversifies the search process, avoiding in this manner the trap of local minima.
- The EEA algorithm performs search in all defined sub-regions of the search space, so that the whole feasible domain is explored.
- The formed final population contains diverse enough individuals, so that it is expected that the final search phase would lead to the global optimal solution.
- The EEA algorithm can be efficient in comparison to other global search algorithms, because it is an elitism based polynomial algorithm. The non elitism procedure based on the same steps would evaluate 2 or three times more solutions at each iteration. Hence it will perform two or three time slower than EEA algorithm.
- The populations used in EEA algorithm have relatively small size, so that no great memory will be necessary for its implementation.
- The EEA algorithm guarantees the feasibility of the obtained solutions.
- A great part of the integer points located near or on the rays forming each cone, which have been explored during the search in the corresponding sub-region at the current iteration, can be used during the exploration of the next sub-region. This may be used for creation of efficient program realizations of EEA algorithm.

The EEA algorithm will be tested on a set of test examples and the results will be compared with that of non elitism based evolutionary algorithms.

Acknowledgement

The authors gratefully acknowledge the support of Bulgarian National Science Fund, Grant No DTK02/71 "Web-Based Interactive System, Supporting the Building Models and Solving Optimization and Decision Making Problems".

References

[1] Damavandi N. and Safavi-Naeini S., "A hybrid evolutionary programming method for circuit optimization", In Proc. *IEEE Transactions on circuits and systems – I: Regular papers*, vol. 52, No. 5, May 2005, pp. 902-910.

[2] Garey M. R. and Johnson D. S., "*Computers and Intractability: A Guide to the Theory of NP-Completeness*", W. H. Freeman, San Francisco 1979.

[3] Glover F., "Tabu Search for Nonlinear and Parametric Optimization (with Links to Genetic Algorithms)", *Discrete Applied Mathematics*, "Viewpoints on Optimization", 1991.

[4] Guliashki, V., (2007) "A Hybrid Population Based Method Solving Convex Integer Optimization Problems", In: Proceedings of Papers of the XLII International Scientific Conference on Information, Communication and Energy Systems and Technologies ICEST2007, (Editor Prof. Dr. Mitrovski, C.), June 24 – 27, 2007, Ohrid, Macedonia, vol. I, pp. 249-252.

[5] Hertz A., Kobler D., „A Framework for Description of Population Based Methods", Tutorials and Research Reviews, 16-th European Conference on Operational Research Brussels, Belgium, 1998, pp. 48-59.

[6] NemhauserG. L. and Wolsey L. A. "*Integer and Combinatorial Optimization*", Wiley, New York. 1988.

[7] Taillard E. D., Gambardella L. M., Gendreau M. and J.-Y. Potvin, "Adaptive Memory Programming: A Unified View of Metaheuristics", *Tutorials and Research Reviews*, 16-th European Conference on Operational Research Brussels, Belgium, 1998, pp. 30-38.

Evolutionary Algorithm for Multiple Objective Convex Integer Problems

Krasimira Genova[1] and Vassil Guliashki[2]

***Abstract* – An interactive evolutionary algorithm is proposed to solve multiple objective convex integer problems. It combines two different strategies helping the Decision Maker (DM) to choose good preference direction and to focus the search process around a desired part of Pareto optimal front. These strategies lead to drastic reducing the number of iterations for moving the population to the set of Pareto-optimal points. An illustrative example is presented.**

***Keywords* – Multi-objective optimization, Evolutionary algorithms, Reference direction, Pareto-optimal front.**

I. INTRODUCTION

In this paper is considered the multiple objective convex integer optimization problem, which can be stated in the following general form:

$$\text{Minimize} \quad \boldsymbol{f}(\boldsymbol{x}) = [\, f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \ldots, f_k(\boldsymbol{x})]^{\mathrm{T}} \qquad \textbf{(1)}$$

$$\text{subject to:} \quad \boldsymbol{g}_j(\boldsymbol{x}) \le 0, \qquad j = 1,2,\ldots, m; \qquad \textbf{(2)}$$

$$\boldsymbol{x}_i^{(L)} \le \boldsymbol{x}_i \le \boldsymbol{x}_i^{(U)}, \quad i = 1,2,\ldots, n; \qquad \textbf{(3)}$$

$$\boldsymbol{x} \in \boldsymbol{Z}^n, \qquad \textbf{(4)}$$

where $\boldsymbol{g}_j(\boldsymbol{x})$, $j = 1,2,\ldots, m$; are convex functions and $f_i(\boldsymbol{x})$, $i = 1,2,\ldots, k$; are nonlinear functions.

Below in the text of the paper we consider the term "solution" as a variable vector and the term "point" as a corresponding objective vector.

A solution $\boldsymbol{x} \in \boldsymbol{Z}^n$ is a vector of n decision variables: $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^T$. The value $x_i^{(L)}$ is the known lower bound and the value $x_i^{(U)}$ is correspondingly the upper bound of variable x_i. The solutions satisfying the constraints (2)-(4) constitute a feasible decision variable space $\boldsymbol{V} \subset \boldsymbol{Z}^n$. The objective functions (1) constitute a k-dimensional space, called objective space $\boldsymbol{F} \subset \boldsymbol{R}^k$. For each solution $\boldsymbol{x}$ in the decision variable space, there exists a point $\boldsymbol{f} \in \mathbf{R}^k$ in the objective space, denoted by $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{f} = (f_1, f_2, \ldots, f_k)^T$.

The problem (1-4) does not posses an analytically defined optimal solution. To solve it vectors, containing values of each objective are compared during an optimization procedure. In this case the usual concepts are those proposed by Pareto [12].

The domination between two solutions is defined as follows (see [1, 3, 10]):

Definition 1. *A solution* $\mathbf{x}^{(1)}$ *is said to dominate the other solution* $\mathbf{x}^{(2)}$, *if both the following conditions are true*:

1. *The solution* $\boldsymbol{x}^{(1)}$ *is no worse(say the operator* $\prec$ *denotes worse and the operator* $\succ$ *denotes better) than* $\boldsymbol{x}^{(2)}$ *in all objectives*, or $f_j(\boldsymbol{x}^{(1)}) \not\prec f_j(\boldsymbol{x}^{(2)})$ for $j = 1,2,\ldots, k$;.

2. *The solution* $\boldsymbol{x}^{(1)}$ *is strictly better than* $\boldsymbol{x}^{(2)}$ *in at least one objective*, or $f_j(\boldsymbol{x}^{(1)}) \succ f_j(\boldsymbol{x}^{(2)})$ for at least one $j \in \{1,2,\ldots, k\}$.

All points which are not dominated by any other point $\boldsymbol{f} \in \boldsymbol{F}$ are called the non-dominated points of class one, or simply the non-dominated points. Usually the non-dominated points together make up a front in the objective space and are often visualized to represent a non-domination front. The points lying on the non-domination front, by definition, do not get dominated by any other point in the objective space, hence they are Pareto-optimal points (together they constitute the Pareto-optimal front), and the corresponding variable vectors are called Pareto-optimal solutions.

There are two ideal goals in the multi-objective optimization:

1. Find a set of solutions which are diverse enough to represent the entire range of the Pareto-optimal front, and

2. Find a set of Pareto-optimal solutions, which satisfy in the best way the DM's preferences.

In this paper is proposed an interactive algorithm to solve the problem (1-4). The interactive algorithms are the most popular in solving multi-objective optimization problems. They consist of two alternate phases: 1. Interaction (dialogue) with the DM and 2. Generating solutions. Usually an appropriate single objective convex integer optimization problem is solved during the second phase. Such problems belong to the class of NP-hard problems (see for example [5, 11]). There does not exist an exact algorithm, which can solve these problems in time depending polynomially on the problem input data length or on the problem size. For this reason many researchers investigate approximate algorithms with polynomial computational complexity, which solve such kind optimization problems. For the past 20 years evolutionary multiple objective optimization (EMOO) methodologies have demonstrated their usefulness in finding a set of near Pareto-optimal solutions [1, 2, 4]. As a sequence many source codes – both commercial and free have been created and the EMO algorithms obtained wide application. During the World Congress of Computational Intelligence (WCCI) in Vancouver 2006, the EMOO has been evaluated as one of the three fastest growing fields of research and application among all computational intelligence topics.

In principle the evolutionary optimization (EO) algorithms use a population-based approach, in which the iterations are performed on a set of solutions (called population) and more than one solution is generated at each iteration. The main positive features making popular the EO algorithms are the following: (i) They do not require any derivative information;

[1]Krasimira Genova is with the Institute of Information Technologies, "Acad. G. Bonchev" str. Bl. 29A, 1113 Sofia, Bulgaria, E-mail: kgenova@iinf.bas.bg

[2]Vassil Guliashki is with the Institute of Information Technologies, "Acad. G. Bonchev" str. Bl. 29A, 1113 Sofia, Bulgaria, E-mail: vggul@yahoo.com

(ii) EO algorithms are relatively simple to implement; (iii) EO algorithms are flexible and robust, i.e. they perform very well on a wide spectrum of problems (see [6]); The use of a population in EO algorithms has a number of advantages (see [1]): 1) it provides an EO procedure with a parallel processing power, 2) it allows EO procedures to find multiple optimal solutions, thereby facilitating the solution of multi-modal and multi-objective optimization problems, and 3) it provides an EO algorithm with the ability to normalize decision variables (as well as objective and constraint functions) within an evolving population using the best minimum and maximum values in the population.

Some important disadvantages of EMOO algorithms are: (i) their convergence to the Pareto-optimal front could be slow and may require large number of iterations; (ii) they face difficulty in solving problems with a large number of objectives, i.e. they could obtain difficult a well representative set of Pareto-Optimal Solutions (see [4]).

The number of objectives as a convergence factor is considered in [13]. Good approach in solving problems with large number of objectives is to use the EMOO methodologies to find a preferred and smaller set of Pareto-optimal solutions, instead of the entire front [4]. In this way the DM can concentrate to explore only the regions of Pareto-optimal front, which are of interest to her/him. An accelerating technique for population based algorithms is proposed in [7]. A technique for quickly moving the population to the Pareto-optimal front is proposed in [8]. Some hybrid EMOO algorithms have been recently proposed to overcome the second mentioned disadvantage (see [4, 9]). They expand the use of classical multi-objective optimization procedures (see [10]) like reference point-, reference direction- and other type methods, proposing new approaches and hybrid techniques.

We propose two strategies for quickly moving the whole population to the Pareto-optimal front. They are included in the proposed algorithm for multiple objective convex integer problems (MOCIP-algorithm) and lead to drastic reduction of iteration number for convergence the search process to Pareto-optimal front. At the same time they deal multiple preference conditions simultaneously and have robust performance for problems with large number of objectives. A specific form of fitness function is proposed for performing a small number of genetic iterations if DM considers that this may help to find better compromise solution. At the final phase of the search process we combine our approach with the reference point approach suggested by Wierzbicki [14].

II. PROPOSED REFERENCE POINT STRATEGIES

The main idea of both proposed strategies is for given reference point provided by the DM to construct direction in the variable space, which is used to move quickly the whole population of solutions to the desired part of Pareto-optimal front – as close as possible to a given reference point.

A. Worst point reflection strategy (WPRS)

Let the reference point *fr* is known and let we have a population P of p solutions. Then perform the following procedure:

Step 1. Calculate the Euclidean distance measure in the criteria space between each point $fi \in P$, for $i = 1,\ldots,p$; and the reference point *fr*:

$$di = \sqrt{\sum_{j=1}^{k} \left(fi_j - fr_j \right)^2} \qquad \textbf{(5)}$$

Step 2. Determine the solution $\boldsymbol{x}_{\max}$, corresponding to the point $fi_{\max}$, having maximal value $di_{\max}$. Set *icount* = 1. Set *itlim* = *const*. Here *const* ≤ *n*+1;

Step 3. Calculate the weight center **c** of all other solutions $\boldsymbol{x}i \in P$, for $i = 1,\ldots,p$-1; and $i \neq$max:

$$\boldsymbol{c} = \frac{\sum_{i=1}^{p-1} x_i}{p-1} \qquad \textbf{(6)}$$

Step 4. Construct a direction in the variable space for moving the whole population to the reference point using reflection of solution $\boldsymbol{x}_{\max}$ towards the weight center *c*:

$$\boldsymbol{y} = \boldsymbol{c} - \boldsymbol{x}_{\max} \qquad \textbf{(7)}$$

Step 5. Move each solution $xi \in P$, for $i = 1,\ldots,p$; in the following manner:

1.) $\boldsymbol{x}i_{\text{new}} = \boldsymbol{x}i + \boldsymbol{y}$. In case the obtained $\boldsymbol{x}i_{\text{new}}$ violates some constraint from the system **(2-4)** change $\boldsymbol{y}$ by: $\boldsymbol{y} = 0.9\boldsymbol{y}$ and repeat 1.)

2.) Round off the obtained solution $\boldsymbol{x}i_{\text{new}}$ to the nearest integer feasible solution and include it in the new population P_{new}.

At this step some newly generated solutions may coincide. In this case the number of population members is reduced by 1.

Step 6. Set *icount* = *icount* + 1 and $P = P_{\text{new}}$. If *icount* > *itlim*, then STOP (the procedure terminates), else go to Step 3.

B. Improving direction strategy (IDS)

Let the reference point *fr* is known and let we have a population P of p solutions. Then perform the following procedure:

Step 1. Calculate the Euclidean distance measure in the criteria space between each point $fi \in P$, for $i = 1,\ldots,p$; and the reference point *fr* according **(5)**. Arrange all solutions $xi \in P$, for $i = 1,\ldots,p$; in a list D according their *di* distance measures in an ascendant order.

Step 2. Calculate the weight center of the first three solutions $\boldsymbol{x}i \in D$, for $i = 1, 2, 3$;:

$$\boldsymbol{cd} = \frac{\sum_{i \in D, i=1}^{3} x_i}{3} \qquad \textbf{(8)}$$

Calculate the weight center $\boldsymbol{c}$ of all other solutions $\boldsymbol{x}i \in D$, for $i = 4,\ldots,p$;

Step 3. Construct an improving direction in the variable space for moving the whole population to the reference point as follows:

$$\boldsymbol{y} = \boldsymbol{cd} - \boldsymbol{c} \qquad \textbf{(9)}$$

Step 4. Move each solution $xi \in P$, for $i = 1,\ldots,p$; in the following manner:

1.) $\boldsymbol{xi}_{\text{new}} = \boldsymbol{xi} + \alpha \boldsymbol{y}$, where α is a positive scalar. In case the obtained $\boldsymbol{xi}_{\text{new}}$ violates some constraint from the system **(2-4)** change α by: $\alpha = 0.9\alpha$ and repeat 1.)

2.) In case the corresponding *di*-value is decreased repeat 1.), otherwise go to 3.)

3.) Round off the obtained solution $\boldsymbol{xi}_{\text{new}}$ to the nearest integer feasible solution and include it in the new population P_{new}.

Step 5. Find all different non-dominated solutions $\boldsymbol{xi} \in P_{\text{new}}$ and arrange them in an ascending order of their Euclidean distances to the reference point *fr*. Then include them in the final population P_{fin}. STOP.

III. NEW EVOLUTIONARY MOCIP-ALGORITHM

The considered problem has a closed feasible domain because there are given lower and upper bounds for each variable (see constraint **(3)**). The Tchebycheff center of the feasible domain can be calculated and can be rounded off to the closest integer feasible point $\boldsymbol{x}^{ch}$, called below rounded Tchebycheff center.

The DM can give feasible or infeasible reference point(s). In both cases the proposed new algorithm finds a set of (near) Pareto-optimal solutions, which are located close to the supplied reference point(s). The found non-dominated points and their corresponding solutions during the search process are stored in an array *LND*, which is updated at each generation of new point.

The proposed MOCIP-algorithm contains the following steps:

Step1. Calculate the points $f^*_1, f^*_2, \ldots, f^*_k$; and their corresponding solutions x^*i, $i=1,\ldots,k$; , where f^*_i is the optimal point for the *i*-th objective. To obtain these points solve *k* single objective problems with the same constraint system **(2-4).**

Step 2. Create an initial population *P*o of size *p*, generating *p* uniformly distributed integer feasible solutions around the rounded Tchebycheff center $\boldsymbol{x}^{ch}$. Here *p* could vary according the number of objectives *k* in the current problem. For example: $p=2k$, when $k \geq 50$, $p=3k$ for $k \in [33 \div 50)$, $p=4k$ for $k \in [25 \div 33)$, $p=5k$ for $k \in [10 \div 25)$, etc.

Step 3. Ask the DM to specify a reference point *fr*.

Step 4. Perform *WPRS*.

Step 5. Ask the DM to specify the most acceptable, as well the most non acceptable obtained point (corresponding *xb* and *xw*) Based on the direction $v = xb - xw$ propose new reference point to DM, or ask her/him to specify a new reference point *fr*.

Step 6. Perform *IDS*. Include all obtained non-dominated solutions in the population P_{fin}.

Step 7. Show the obtained non-dominated points, corresponding to the non-dominated solutions $xi \in P_{fin}$, for $i = 1,\ldots,p_{fin}$; to the DM.

Step 8. Ask the DM if she/he wishes to perform small number genetic iterations using P_{fin}. If yes calculate for each point from P_{fin} utility coefficients $\eta_i = \dfrac{fi(x) - fi^*}{fi_{\max} - fi^*}$, $i=1,\ldots k$; where $fi_{\max}$ is the obtained maximal value of *i*-th objective during the search process. Calculate fitness values for the solutions in P_{fin}, and for the new generated offspring solutions, minimizing the following fitness function:

$$\boldsymbol{F}_i = \eta_i . di . (nd_i + 1) \qquad \textbf{(11)}$$

where *di* are the Euclidean distance measures in the criteria space between the point $fi \in P_{fin}$, for $i = 1,\ldots,p_{fin}$; and the reference point $\boldsymbol{fr}$. The value nd_i corresponds to the number of points, which eventually dominate the point *fi*.

Step 9. Show the solutions and their corresponding points in the population P_{fin} to the DM. Ask DM to enter a last reference point $\boldsymbol{fr}$, close located to the non-dominated solution $\bar{x} \in P_{fin}$.

Step 10. Perform the procedure, suggested by Wierzbicki [14], to create *k* new reference points around $\boldsymbol{fr}$ as follows:

$$\boldsymbol{fj} = \boldsymbol{fr} + (\bar{x} - \boldsymbol{fr}) . e^{(j)}, \; j=1,\ldots,k; \qquad \textbf{(12)}$$

where $e^{(j)}$ is the *j*-th coordinate direction vector.

Step 11. Perform simple search procedure for a better point along the direction $(\bar{x} - \boldsymbol{fr})$, starting with each created reference point $\boldsymbol{fj}$ $j=1,\ldots,k;$.

Step 12. Ask the DM to choose final solution among the non-dominated solutions found during the search process, stored in *LND*. STOP.

IV. ILLUSTRATIVE EXAMPLE

The performance of MOCIP-algorithm is illustrated on the following test example:

$$\textbf{Min } \boldsymbol{f}(\boldsymbol{x}) = [\boldsymbol{f1}(x), \boldsymbol{f2}(x), \boldsymbol{f3}(x)]^{\mathrm{T}},$$

$$\boldsymbol{f1}(x) = x_1^2 + 3x_2^2; \; \boldsymbol{f2}(x) = 5x_1^2 + x_2^2; \; \boldsymbol{f3}(x) = 2x_1^2 - x_2;$$

subject to:

$$-x_1 - x_2 + 11 \leq 0$$
$$0 \leq x_1 \leq 16,$$
$$0 \leq x_2 \leq 16,$$
$$x_1, x_2 \text{ – integer;}$$

At Step 1. The points f^*_1, f^*_2, f^*_3; are: $f^*_1 = (91, 329, 125)$, $x1^* = (8,3)$. $f^*_2 = (247,101,-1)$, $x2^*=(2,9)$, $f^*_3 = (768,256,-16)$, $x3^* = (0,16)$.

At Step 2. The rounded Tchebycheff center $\boldsymbol{x}^{ch} = (10,10)$. The following initial population is created: $x^{(1)} = (5,12)$, $x^{(2)} = (8,12)$, $x^{(3)} = (11,12)$, $x^{(4)} = (14,12)$, $x^{(5)} = (7,10)$, $x^{(6)} = (10,10)$, $x^{(7)} = (13,10)$, $x^{(8)} = (5,8)$, $x^{(9)} = (8,8)$, $x^{(10)} = (11,8)$, $x^{(11)} = (14,8)$;

At Step 3. The DM specifies the reference point $\boldsymbol{fr} = (100, 150, 45)$.

At Step 4. The worst point in the population is $x^{(4)} = (14,12)$, $f^{(4)} = (628,1124,380)$ with Euclidean distance measure $d4=1157.4476$. The weight center $\mathbf{c} = (9.2, 9.8)$. The direction for moving the whole population is $\mathbf{y} = (-4.8, -2.2)$. The obtained new population is: $x^{(1)} = (1,10)$, $x^{(2)} = (3,10)$, $x^{(3)} = (6,10)$, $x^{(4)} = (9,10)$, $x^{(5)} = (3,8)$, $x^{(6)} = (5,8)$, $x^{(7)} = (8,8)$, $x^{(8)} = (4,7)$, $x^{(9)} = (5,6)$, $x^{(10)} = (6,6)$, $x^{(11)} = (9,6)$; The worst point in this population is $x^{(4)} = (9,10)$, $f^{(4)} = (381,505,152)$ with

Euclidean distance measure $d4$=465.2258. The weight center **c** = (5, 7.9). The direction for moving the whole population is **y** = (-4, -2.1). The obtained new population is: $x^{(1)}$ = (1,10), $x^{(2)}$ = (2,9), $x^{(3)}$ = (3,8), $x^{(4)}$ = (5,8), $x^{(5)}$ = (4,7), $x^{(6)}$ = (5,6), $x^{(7)}$ = (6,5);

At Step 5. The DM specifies as most acceptable the solution $x^{(7)}$ = (6,5) with $f^{(5)}$ = (111,205,67); The most non acceptable point is $x^{(1)}$ = (1,10) with $f^{(1)}$ = (301,105,-8); The improvement direction between them is y = (5, -5). The new proposed reference point is ***fr*** = (100, 500, 200) with corresponding solution x = (10,0),. The weight center ***cd*** = (5,6). The weight center ***c*** = (2.75, 8.75). The direction for moving the whole population is **y** = (2.25, -2.75). The parameter α =1. The obtained new population is: $x^{(1)}$ = (3,8), $x^{(2)}$ = (4,7), $x^{(3)}$ = (5,6), $x^{(4)}$ = (7,6), $x^{(5)}$ = (6,5), $x^{(6)}$ = (7,4), $x^{(7)}$ = (8,3); During the repeated step in the same direction also the non-dominated solutions $x^{(8)}$ = (9,2), $x^{(9)}$ = (10,1) have been generated.

At Step 7. The obtained non-dominated solutions are: $x^{(1)}$ = (1,10), $x^{(2)}$ = (2,9), $x^{(3)}$ = (3,8), $x^{(4)}$ = (4,7), $x^{(5)}$ = (5,6), $x^{(6)}$ = (6,5), $x^{(7)}$ = (7,4), $x^{(8)}$ = (8,3), $x^{(9)}$ = (9,2), $x^{(10)}$ = (10,1);

At Step 8. the DM does not want to perform a genetic search procedure.

At Step 9. the DM refuses to enter a last reference point ***fr***.

At Step 10. and Step 11. don't have been generated any better points.

At Step 12. the DM chooses the solution $x^{(7)}$ = (7,4) with $f^{(7)}$ = (97,261,94) as the best compromise solution.

In case DM wishes, she/he can start the algorithm specifying another reference point. In this way DM could explore different parts of Pareto-optimal front.

V. CONCLUSION

The proposed MOCIP-algorithm has the following basic characteristics:

- It is designed to find a preferred set of solutions instead of the entire Pareto-optimal set.
- It can quickly converge to the desired part of Pareto-optimal front.
- It is indifferent to the shape of Pareto-optimal front.
- It is applicable to problems with large number of objectives and large number of variables.
- It does not put great demands to the DM.
- It is an interactive evolutionary method and could generate a number of solutions in the region of interest, so that the DM would be able to find without great efforts the satisfactory non-dominated solution among them.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the support of Bulgarian National Science Fund, Grant No DTK02/71 "Web-Based Interactive System, Supporting the Building Models and Solving Optimization and Decision Making Problems" and of European Social Fund and Bulgarian Ministry of Education, Youth and Science under Operative Program "Human Resources Development", Grant BG051PO001-3.3.04/40.

REFERENCES

[1] Branke J., Deb K., Miettinen K., Słowiński R., *„Multiobjective Optimization. Interactive and Evolutionary Approaches*", Springer-Verlag Berlin Heidelberg, 2008

[2] Coello C. A. C., D. A. Van Veldhuizen, and G. Lamont *Evolutionary Algorithms for Solving Multi-Objective Problems*, Boston, MA: Kluwer Academic Publishers, 2002.

[3] Deb K., *„Multi-Objective Optimization Using Evolutionary Algorithms*", Wiley-Interscience Series in Systems and Optimization. John Wiley & Sons, Chichester, 2001

[4] Deb K., J. Sundar, U. B. Rao N., S. Chaudhuri, "Reference Point Based Multi-Objective Optimization Using Evolutionary Algorithms", *International Journal of Computational Intelligence Research*, Vol. 2, No. 3 (2006), pp. 273-286.

[5] Garey M. R. and Johnson D. S. *Computers Intractability: A Guide to the Theory of NP-Completeness*, W. H. Freeman, San Francisco 1979.

[6] Goldberg, D. E. Genetic Algorithms in Search, Optimization and Machine Learning, Reading Mass, Addison Wesley, 1989.

[7] Guliashki V., (2008) „An Accelerating Technique for Population Based Algorithms", Proceedings of Papers of IX Internatrional Conference on Computer Systems and Technologies CompSysTech'08, Gabrovo, Bulgaria, 12-13 June, 2008, pp. -IIIB.11-1- ÷ -IIIB.11-6-.

[8] Guliashki V., L. Kirilov, "SPEA-Based Method for MCDM Convex Integer Problems", *Cybernetics and Information Technologies*, Vol. 9, No. 4 (2009), pp. 93-101.

[9] Ivanova J, "An Interactive Hybrid Algorithm of Integer Multiobjective Optimization", International Conference Automatics and Informatics'08, In: Proceedings "School for young Scientists Business Process management", (Ed. P. Ruskov), 2008, pp. 45-48.

[10] Miettinen K., P. Neittaanmäki, M. M. Mäkelä, J. Périaux, "*Evolutionary Algorithms in Engineering and Computer Science: Recent Advances in Genetic Algorithms, Evolution Strategies, Evolutionary Programming, Genetic Programming and Industrial Applications*", John Wiley & Sons, Chichester, 1999

[11] Nemhauser G.L. and Wolsey L.A. *Integer and Combinatorial Optimization*, Wiley, New York 1988.

[12] Pareto V. "*Cours d'Economie Politique*", Rouge, Lausanne, Schwitzerland, 1896.

[13] Praditwong, K.; Xin Yao, „How well do multi-objective evolutionary algorithms scale to large problems", *Evolutionary Computation*, 2007. CEC 2007. *IEEE Congress on*, 25-28 Sept. 2007, pp. 3959 – 3966.

[14] Wierzbicki A. P., "The use of reference objectives in multiobjective optimization". In G. Fandel and T. Gal, editors, Multiple Criteria Decision Making Theory and Applications, Berlin: Springer-Verlag, pp. 468-486.

End-to-End Queue Dimensioning in IP Network

Rossitza Goleva [1] and Seferin Mirtchev [2]

***Abstract* – This paper proposes analytical and simulation approach for QoS management in IP networks. It uses flexible queueing bounds for delay and loss combined with priorities. The analytical solution uses Priority Queueing. It can be applied also to Weighted Fair Queueing, Round Robin or other technique. The simulation is done for IntServ – RSVP and DiffServ. The results demonstrate better traffic shaping on real time traffic.**

***Keywords* – Queueing, IP Quality of Service management, Shaping, Wait and loss bounds.**

I. Quality of Service Management

Internet traffic is changing continuously and requires dynamic Quality of Service (QoS) estimation and management [1], [2]. TCP and UDP segments are different by size and QoS requirements. Packet sizes between 20 and 1500 bytes influence the servicing time. Output scheduling algorithms like Priority Queuing (PQ), Weighted Fair Queuing (WFQ), Round Robin (RR) or other applied in the queues change the distribution of the packet flows [4], [5]. Traffic policing and traffic shaping mechanisms, packet dropping and early random detection can change the characteristics of the flows.

The traffic profile is changing from hour to hour, weekly and due to the season. Effects like self-similarity and long tailed queues are a matter of long research. Internet Service Providers need to configure their equipment dynamically in order to meet the Quality of Service requirements. The mixture of voice and data traffic requires specific prioritization and reservation scheme that will allow conformance to the QoS requirements. Wired and wireless parts of the connections, access and transport devices gather the traffic in a different way. It is difficult to predict end-to-end QoS in a typical IP connection with many hops.

In his paper, Atov [2] has presented a combined DiffServ/MPLS approach that classifies the traffic depending on the delay and delay jitter requirements. MPLS is used for resource reservation to support delay bounds. DiffSerf is applied for prioritization of different aggregated traffic.

A comparison between Weighted Round Robin (WRR) and Weighted Fair Queueing (WFQ) for resource allocation is shown in [8]. A demonstration of the FTP traffic requirements and its management to fulfil the QoS requirements and channel capacity utilization criteria is presented in [10].

UDP and TCP traffic sources can generate symmetric or asymmetric traffic. UDP is symmetric in VoIP and asymmetric in TFTP service. UDP is also simplex in IPTV. TCP traffic is asymmetric in HTTP, email, FTP. It can be also symmetric for client-server applications. Bounds on delay in real time and non real time services can vary in quite high ranges. The delay can be close to the upper limits depending on the number of hops, processing delay, scheduling in the queues, credit fluctuations.

IntServ – RSVP and DiffServ occupy waiting places in the queue in different ways. Patchararungruang [10] proposes a simplified method of router representation applying fuzzy logic. His results are approximate. The model is not applicable for fast calculation and dynamic resource reservation. The capability of Pareto distribution to model data traffic and especially heavy tailed effect in the router interfaces is demonstrated in [1]. Chen [11] shows the capability of Multi-Reservation Multiple Access (MRMA) scheme to guarantee the delay bounds in access networks.

End-to-end delay and delay variation in TCP and UDP services under bursty traffic depends strongly on the traffic distribution, policing and shaping applied [6], [8]. Series length is important in delay and queueing places bounds. In paper [5], numerical results after simulation for QoS techniques IntServ – RSVP and DiffServ are presented. Similar approach can be also seen in [8].

The aim of this paper is to show course-grained and fne-grained approach for traffic shaping that will allow dynamic configurations and end-to-end QoS management. We show how the QoS requirements can be kept using combined DiffServ and Priority queueing approach.

II. Traffic Sources and Simulation Model

Three types of traffic sources are assumed in the network model – Voice over IP, LAN emulation, email. LAN traffic has lower priority in comparison to the VoIP traffic and higher priority in comparison to the email. In VoIP service silence and talk intervals are exponentially distributed. On-off model is applied. Short packets are considered as between 64 and 100 bytes. They carry up to 15 ms voice [12]. LAN emulation is specific with its TCP sessions. Sessions are established for any Internet connections [5]. LAN packets are between 20 and 1500 bytes. Emails are specific with packet exchange mostly in one direction. Packets are taken to be as long as 1500 bytes.

End-to-end Quality of Service requirements for real time services allow up to 150 ms end-to-end delay and up to 30 ms

[1] Rossitza Goleva is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, e-mail: rig@tu-sofia.bg

[2] Seferin Mirtchev is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, e-mail: stm@tu-sofia.bg

delay jitter. Packet loss probability is considered to be less that 0,1. The loss and delay bounds limit the total number of hops. The bounds for waiting times for TCP traffic are calculated on slow start timer that has typical value of 3 seconds. Servicing times per packets are simulated on 100 Mbps line interface.

III. Networks of Queues Approximation

In typical router interface, few queues classify the traffic depending on the port number. Many priority levels can be applied according to different Quality of Service mechanisms like IntServ – RSVP and DiffServ. All of them classify the traffic in input queue and identify scheduling technique to the output line (Fig. 1).

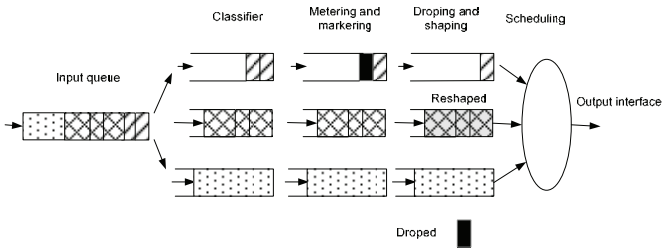


Fig. 1. Parallel Queues

Typical IP connection has many hops. It can be modelled with cascaded Leaky Bucket queues with priorities (Fig. 2). The sequence of queues can be considered to have equal characteristics with exceptions in access queues at both ends of the connection. In access part of the model limits for waiting place and waiting times should be more relaxed in comparison to the others. Access routers are usually less powerful than core network routers.

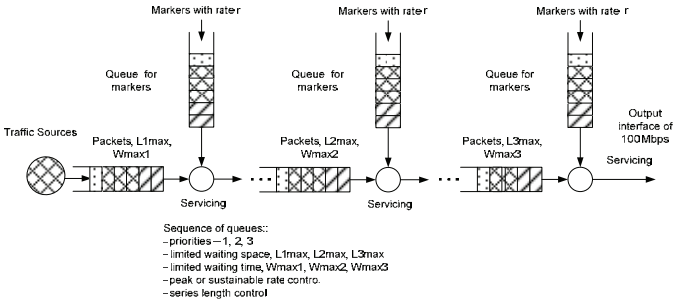


Fig. 2. Sequence of Queues

The output interface can be easily considered as single Leaky Bucket queue with priorities (Fig. 3). Limits to the queue length L_{imax} and waiting times W_{maxi} are calculated in order to satisfy requirements for Quality of Service. L_i are the current lengths of the queue fractions depending on the type of packets.

Cascaded queues accumulate waiting times and end-to-end losses. The overall throughput of the connection is the minimal throughput of the nodes in the connection. All routers in the connection see round trip delay and are capable to count total loss. So, the structure of the model shown on Fig. 2 can be applied end-to-end keeping in mind that the delay and loss bounds are cumulative products of the delays and losses bounds in the nodes of the connection.

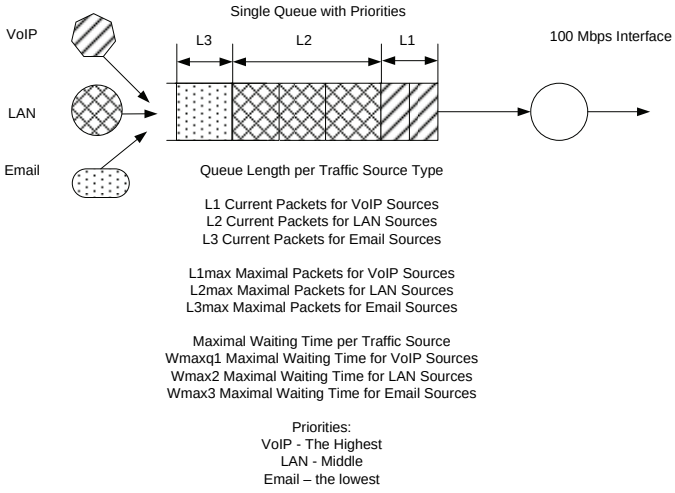


Fig. 3. A single node QoS model and end-to-end QoS model

Different scheduling techniques like Priority Queuing (PQ), Weighted Fair Queuing (WFQ) and Round Robin (RR) can be considered for scheduling in parallel queues. Priority queueing technique is presented by Kleinrock in [3]. He derived the formula for the priority scheme with preemptive and non preemptive disciplines taking into consideration second moments of the queueing delay and propagation delay as well as distribution of the packets. The idea of Kleinrock is not directly applicable to the router interfaces because of its complexity.

Below, we try to simplify the idea and make it more applicable for dynamic interface dimensioning. We show that queue scheduling influence significantly the end-to-end behaviour. The model shown on Fig. 3 is considered. Delay bounds depend on slow start procedure for TCP traffic and real time transmission requirements for VoIP. Calculation of the waiting times depends on type of priority, IntServ-RSVP/DiffServ application and scheduling algorithms. For UDP traffic short queue for VoIP is modeled. Long queue is applied for LAN emulation traffic that is mostly http. SMTP traffic can be transmitted with lower priority in comparison to the http traffic.

Parameters of the Leaky bucket queues depend on type of services. In the simulation model, bigger packetization and depacketization delays in the access routers are considered. By changing the delay and loss bounds and priorities it is possible to adjust router interfaces behaviour depending on the real time circumstances in the networks. It can be done per aggregated service and aggregated packet level. The model applies three priorities.

IV. Results from Simulation

Simulation is performed on C++ language. The waiting time and loss bounds are calculated in accordance to DiffServ [14], [15]. Queue behaviour is complex due to the priorities

and limits on waiting times and places. Many parameters have been derived from the model including probability of packet loss due to the lack of place in the queue, probability packet to be dropped due to the exceed of waiting limit, probability to wait for different types of traffic, observations on of the distribution of packets intervals, queue lengths, delay, delay jitter. Statistical accuracy of the derived results is proven by batch mean method for output results analysis with Student distribution and confidence probability 0.95 (Table I).

TABLE I. NUMERICAL RESULTS ON UTILIZATION AND LOSS PROBABILITIES FOR DIFFSERV

VoIP maximal delay, s	0.002	0.001	0.00004	0.0001	0.0001
VoIP queue limit	10	5	7	3	2
LAN maximal delay, s	0.07	0.06	0.0004	0.04	0.03
LAN queue limit	50	25	24	5	5
Email maximal delay, s	0.09	0.08	0.0004	0.06	0.04
Email queue limit	50	25	24	5	5
Queue size	110	55	55	13	12
Probability of losses due to the lack of place	0	0	0.18272	0.0265	0.0881
Prob. of place losses for VoIP	0	0	0	0.0001	0.0056
Prob. of place losses for LAN	0	0	0.18038	0.0309	0.0978
Prob. of place losses for email	0	0	0.54851	0.0017	0.1389
Prob. of losses due to the exceed of waiting limit	0	0	0.00009	0	0
Prob. of waiting losses for VoIP	0	0	0.00062	0	0
Prob. of waiting losses for LAN	0	0	0	0	0
Prob. of waiting losses for email	0	0	0	0	0
Prob. to wait observed	0.495	0.4673	0.6945	0.6436	0.7175
Prob. to wait observed for VoIP	0.522	0.5097	0.88853	0.6934	0.8259
Prob. to wait observed for LAN	0.492	0.4634	0.69823	0.6380	0.7065
Prob.to wait observed for email	0.489	0.4643	0.29602	0.6312	0.6336
Utilization	0.473	0.4585	0.82179	0.6132	0.752
Utilization for VoIP	0.045	0.0437	0.13892	0.0681	0.0997
Utilization for LAN	0.423	0.41	0.64798	0.523	0.6150
Utilization for email	0.005	0.0048	0.03489	0.0221	0.0373

Interesting results that influence directly interfaces and queue management are derived based on queue length per service type. The queue fractions of the three services are observed. Under almost the same utilization factor, the fractions of the queue per service is changed. Scenario limits are shown on the first six lines.

On the third column of the table, it can be seen that tight bound in waiting time of "0.00004" will keep the VoIP queue short with probability of waiting time loss of "0.00062". The total load on the interface is significant (0.82179) but the occupation of the interface by VoIP packets is very low in comparison to other cases (0.13892). This means that the influence of the waiting time limit is significant in the system. The system also shows the deterministic nature of the IP traffic. The same effect can be seen on the last column.

The adjustments of the waiting time and waiting place bounds are considered equal along the connection with exception for access routers. It can be done also proportionally depending on the results for round trip delay from traceroute measurement.

V. ANALYTICAL MODEL FOR QUALITY OF SERVICE MANAGEMENT

The simulation model is used for the end-to-end cumulative QoS analyses of the model in Fig. 2. The proposed simple analytical algorithm supports end-to-end maximal limits calculations per packets, per aggregated traffic and per scheduling algorithm.

Kleinrock in his study on Internet traffic [3] has derived formulae for different packet traffic. His derivations are based on end-to-end round trip delay. They can be applied for end-to-end analyses of the traffic with known distribution and profiles. It cannot be applied dynamically at the router interfaces. The proposed approach is simpler and applicable for fast dimensioning.

Queue limits for TCP traffic is calculated from number of hops and slow start timer. ICMP requests and traceroute requests return the minimal, maximal and mean delay. Upper limits of waiting time for VoIP traffic is also calculated using number of hops and 150 ms end-to-end maximal bound. Waiting time bounds calculated analytically are compared to the results from simulation. VoIP traffic is considered to be modelled like ON-OFF source. 1250 VoIP on-off traffic sources with intensity of 0.05 Erl are aggregated.

In Priority Queuing (PQ) mechanism, every packet from the queue with higher priority will be send before any packet from the queue with lower priority. The nonpreemptive queueing is considered. In case the VoIP traffic is dominated in the network, other traffic will suffer of big delays. Minimal delay on fragmentation at both ends is equal to the voice buffer, i.e. 20-30 milliseconds. Maximal waiting time per queue $W_{maxVoIP}$ is calculated with Eq. (1) (Table II). Maximal waiting time bound for LAN packet $W_{\max LAN}$ depends on existing number of VoIP packets because they are of higher priority. In equations, t_1 and t_2 are the time necessary for the router interface to send VoIP and LAN packet, n is the number of VoIP packets, m is the number of LAN packets, p is the number of email packets in the queue. The delay bound for real time services is kept on the favour of the non real time services. Every packet of type 3 will wait packets from type 1, 2 and 3 in the schedule to be served. Furthermore, new packets from type 1 and 2 arriving in the queue input will be served before the waiting packet of type 3. This discipline is serving high priority traffic in a very good way.

$$W_{\max VoIP} \geq nt_1 \quad W_{\max LAN} \geq nt_1 + mt_2 \qquad (1)$$

$$W_{\max email} \geq nt_1 + mt_2 + pt_3 .$$

The limit for queue length calculated as an upper one is considered as course-grained bound. The waiting time limit is considered to be fine-grained bound. Two simple instruments for interface adjustments are obtained. The maximal waiting time for LAN traffic is corrected with the maximal waiting time for VoIP traffic. The value is less that the one practically calculated from RTT. The limits for email and other lower

priority services are shared with the other TCP traffic. Time for acknowledgement service is considered negligible. Calculations are made for 100 Mbps interface.

TABLE II.

NUMERICAL RESULTS ON DELAY BOUNDS IN PQ

VoIP service (traffic)		
Low delay bound	7.62939E-06	sec
High delay bound	0.0048	sec
Servicing time for 100 bytes packet	7.62939E-06	sec
Number of hops	25	
Packetization delay	0.03	sec
Upper number of packets in queue	629.1456	packets
n –VoIP packets	600	
Queue length for VoIP packets	62914.56	bytes
Waiting time for VoIP more than	0.004577637	sec
LAN service traffic		
m - 200 LAN packets	1460.4352	packets
Packets length	1000	bytes
Servicing time per packet	7.62939E-05	sec
Low delay bound	7.62939E-05	sec
High delay bound per router	0.111422363	sec
Waiting time for LAN packets more than	0.116	sec
Queue length for LAN	2190652.8	bytes
p - email packets	50	
Servicing time packet	7.62939E-05	sec
Low delay bound	7.62939E-05	sec
High delay bound per router	0.111422363	sec
Waiting time for email packets more than	0.119814697	sec
Queue length for email	75000	bytes

VI. CONCLUSION

In this paper, we demonstrate similar to [2] approach for aggregated DiffServ traffic without use of MPLS capabilities. We show that due to the delay bounds in the queues techniques like Weighted RED algorithms can be avoided. Instead of WRED packets that delay will be dropped. The approximate approach to network dimensioning ([6], [8]) with single server queue with priorities and place and time bounds is proven to be effective. The approach can be used for TCP and UDP traffic as well as for sequence of queues and parallel queues.

The capability of the algorithm at router interface to shape traffic from different traffic sources is shown. The approach allows some services to obtain better service conditions on the favour of others. Limits criteria for queue management as coarse-graned and fine-grained adjustments are evaluated. Cumulative traffic with different distribution and without specific distribution of the packets is studied. Combination with PQ prove possibility to implement the ideas on traffic shaping on the interface and end-to-end.

With numerical results, we also prove that shaping is effective under interface load above 0.6. Therefore, it can be applied as congestion management technique. Because of the simplicity of the proposed algorithm, it can be implemented in real time queues for fast reconfiguration and QoS dynamic management. The derived results are applicable to the routers that are capable to classify packets and manage dynamically queues and priorities.

ACKNOWLEDGEMENT

This paper is sponsored by the COST Action IC0703 “Data Traffic Monitoring and Analysis (TMA), research programme at Technical University of Sofia, and Mtel Ltd., Bulgaria.

REFERENCES

[1] Ariffin, S.: Accelerated Simulation for Packet Buffer with a Non FIFO Scheduler. Submitted for the Degree of Doctor of Philosophy, Queen Mary University of London, March (2006)

[2] Atov, I., Harris, R.J.: Capacity Provisioning in DiffSrv/MPLS Networks Based on Correlated Traffic Inputs. Performance, Computing, and Communications, 2004 IEEE International Conference, pp. 187- 193 (2004)

[3] Kleinrock, L.: Queueing Systems. Volumes I and II, John Wiley and Sons (1976)

[4] Ralsanen, V.: Implementing Service Quality in IP Networks. John Wiley & Sons, Ltd. (2003)

[5] Goleva, R., Atamian, D.: End-to-End Delay Analyses in IP Networks. ICEST 2008, pp. 40-43, Nish, Serbia (2008)

[6] Peng, G.: Availability, Fairness, and Performance Optimization in Storage Virtualization Systems. A dissertation, The Graduate School, Stony Brook University (2006)

[7] Iversen, V.: Teletraffic Engineering Handbook. ITU-D (2005)

[8] Victor, V., Záruba, G., Balasekaran, G.: RECHOKe: A Scheme for Detection, Control and Punishment of Malicious Flows in IP Networks. IEEE GLOBECOM (2007)

[9] Fernekes, A., Klein, A.: Influence of High Priority Users on the System Capacity of Mobile Networks. WCNC (2007)

[10] Patchararungruang, S., Halgamuge, S., Shenoy, N.: Optimized Rule-Based Delay Proportion Adjustment for Proportional Differentiated Services. IEEE Journal on Selected Areas in Communications, Vol. XX, No. Y, January (2005)

[11] Chen, H., Fei Yu, Chan, Henry C. B., Leung, Victor C. M.: A Novel Multiple Access Scheme over Multi-Packet Reception Channels for Wireless Multimedia Networks. IEEE Transactions on Wireless Communications, Vol. 6, No. 4, April, pp.1501 (2007)

[12] Mirtchev, S.: Stud y of Queueing Behaviour in IP Buffers. iTech’07 Conference, Varna, St. Kostantin and Elena, June, pp. 187-193 (2007)

[13] Mirtchev, S., Goleva, R.: A Discrete Time Queueing Model with a Constant Packet Size. ICEST 2009, pp. 173-176, V. Tarnovo, Bulgaria (2009)

[14] Mirtchev, S., Goleva, R.: Study of Discrete Time Pareto/D/1/k Queue by Simulation. Telecom 2009, 8-9 October, pp. 306-313, St. Constantine, Varna, Bulgaria (2009)

[15] Mirtchev, S., Goleva, R.: Evaluation of Pareto/D/1/k Queue by Simulation. International Book Series “Information Science&Computing” No.1, Supplement to the International Journal „Information Technologies & Knowledge”, Vol.2, pp.45-52 (2008)

Application of Support Vector Machines On Medium Term Load Forecasting

Miloš B. Stojanović[1], Miloš M. Božić[2] and Milena M. Stanković[3]

***Abstract* – Medium term load forecasting using Support Vector Machines (SVM) is presented in this paper. The forecasting is performed for the electrical maximum daily load of the period of one month. The data considered for forecasting consist of quarter hours daily load for the City of Niš for one year, daily average temperature and daily average wind speed. An analysis of available data was performed and most important attributes for the formation of the SVM model are selected. It was shown that the size and structure of the training set may significantly affect the accuracy of load forecasting. The results obtained from training set with and without involved average wind speed data are presented and the results shows the approach without wind speed data is more accurate.**

***Keywords* – Load forecasting, Support vector machines, Regression, Time series.**

I. INTRODUCTION

Load forecasting has always been important for economic and reliable operation of an electric utility. However, with the deregulation of the energy industries, load forecasting is even more important. With supply and demand fluctuating and the changes of weather conditions and energy prices increasing more during peak situations, load forecasting is vitally important for utilities. Load forecasting helps an electric utility to make important decisions including decisions on purchasing and generating electric power, load switching, and infrastructure development. Load forecasts can be divided into three categories based on prediction time period: long term, medium term and short term. Long term forecasts are related to time period which are longed then a year, medium forecasts are usually from a week to a year and short term forecasts which are usually from one hour to one week.

Load forecasting methods mostly use statistical techniques or artificial intelligence algorithms such as regression, neural networks, fuzzy logic, and expert systems. The use of Artificial Neural Networks (ANN) [7] has been a widely studied electric load forecasting technique. More recent powerful technique for use in electric load forecasting is SVM [4]. Unlike ANN, which try to define complex functions of the input feature space, SVM perform a nonlinear mapping (by using so-called kernel functions) of the data into a high dimensional space. SVM then use simple linear functions to create linear decision boundaries in the new space. The problem of choosing an architecture for a ANN is replaced in SVM by the problem of choosing a suitable kernel.

Medium term load forecasting using SVM are presented in this paper. The paper is organized as follows. In Section II, SVM techniques are presented. Data analysis is given in Section III. Section IV explains and describes our model in detail. Section V presents experimental results and provides model evaluation, and the conclusions are given in Section VI.

[1]Miloš B. Stojanović is with the Faculty of Electronics Engineering, A. Medvedeva 14, 18000 Niš, Serbia, E-mail: milos.stojanovic@vtsnis.edu.rs

[2]Miloš M. Božić is with the Faculty of Electronics Engineering, A. Medvedeva 14, 18000 Niš, Serbia, E-mail: milos1bozic@yahoo.com

[3]Milena M. Stanković is with the Faculty of Electronics Engineering, A. Medvedeva 14, 18000 Niš, Serbia, E-mail: milena.stankovic@elfak.ni.ac.rs

II. SVM

SVM are developed by Vapnik and assistants in 1995. to resolve the issue of data classification. Two years later, the version of SVM is proposed that can be successfully applied to the data regression problem. This method is called Support Vector Regression (SVR) and it is the most common form of SVM that is applied in practice [1]. SVM are based on the principle of structural risk minimization (SRM), which is proved to be more efficient than the empirical risk minimization (ERM), which is used in neural networks. SRM minimizes an upper bound of expected risk as opposed to ERM that minimizes the error on the training data [2].

A brief review of SVR method is given, which is successfully used to predict the data that are organized as time series. The goal of SVR is to generate a model (function) which will perform prediction of unknown output values based on the known input parameters. In the training phase the formation (training) of the model is performed based on the known training data (x_1, y_1), (x_2, y_2), …, (x_n, y_n), where x_i are input vectors, and y_i outputs associated to them. Each input vector consists of numeric attributes. In the test phase the trained model on the basis of new inputs $x_1, x_2, ...,x_n$ makes prediction of output values $y_1, y_2, ...,y_n$. SVR is an optimization problem in which is needed to determine the parameters ω and b to minimize [3]:

$$\min_{\omega,b,\xi,\xi^*} \frac{1}{2}\omega^T\omega + C\sum_{i=1}^{n}(\xi_i + \xi_i^*), \quad (1)$$

with conditions:

$$y_i - (\omega^T - \phi(x_i) + b) \le \varepsilon + \xi_i, \quad (2)$$

$$(\omega^T\phi(x_i) + b) - y_i \le \varepsilon + \xi_i^*,$$

$$\xi_i, \xi_i^* \ge 0, i = 1, 2, \dots, n,$$

where x_i is mapped in the multi-dimensional vector space with mapping ϕ, ξ_i is the upper limit of training error and

$\xi_i^*$the lower. Mapping ϕ is also known as kernel function. The parameters that control the quality of the regression are mapping ϕ, and the constants C and ε. C is a constant which determines the "cost of error", i.e. determines the tradeoff between the model complexity and the degree to which deviations larger than ε are tolerated. Parameter ε controls the width of ε insensitive zone, used to fit the training data [9]. The goal of SVR is to place as many input data (vectors x_i) inside the tube $|y - (\omega^T\phi(x) + b)| \leq \varepsilon$, which is shown on Fig 1. If x_i is not inside the tube, an error occurs ξ_i or ξ_i^*.

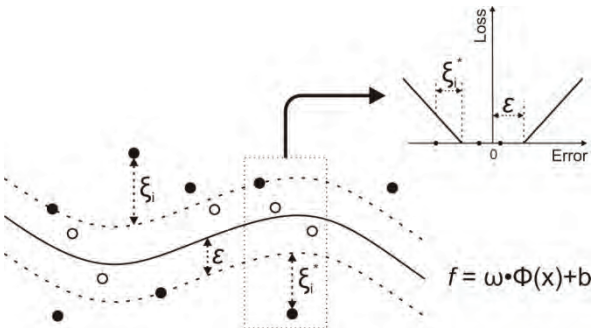


Fig. 1. ε tube of nonlinear SVR

In the least-square regression ε is always zero and the data are not mapped in a multidimensional space. Therefore, SVR provides greater flexibility and can be applied to the various sets of problems [4].

For the experiments in this paper we used a publicly available library LibSVM - A Library for Support Vector Machines [5] in C# version.

III. Data Analysis

Before we proposed the solution, some observations about properties of load demand are examined first. Some relations between load demand and other information like time and climate influence are identified.

Behavior of load demand: Load demand data given are quarter hours recorded from 2008. to 2009. for City of Niš. On Fig 2. maximum daily loads from 2008. to 2009. are presented.

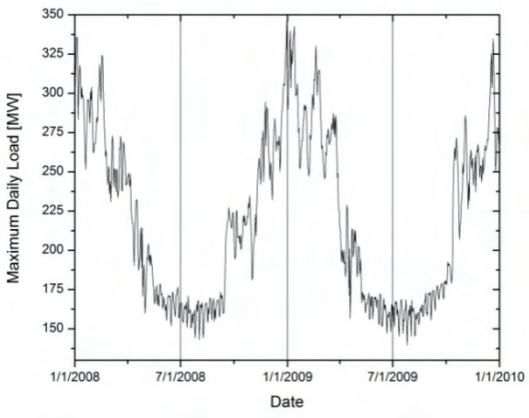


Fig. 2. Load pattern during period of two years

Relations between electricity usage and weather conditions in different seasons are observed from data on Fig 2. Load demand in winter period is higher compared to summer period. Additionally, can be observed that the seasonal load pattern is almost the same for both years. Also, load pattern periodicity exists in every week. Usage of electricity is relatively constant during week days and drops in the weekend.

Climate influence: Relation between climate and load demand also was shown in previous works on short term load forecasting [4,7]. Climate conditions influence on load demand pattern analysis may include temperature, wind speed, humidity, pressure and illumination. In this paper two climate parameters are considered, temperature and wind speed.

From Fig 3. can be easily observed negative correlation between maximum daily load and temperature. The correlation coefficient between maximum daily load and temperature is -0.942, which tells us that there is very strong connection between these two variables. Reason for that is heating use in winter period of year, higher temperature causes lower load demands.

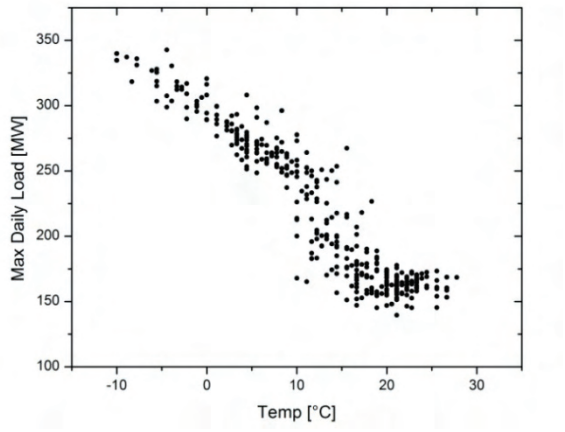


Fig. 3. Correlation between the max daily load and the temperature

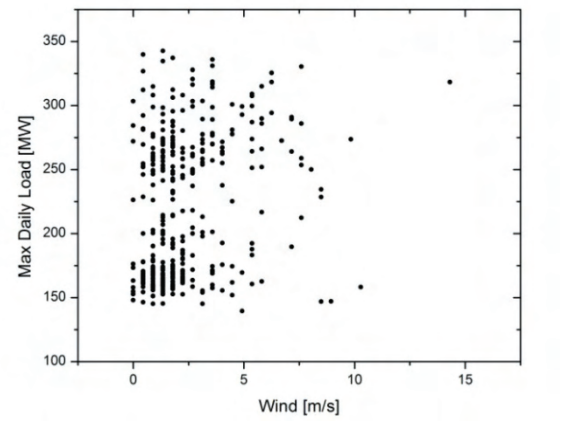


Fig. 4. Correlation between max daily load and the wind speed

Another climate parameter which may have significant influence on load demand is wind speed. Correlation between wind speed and maximum daily load is presented on Fig. 4. and this correlation coefficient is 0.263, which implies weak connection between these variables, although it was not initially expected. Because of that, two different models are

built, one with wind speed included as feature and one without wind speed.

Holiday influence: Holidays and other local events also may affect the load demand. These events are usually local and their influences highly depend on the customs of the area. Major holidays such as Christmas or New Year have more influences on load demand then other holidays.

IV. Data Preparation and Model Formation

Here, we considered what kind of information should be included to prepare datasets to build SVM models. Each component of input vectors is called a feature (attribute), and it is very important to identify real features.

Calendar attributes. In Section III, weekly pattern of load demand is discussed. We pointed out earlier that, the load demand on week days is relatively constant and higher than in the weekend. Also, the load demand on holidays is lower than on non-holidays. This information (weekdays and holidays) is useful for encoding in training entries. Actually, many works [4,7,8] have used the calendar information to model the problem.

Temperature. The temperature data is another possible feature, since load demand and temperature have causal relation in between. Weather information which includes temperature, wind speed, sky cover, humidity and etc., has also been used in most load forecasting works. But there is one difficulty: for mid-term load forecasting, temperature data for several weeks away are needed.

Wind speed. Beside temperature, wind speed is another climate parameter which has important influence on load demand. As with temperature date, if we want to encode the wind speed in our training entries, we will also need forecasted wind speed data for several weeks away.

Time series style. Another information which we considered to encode as the feature is the past load demand. With this approach, concept of time-series is introduced into our models [8]. If y_i is the target value for prediction, the vector x_i includes several previous target values $y_{i-1},\ldots,y_{i-\Delta}$ as attributes. In the training phase all y_i are known but for future prediction, $y_{i-1},\ldots,y_{i-\Delta}$ can be values from previous predictions. This approach is used because we know, the past load demand could affect and imply the future load demand. In this paper, we used Δ=7, which means that seven past daily load demands are used for attributes.

The various factors that affect the load forecast are analyzed and appropriate features are chosen, from which we form vectors which will be used as inputs for SVM. Our proposed model is shown on Fig 5. Input vectors for SVM model is composed of the following features:

- Maximum daily load for past seven days (L_{i-k}), k=1,…,7,
- Average daily temperatures (T_i),
- Average daily wind speed (W_i),
- Day Of the Week (DOW).

On the basis of established vectors and the known values of the load in the selected time interval SVM generates model which forecasts the load for a period of one month (training of the model). Trained model is forecasting the load for a period of one month (the application of the model).

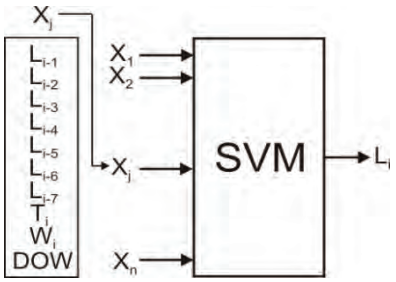


Fig. 5. Proposed architecture

In addition to the selected attributes, in order to get "good" model, we should determine the following parameters:

1. Mapping ϕ (kernel),
2. Constant C ("cost of error"),
3. Constant ε (width of tube).

For mapping ϕ RBF kernel is selected (Radial Basis Function) which is the most widespread in practical use:

$$\phi(x_i)^T\phi(x_i) = e^{-\gamma\|x_i-y_i\|^2}, \gamma > 0.$$

For constant ε it was chosen the value of 0.1. It was found that the choice of ε (tested values were in range of 0.01 to 0.5 with the step of 0.01) does not affect significantly on the accuracy of load forecasts.

It is necessary to determine the parameters C and γ. It is not known in advance which values of these parameters are best choice for a given problem. The parameters were determined using Cross - Validation procedure and Grid - Search [6].

The training set is randomly divided into training and testing parts, for example in relation 1:9. Then the learning algorithm is applied to training part and evaluation of the quality of prediction is performed on testing part. This procedure is repeated n times e.g. 10 and we selected a pair C and γ with which the best accuracy is achieved. Couples C and γ should be search exponentially, for example, $C = 2^{-5}$, 2^{-3}, ... , 2^{20} , $\gamma = 2^{-15}$, 2^{-13}, ... , 2^{3} [6]. Obtained large values for C in range 2^{18} - 2^{22} generates models that very well performs data fitting, i.e. tolerates very few errors. Values for γ in range 2^{-5} - 2^{-7} are near to default values proposed by LibSVM (γ=1/n, where n is number of attributes in vector).

V. Experimental Results

In electricity load forecasting, the prediction accuracy is generally evaluated using Mean Absolute Percentage Error (MAPE) [10]. The equation describing this error is:

$$MAPE = 100 \cdot \frac{1}{n}\sum_{i=1}^{n}\left|\frac{P_i - \hat{P}_i}{P_i}\right|, \qquad (3)$$

where P_i and $\hat{P}_i$ are the real and the predicted value of maximum daily electrical load on the i^{th} day and n is the number of days in the month.

To evaluate the accuracy of the model, maximum daily load forecasting for January 2010 is done. Training model was committed with several different data encodings and segmentations.

TABLE I
MAPE USING DIFFERENT DATA PREPARATION

Segments	With wind speed	Without wind speed
All	4.25%	3.84%
Winter	4.42%	3.56%
Dec-Jan-Feb	5.49%	3.44%
Summer	15.68%	15.28%

Table I shows MAPE errors generated by different data encodings and segmentations. The first column represents the data segments used, and then the next two columns show the prediction with and without the wind speed feature. In Table I, it can be observed that model built without wind speed feature using the "Dec-Jan-Feb" data, exceeds all others. Trained models with segments formed from "Winter" and "Dec-Jan-Feb" data generally have smaller errors compared to training models with "All" and "Summer" data segments.

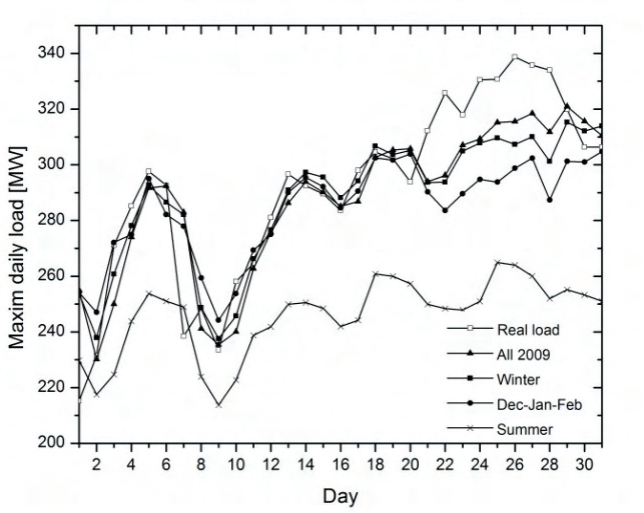


Fig. 6. Estimates and real load demand in Jan. 2010, with all features

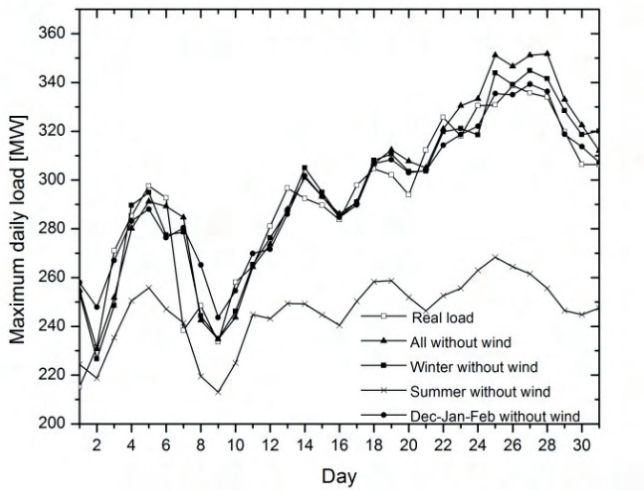


Fig. 7. Estimates and real load demand in Jan. 2010, without wind speed feature

In Fig 6. and Fig 7., the estimates and real load demand in January 2010 are shown, with all features and without wind speed feature respectively. An interesting observation can be made from the results: models built without the wind speed feature generally perform better than those built with. Also, the most deviations from real load are in days 1 and 7. Reason for that is these days are holidays, Ney Year and Christmas. As we mentioned earlier, holiday events also affect the load demand and firstly holiday information have been involved in our model. The problem is that the historical load data on holidays are not enough to train the model well, and this makes the behavior of load demand on these days hard to predict. However, later we remove holiday information from our model.

VI. CONCLUSION

In this paper several SVM models are made with different data segments. We find that choosing appropriate data segments may improve performance of models, since the load demand appears to have different distribution in different seasons. Furthermore, models build with climate information (temperature and wind speed) require future climate data for several weeks away. This difficult may lead to inaccurate prediction and because of that models without wind speed feature were build. Rejection wind speed feature from models was significantly increase models performance. However, the use of causal relation between temperature and load demand might need further consideration.

REFERENCES

[1] D. Basak, S. Pal, D.C. Patranabis, "Support Vector Regression", Neural Information Processing – Letters and Reviews, Vol. 11, No. 10, October 2007.

[2] S. R. Gunn, "Support Vector Machines for Classification and Regression", Technical Report, Image Speech and Intelligent Systems Research Group, University of Southampton, 1997.

[3] D. Basak, S. Pal, D. C.Patranabis, "Support Vector Regression", Neural Information Processing – Letters and Reviews, Vol. 11, No. 10, October 2007.

[4] A. Jain, B. Satish, "Clustering based Short Term Load Forecasting using Support Vector Machines", Bucharest Power Tech Conference June 28th - July 2nd 2009.

[5] C. C. Chang, C. J. Lin, "LibSVM: a library for support vector machines", Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm/.

[6] C. W. Hsu, C. C. Chang, C. J. Lin, "A Practical Guide to Support Vector Classification", Department of Computer Science National Taiwan University, Taipei 106, Taiwan.

[7] D. Ortiz-Arroyo, M. Skov, Q. Huynh, "Accurate Electricity Load Forecasting with Artificial Neural Networks", Proceedings of the 2005 International Conference on Computational Intelligence for Modeling, Control and Automation.

[8] N. Amjady, "Short-term hourly load forecasting using time-series modeling with peak load estimation capability", IEEE Transactions on Power Systems 16(3):498–505, August 2001.

[9] V.Cherkassky, Y.Ma, "Selection of meta-parameters for Support Vector Regression", International Conference Madrid, Spain, August 28-30, 2002 Proceedings.

[10] P. Bunnoon, K. Chalermyanont, C. Limsakul, "A Computing Model of Artificial Intelligent Approaches to Mid-Term Load Forecasting: a state-of-the-art survey for the researcher", International Journal of Engineering Technology, Vol. 2, No. 1, February 2010.

SESSION CSIT II

Computer Systems and Internet Technologies II

Supply Chain Management Information System

Vladimir D. Simić[1] and Branka S. Dimitrijević[2]

***Abstract* – It is no surprise that IT played a big role in enabling many processes and ideas in supply chain management (SCM) that seemed impossible in earlier years. Information system (IS) is no longer simply a support for supply chain (SC) operations; it is at the heart of its operations. In order to achieve better results in the SC, supply chain management information system (SCM IS) need to be developed.**

SCM IS is information system used to coordinate information flow between internal and external customers, suppliers, distributors, and other participants in a SC.

The performance of the SCM IS greatly relies on its infrastructure. The proposed SCM IS embraces the concept of distributed object technology to enable efficient data exchange among various SC participants which may reside in distributed platforms.

***Keywords* – Information technology, Information system, Supply chain management.**

I. INTRODUCTION

The adoption of information technology (IT) along a supply chain has become a necessity for enhancing supply chain (SC) performance [1].

IT played a big role in enabling many processes and ideas in supply chain management (SCM) that seemed impossible in earlier years. It is a critical enabler of effective SCM. Indeed, much of the current interest in SCM is motivated by the opportunities that appeared due to the abundance of data and the savings that can be achieved by sophisticated analysis of these data [2].

IT-based SCM systems, such as supply chain management information system (SCM IS) coordinate and integrate the flow of products, information, and finances between supplier, manufacturer, wholesaler, retailer and end consumer. They serve to create a multiplicity of digital options to be applied for operational, tactical and strategic purposes.

The task of SCM IS is to collect, exchange, synthesize and analyze the life cycle information necessary to support decision making process and help in identifying and implementing features that will maximize economical effect and minimize environmental degradation.

Successful implementation of the SCM IS requires a distributed computing and controlling infrastructure to connect with the computer systems of all SC participants. To meet this objective, the distributed object-oriented technology can play an important role in terms of easy access and sharing of information across a distributed IS. In this paper, we preferred distributed component object model (DCOM)-based infrastructure of the SCM IS.

The remainder of the paper is structured as follows: Section II gives some comments about SCM topic. Section III introduces basic concepts and framework of supply chain management information system. Finally, concluded remarks are given in Section IV.

II. SUPPLY CHAIN MANAGEMENT BASICS

Interest in SCM has grown rapidly over the past several years, and continues to grow. At the same time, information and communication systems have been widely implemented, providing access to comprehensive data from all components of the SC [2].

SCM has many definitions, all with a similar underlying theme of integrating the firm's internal processes with suppliers, distributors, and customers. Perhaps the most often cited definition comes from the Council of supply chain management professionals (CSCMP) [3]: "SCM encompasses the planning and management of all activities involved in sourcing and procurement, conversion, and all logistics management activities. Importantly, it also includes coordination and collaboration with channel partners, which can be suppliers, intermediaries, third-party service providers, and customers. SCM integrates supply and demand management within and across companies."

SCM is an integrating function with primary responsibility for linking major business functions and business processes within and across companies into a cohesive and high-performing business model. It includes all of the logistics management activities, as well as manufacturing operations, and it drives coordination of processes and activities with and across marketing, sales, product design, finance, and information technology [4].

The existing literature on SCM is extensive (comprehensive review is given, for example, in [5]) and mainly proposes that integrated control of these "multi-company networks" can provide significant benefits.

The ability of customers to access information is becoming an essential requirement in SCM, because visibility of information is what an increasing number of customers expect. The Internet enables these capabilities, and companies will need to invest in SCM IS that supports it.

[1]Vladimir D. Simić is with the University of Belgrade, Faculty of Transport and Traffic Engineering, Vojvode Stepe 305, 11000 Belgrade, Serbia, E-mail: vsima@sf.bg.ac.rs

[2]Branka S. Dimitrijević is with the University of Belgrade, Faculty of Transport and Traffic Engineering, Vojvode Stepe 305, 11000 Belgrade, Serbia, E-mail: brankad@sf.bg.ac.rs

III. SUPPLY CHAIN MANAGEMENT INFORMATION SYSTEM

An IS can be defined as a set of interrelated components that collect (or retrieve), process, interpret, store, filter and distribute information to support decisions within and across partners [6].

Efficient information handling in a SCM is of the most significant importance; i.e., information system (IS) is no longer simply a support for SC operations; it is at the heart of its operations [7].

The information environment is the backbone of SCM. Information flows connect SC participants, SC functions both vertically and horizontally and management decision levels [8].

IS and new technologies could be used to improve a SCM operations. It has been identified that the use of appropriate IS could lead to the creation of differential business value [9].

In SCM area, there are widespread adoption of IS like: material requirements planning (MRP), manufacturing resource planning (MRPII), enterprise resource planning (ERP), supplier relationships management (SRM), customer relationships management (CRM) [10]. However, due to their incapability to deal with uncertainty, other systems and/or technologies are needed to improve product traceability in a SCM. As a result, advanced technologies such as RFID, global positioning system (GPS), and mobile technology have recently been applied in the SCM. They all together could to ensure improved efficiency in information processing, improved security, fast ordering, improved customer relationships, better control of supplies, etc.

SCM IS is information system used to coordinate information between internal and external customers, suppliers, distributors, and other participants in a SC [11]. It plays an increasingly critical role in the ability of firms to reduce costs and increase the responsiveness of their SC.

In the remainder of this Section, basic concepts and infrastructural framework of SCM IS are introduced.

A. Supply chain management information system: basic concepts

It must be said that the IT provide a new level of coordination capabilities in SCM and enable a breakthrough in SCM responsiveness and flexibility. IT, on one hand, serves as an environment to support SCM; while on the other hand, it is the enabler of much advancement in SCM. Modern IT can potentially enable almost any coordination concept.

In addition, there is a wealth of literature on IT in SCM; i.e., [1], [6], [7], [9], [10], [12], [13], [14], [15], [16], [17]. The comprehensive literature review from above identifies that IT is expected to have a pivotal role in SCM, now and in the future. In fact, it seems that the use of IT is crucial, particularly for managing contemporary SCs [17].

IT-based SCM systems differ from the traditional supply chain function due to computational speed and large data stores [14]. Although increased computational capacity and access to large data stores are valuable, the primary advantage of IT-based SCM systems comes from the ability to share information across all supply chain functions and, increasingly, across all SC's participants.

IT-based SCM systems deliver information to decision makers when they need it and in the format, they need it. Decision makers can create, customize, and deploy reports easily and efficiently as required. IT-based SCM systems allow monitoring of the extended supply chain, providing entire supply chain visibility.

According to [18] the *objectives of IT in SCM* are:

- Providing information availability and visibility;
- Enabling a single point of contact for data;
- Allowing decisions based on total SC information; and
- Enabling collaboration with SC partners.

IT-based SCM systems, such as SCM IS, coordinate and integrate the flow of products, information, and finances between supplier, manufacturer, wholesaler, retailer and end consumer. They serve to create a multiplicity of digital options to be applied for operational, tactical and strategic purposes. The task of such SCM IS is to collect, exchange, synthesize and analyze the life cycle information necessary to support decision making process and help in identifying and implementing features that will maximize economical effect and minimize environmental degradation. In addition, development of the SCM IS is needed in order to: Support operational decisions made in day-to-day processing; and to provide an IT infrastructure.

It is important that SCM IS has the ability to integrate not only several actor groups, but all of them across the SC. In fact, the lack of integration is a great obstacle to building an effective IS [19].

New Auto-ID technologies (like RFID, for example) enables *intra-* and *inter-organizational* communication. It is adopted for intra-organizational processes, such as inventory management, point-of-sales management, and asset management; for inter-organizational processes, such as tracking and tracing the status of item in SC [1]. For example, classical example of internal (i.e., intra- organizational) IS is ERP system. On the other hand, inter-organizational IS are used for information sharing and/or processing across organizational boundaries. There are three categories for the use of inter-organizational IS in SCM [20]:

1). Transaction processing;
2). Supply chain planning and collaboration; and
3). Order tracking and delivery coordination.

B. Supply chain management information system framework

In general, a SC logistic network comprises the *physical element* (physical SC) and the *information element* (information SC). Information part supplements the physical one, in order to achieve the efficient performance of the whole SCM. The SCM IS extends the scope and efficiency of the

SCM by providing an infrastructure to facilitate the efficient exchange of data among all SC participants (see Fig. 1).

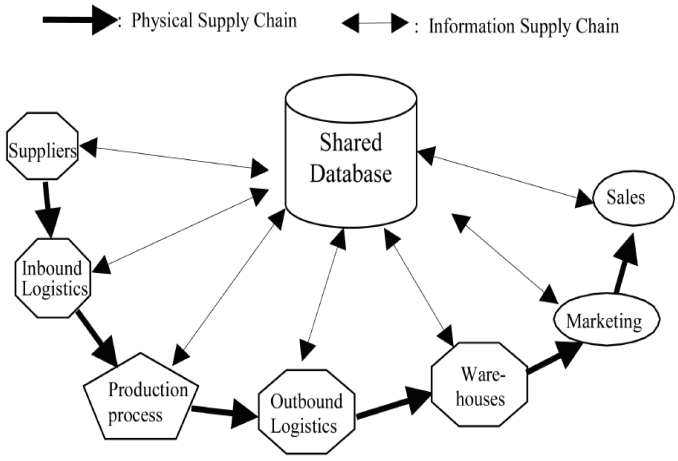


Fig. 1. Internet-based SCM [21]

All SC participants need better information flow together with easier and quicker access to required information. As SC participants begin to demand more out of the computer system to manage their everyday operations, they begin to suffer from restrictive connectivity and integration constraints. Even if a small portion of a large customer base chooses to get information over the Internet (which can include millions of inquiries per month) that would stretch the capability of the server to the limit.

The performance of the SCM IS greatly relies on its infrastructure. Studies concerned with the detailed infrastructure of a SCM IS, have not received the attention they deserve. However, there are some bright examples like [21], who elaborated on an infrastructure of integrated component-based SCM IS. The proposed SCM IS embraces the *concept of distributed object technology* to enable efficient data exchange among various data objects, which may reside in distributed platforms. In addition, the significance of this paper represents the provision of a cross-platform data exchange.

C. Supply Chain Management Information System: Infrastructural framework

Successful implementation of the SCM IS requires a distributed computing and controlling infrastructure to connect with the computer systems of all SC participants. To meet this objective, the *distributed object-oriented technology* can play an important role in terms of easy access and sharing of information across a distributed IS.

Any Internet server have to use some of distributed technologies, such as for example, the *common object request broker architecture* (CORBA) standard [22] or the *distributed component object model* (DCOM) [23]. CORBA and DCOM technologies make it possible to distribute information across virtually any number of physical servers located on a local area network (LAN) and/or wide area network (WAN), which form the infrastructure of the SCM. In addition, CORBA and DCOM-based systems avoid the single server bottlenecks that are frequent in traditional client-server systems [21]. In addition, important feature of using distributed object technology, as CORBA and DCOM are, is to ensure interoperability between applications on different machines in a heterogeneous distributed environment.

CORBA supports different computing languages and runs on different machines in heterogeneous distributed environments, and it is a well-accepted standard. There are a number of specifications and standards associated with CORBA. The core ones are as following:

- ORB is a middleware with which is able to access data storage on remote systems;
- Internet inter-ORB protocol (IIOP) is the protocol that ORB use to communicate over TCP/IP networks;
- Interface description language (IDL) is used to specify the interface between the client ORB and the server ORB; and
- Business objects, defined as high-level representations of things that exist in a business domain [22].

On the other hand, Microsoft has released its *DCOM* architecture. In brief, DCOM is an architecture that enables components (processes) to communicate across a network in a distributed way. DCOM provides distributed messaging services, distributed transaction services, data connectivity services, etc.

We preferred DCOM-based infrastructure of the SCM IS (see Fig. 2), because of two reasons: its acceptance is growing due to the wide Windows customer base; and DCOM is comparatively cheaper than other similar products (it arrives totally ”free” with Windows NT program package).

The SC participants exchange information with the ”main actor” (for example, before point-of-sale paramount participant (or main actor) is producer) based on two approaches. The first is through the DCOM directly; and the second is through Internet browser with HTTP protocol.

Business object is regarded as the organized data about certain SCM activity, which is encapsulated in the form of an entity called business object [21]. With the help of object technology, all kinds of SCM activities can be represented as business objects and incorporated into the SCM IS.

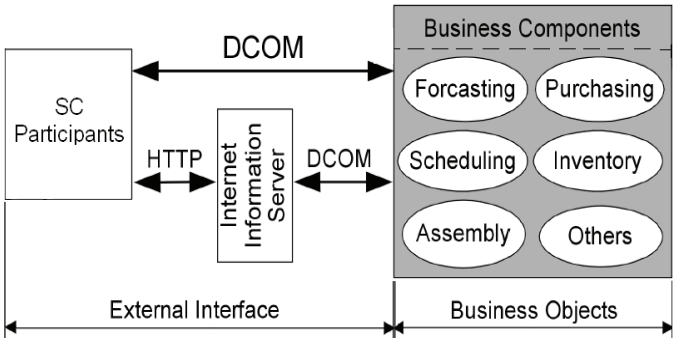


Fig. 2. DCOM-based infrastructure of the SCM IS

IV. CONCLUSION

IT-based SCM systems, like SCM IS, deliver information to decision makers when they need it and in the format, they need it. SCM IS should be regarded as an essential ingredient to provide SC participants with the ability to see, know, anticipate, model, link, and trade off available resources.

The SCM IS is aimed at making use of an efficient information flow mechanism for coordinating and monitoring the data interchange among various SC participants. SCM IS is characterized by its ability to provide accurate and relevant information to enhance the performance of the SCM processes. In addition, it can be regarded as a fully integrated and distribution-based IS.

The performance of the SCM IS greatly relies on its infrastructure. The proposed SCM IS embraces the concept of distributed object technology to enable efficient data exchange among various SC participants, which may reside in distributed platforms.

We preferred DCOM-based infrastructure of the SCM IS, because of two reasons: its acceptance is growing due to the wide Windows customer base; and DCOM is comparatively cheaper than other similar products.

REFERENCES

[1] K. Lai, C.W.Y. Wong and T.C.E. Cheng, "Institutional isomorphism and the adoption of information technology for supply chain management", Computers in Industry, vol. 57, no. 1, pp. 93–98, 2006.

[2] D. Simchi-Levi, P. Kaminsky and E. Simchi-Levi, Managing the Supply Chain: The Definitive Guide for the Business Professional, McGraw-Hill, USA, pp. 1-16, 2004.

[3] CSCMP and Supply Chain Visions. Supply Chain Management Process Standards: Enable Processes, Council of Supply Chain Management Professionals: Oak Brook, IL, 2004.

[4] J.L. Sutherland, "Logistics from a Historical Perspective", in Logistics Engineering Handbook, (Editor: M.G. Gaukler), pp. 4, Taylor & Francis Group, Boca Raton, Florida, 2008.

[5] K.C. Tan, "A framework of supply chain management literature", European Journal of Purchasing & Supply Management, vol. 7, no. 1, pp. 39-48, 2001.

[6] N.R. Sanders, R. Premus, "IT applications in supply chain organizations: A link between competitive priorities and organizational benefits", Journal of Business Logistics, vol. 23, no. 1, pp. 65-83, 2002.

[7] J.V. Pereira, "The new supply chain's frontier: Information management", International Journal of Information Management, vol. 29, pp. 372–379, 2009.

[8] D. Ivanov, B. Sokolov, Adaptive Supply Chain Management, Springer-Verlag, London, pp. 45-57, 2010.

[9] A. Radhakrishnan, X. Zu and V. Grover, "A process-oriented perspective on differential business value creation by information technology: an empirical investigation", Omega, vol. 36, pp. 1105–1125, 2008.

[10] P.H. Ketikidis, S.C.L. Kohc, N. Dimitriadisa, A. Gunasekarand and M. Kehajova, "The use of information systems for logistics and supply chain management in South East Europe: Current status and future direction", Omega, vol. 36, pp. 592–599, 2008.

[11] T.S. McLaren, M.M. Head and Y. Yuan, "Supply chain management information systems capabilities. An exploratory study of electronics manufacturers", Information Systems and e-Business Management, vol. 2, pp. 207–222, 2004.

[12] A. Gunasekaran, E.W.T. Ngai, "Information systems in supply chain integration and management", European Journal of Operational Research, vol. 159, pp. 269-295, 2004.

[13] C. Chandra, J. Grabis, "Information technology support for integrated supply chain modeling", Human Systems Management, vol. 27, no. 1, pp. 3-13, 2008.

[14] B. Dehning, V.J. Richardson and R.W. Zmud, "The financial performance effects of IT-based supply chain management systems in manufacturing firms". Journal of Operations Management, vol. 25, pp. 806–824, 2007.

[15] E.A. Williamson, D.K. Harrison and M. Jordan, "Information systems development within supply chain management", International Journal of Information Management, vol. 24, pp. 375-385, 2004.

[16] P.K. Humphreys, M.K. Lai and D. Sculli, "An inter-organizational information system for supply chain management", International Journal of Production Economics, vol. 70, pp. 245-255, 2001.

[17] J. Auramo, J. Kauremaa and K. Tanskanen, "Benefits of IT in supply chain management: an explorative study of progressive companies", International Journal of Physical Distribution & Logistics Management, vol. 35, no. 2, pp. 82-100, 2005.

[18] D. Simchi-Levi, P. Kaminsky, and E. Simchi-Levi, Designing and Managing the Supply Chain: concepts, strategies, and case studies, McGraw-Hill, New York, NY, pp. 247, 2003.

[19] S.R. Croom, "The impact of e-business on supply chain management: An empirical study of key developments", International Journal of Operations & Production Management, vol. 25, no. 1, pp. 55–73, 2005.

[20] M. Kärkkäinen, S. Laukkanen and S. Sarpola, "Roles of interfirm information systems in supply chain management", International Journal of Physical Distribution & Logistics Management, vol. 37, no. 4, pp. 264-286, 2007.

[21] H.C.W. Lau, W.B. Lee, "On a responsive supply chain information system". International Journal of Physical Distribution and Logistics Management, vol. 30, no. 7/8, pp. 598-610, 2000.

[22] Object Management Group, "The Common Object Request Broker: Architecture and Specification". Available from: http://www.omg.org/ [accessed 12.04.2010].

[23] N. Brown, C. Kindel, "Distributed Component Object Model Protocol ± DCOM/1.0", Microsoft Corporation, Network Working Group, 1996. Available from: http://tools.ietf.org/pdf/draft-brown-dcom-v1-spec-00.pdf [accessed 12.04.2010].

Innovation in Information Management Using RFID Technology Throughout Product Life Cycle

Vladimir D. Simić[1] and Branka S. Dimitrijević[2]

***Abstract* – Using radio frequency identification (RFID) in the whole product life cycle is becoming increasingly important. Such innovation enables us to track products in production processes, automatically identify, record, transmit and search product information throughout its life cycle. Implementing RFID systems in the entire product life cycle model leads to the great improvement of decision support system (DSS), since it makes accurate, real-time, and complete product information easily available. Besides all, RFID-based Information Management Framework of product life cycle is introduced.**

***Keywords* – Information management framework, Product life cycle, RFID.**

I. INTRODUCTION

Using RFID technology in the whole product life cycle is becoming increasingly important. RFID is a technology that allows remote interrogation of parts using radio waves [1]. This has several advantages, since many tagged parts of a single product could be simultaneously identified in an automated manner.

The primary task of information management is to collect, exchange and analyze the product life cycle information.

With the high complexity of supply chain (SC), the need for accurate and detailed information is becoming essential. Therefore, three-dimensional model of the information needed for successful managing of SC operations is described.

Finally, RFID-based Information Management Framework of product life cycle is introduced.

The remainder of the paper is structured as follows: Section II comprises a brief description of the RFID technology. Section III is dealing with the product life cycle issue. Three-dimensional SC information model is described in Section IV, while the basic concepts of current product life cycle management systems are given in Section V. RFID-based Information Management Framework of product life cycle is introduced in Section VI. Finally, concluding remarks are given in Section VII.

[1]Vladimir D. Simić is with the University of Belgrade, Faculty of Transport and Traffic Engineering, Vojvode Stepe 305, 11000 Belgrade, Serbia, E-mail: vsima@sf.bg.ac.rs

[2]Branka S. Dimitrijević is with the University of Belgrade, Faculty of Transport and Traffic Engineering, Vojvode Stepe 305, 11000 Belgrade, Serbia, E-mail: brankad@sf.bg.ac.rs

II. RADIO FREQUENCY IDENTIFICATION BASICS

Radio frequency identification is an automatic, wireless identification technology capable of gathering data without human intervention and line of sight requirement. RFID increases efficiency and enables new fields of application by introducing the following abilities:

- To read multiple product identities simultaneously;
- To read identities from distance without having a line of sight;
- To provide the real time information service;
- Processing tags with rewritable and extendable memory features.

RFID systems use radio waves to capture data from tags. They have three *primary components* [2]:

- The tag or transponder;
- The reader; and
- The host computer.

RFID tags are the chips embedded in the product. They are made of a hard copper coil consisting of an integrated circuit (IC) attached to an antenna, then packaged into a housing device appropriate for the application. RFID tags are very rugged and come in several forms and sizes; e.g., some can be as small as a grain of rice. Data is stored in the IC and transmitted through the antenna to a reader. RFID tags can be "*passive*" (no battery) or "*active*" (self-powered by a battery). Passive tags are more popular, less expensive, with a virtually unlimited life span. They do not have their own power supply but use the radiated energy from RFID readers to transmit information.

RFID reader is radio frequency transmitter and receiver controlled by a microprocessor that communicates with the tags. In passive systems, readers transmit an energy field that "wakes up" the tag and provides the power for the tag to operate.

Next, great volume of data collected from tags is wirelessly passed to host computer for further processing.

III. A DETAILLED VIEW ON PRODUCT LIFE CYCLE

The model of product life cycle is characterized by the three phases (see Fig. 1):

1. *Beginning-of-Life* (BOL) phase, including design and production;
2. *Middle-of-Life* (MOL) phase, including use and maintenance; and

3. *End-of-Life* (EOL) phase, characterized by various product-processing scenarios such as [3]:

• *Reuse* - any operation by which components of EOL products are used for the same purpose for which they were conceived;

• *Recycling* - the reprocessing in a production process of the waste materials for the original purpose or for other purposes;

• *Remanufacturing* - the process of disassembly of products during which time parts are cleaned, repaired or replaced and then reassembled to sound working condition; and

• *Disposal* - the process of getting rid of EOL products, such as landfill, incineration, etc.

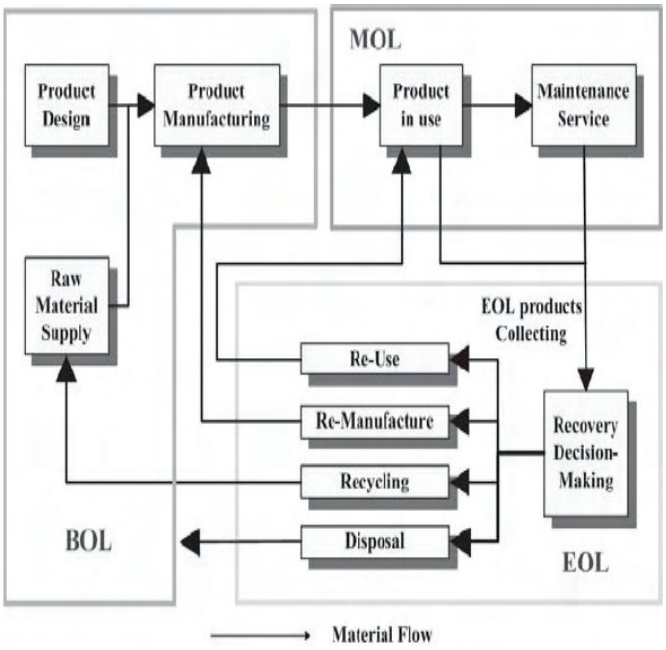


Fig. 1. The Model of product life cycle: SC point of view

Incomplete information on the product makes EOL decision-making process very complicated. Take hazardous materials as an example; there is no sufficient information about specifications and locations of such materials in the recycling processes.

IV. THREE-DIMENSIONAL SC INFORMATION MODEL

The task of SC information management is to collect, exchange and analyze the life cycle information necessary to support decision making process and help in identifying and implementing features that will maximize economical effect and minimize environmental degradation. In fact, with the high complexity of SC, the need for accurate and detailed information is becoming increasingly essential.

Three-dimensional model of the information needed for successful managing of SC operations is shown in Fig. 2. Firstly, we can classify the information into three categories [4]:

1. Economical. It is crucial for measuring the profitability of SC activities;

2. Product & Production. It is vital for the efficiency of SC activities and is referring to product technical data (e.g., material content); and

3. Environmental. It is important for measuring the physical impacts (e.g., energy use, waste generated).

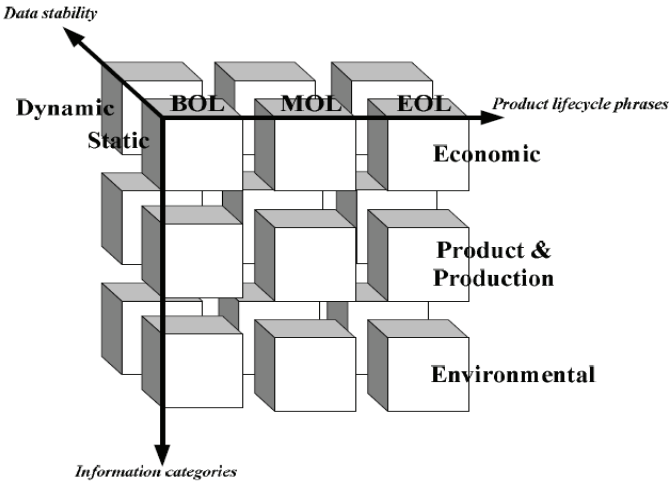


Fig. 2. 3D SC information model

Next, the information state can be classed into static (refers to the information that never changes) and dynamic (refers to the information that is always changing during life cycle). Finally, the information is distinguished into three phrases of product life cycle: BOL, MOL and EOL phase. BOL data is usually static and can be gathered from producers. MOL data is highly dynamic and keeps growing over time. It captures information during the product use and maintenance. The source of EOL data could be certified collectors, recycling facilities and/or eventual disposer.

A. Information flow within the EOL phase

With the increasing pressure on producers to manage their product processing operations, the availability of information to improve processing decisions is becoming crucial [5]. In fact, a fundamental obstacle in making efficient decisions is the loss of information associated with the product after the point-of-sale.

The information required for efficient product processing can be classified into the following six categories:

1. *Product* information. Examples of such information are product identification and product dismantling scheme. Product related information is *static* in nature (i.e., the information does not change over the product life cycle);

2. *Location* information, relates to the specific location of every part and its quantity;

3. *Utilization* information, relates to the use of the product over its entire life cycle (e.g., amount of usage). This information, which is *dynamic* in nature, is required to assess the quality of the product and its potential value. Automatic identification technology, like RFID, is making dynamic data

collection possible, where product identification is build into a product, giving it a unique "footprint";

4) *Legislative* information;

5) *Market* information. Finding markets for used products can be very difficult. Product processing critical objective is to receive the highest possible value for every potentially processed product. In order to achieve this objective, all participants in backward section of SC must have timely and accurate access to market information, concerning product demand, price and availability. E-marketing is helped to increase sales of used parts and introduced an important element of stability to the secondary part market. Maybe, the best example of successful E-marketing is American Recycling Association case [6]. Fifteen years ago, the American Automotive Recycling Association established an Internet based marketing system called "The International Database" which aims to increase the availability and use of recycled automotive parts. Today this database is the largest source of information about recycled vehicle parts in North America; and

6) *Process* information, generated within EOL processing activity. The two main categories of information generated "internally" are [7]: *storage* information (e.g., the availability, quality, stock currently held) and *sales* information.

V. Contemporary Information Management Systems

In order to make effective processing decisions, it is critical to have access to information generated throughout the entire product life cycle. In this section, we will briefly describe the some approaches for managing product information throughout its life cycle, and their pros and cons.

Information Management Systems can be divided into three categories [8]:

- *Design/dismantle data sharing system*;
- *Life cycle data management system*; and
- *RFID-based approach* for managing product information.

Design/dismantle data sharing systems are designed to allow producers to share design and dismantle-related information with certified collecting facilities, dismantlers and recyclers. These systems are only able to capture and share the static information. Evidently, they are not capable of providing detailed product life cycle information. They have the following characteristics in common:

- Manufacturers provide design data and disassembly instructions to product processing facilities; and
- Every authorized SC actor is able to access part information through the Internet.

Life cycle data management system [9] is designed to store and share static and dynamic information. However, the major drawback of this approach is that it fails to capture the dynamic nature of the static information. This results in the inability of providing accurate information about the part status at the EOL phase. In general, this system displays the following characteristics:

- Enable unique identification of every part;
- Provide design data and disassembly instructions;
- Monitor and store important product parameters throughout its life cycle; and
- Product processing facilities are provided access to the information stored with the part and linked to the static information through the Internet.

A recent breakthrough in enabling affordable widespread global deployment of RFID is the emergence of approaches connecting a product tagged with an RFID to a network. The key features of a networked RFID approach for managing product life cycle information are [8]:

- Unique identification;
- Wireless communication between item and reader;
- Networked reader; and
- Common database access approach.

VI. Innovation in Information Management using RFID Technology throughout Product Life Cycle

Implementing RFID systems in the entire product life cycle model (see Fig. 3) enables us to track products in production processes, automatically identify, record, transmit and search product information throughout its whole life cycle.

As showed in Fig. 3, the BOL material flow begins with linking *Raw material supplier* with *Product manufacturing department*. This is paralleled with RFID information flow comprised from design information sent from *Product design* to *Product manufacturing department*. Afterwards, the two combined flows are distributed to different *Customers* in *Product in use*, where the product reaches its MOL phase. Relevant product information is also processed and transmitted to an information management network named *Product Life Cycle Information Systems* (PLC IS). PLC IS contains *BOL Information Systems* (BOL IS), *MOL Information Systems* (MOL IS) and *EOL Information Systems* (EOL IS). It provides the immediate, automatic identification and sharing of item-level information in product life cycle management.

The information flow consist all crucial product information during the whole product life cycle, i.e.:

- Design information;
- Production information;
- Usage information (e.g. data collected from sensors and recorded in RFID tags);
- Maintenance information; and
- EOL instructions and eventually processing record.

Implementing RFID systems in the entire product life cycle model leads to the innovation of decision support, since it makes accurate, real-time, and complete product information available throughout the whole product life cycle. For example, the information needed in EOL phase can be retrieved from the PLC IS according to the unique ID stored in the RFID tag.

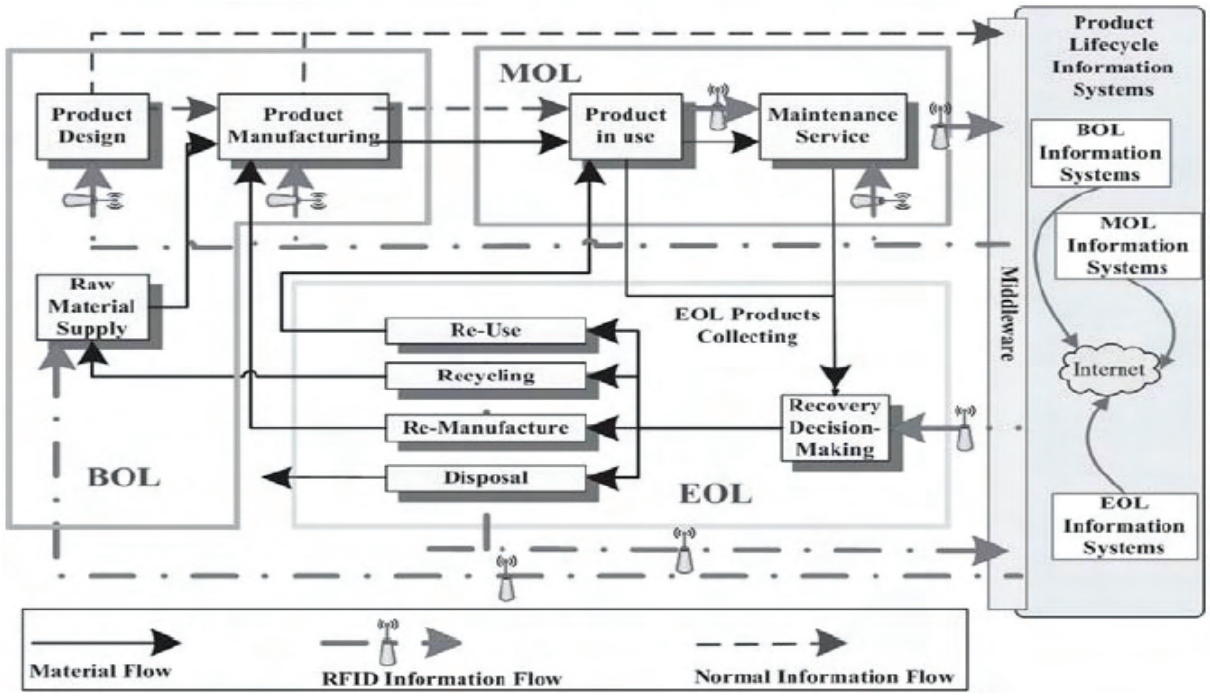


Fig. 3. RFID-based Information Management Framework in product life cycle

VII. CONCLUSION

In this paper, we give an insight into how advanced identification technology, such as RFID, could provide the information necessary to improve the quality of decisions made during product processing activities.

It is concluded that RFID is securing the ability to extract product information in a timely manner. In addition, RFID brings following twofold benefits: *processes improvement*, that upgrading the efficiency and cost effectiveness of SC operations; and *decision improvement*, that leads to higher profit.

From the long-term point of view, the "full" RFID scenario, i.e., implementing RFID in the entire product life cycle should become a reality. The realization of such a concept would enable the complete transparency of one product, the state of its containing parts and would provide direct access to stored information.

REFERENCES

[1] A.K. Parlikad, D.C. McFarlane and A.G. Kulkarni, "Improving product recovery decisions through product information", in *Innovation in Life Cycle Engineering and Sustainable Development*, (Editors: D. Brissaud et al.), pp. 153-172, Springer-Verlag, Netherlands, 2006.

[2] M. Attaran, "RFID: an enabler of supply chain operations", Supply Chain Management: An International Journal, vol. 12, no. 4, pp. 249-257, 2007.

[3] European Union: Directive 2000/53EC of the European Parliament and of the Council of 18 September 2000 on end-of-life vehicles, Official Journal of the European Union L 269, pp. 34-42.

[4] Z. Zhang, "Supporting End-of-life Product Recovery Processes in Closed Loop Supply Chains", Proceedings of the Second International Symposium on Intelligent Information Technology Application, pp. 804-808, Shanghai, China, December 21–22, 2008.

[5] V. Simić, B. Dimitrijević and B. Ratković, "Closed-Loop Supply Chain Of End-of-life Vehicles", Proceedings of the XIX Triennial Conference on Material Handling, Constructions and Logistics - MHCL09, pp. 189-194, Belgrade, Serbia, 15-16 October, 2009.

[6] American based Automotive Recycling Association. Available from: http://www.autorecyc.org/ [accessed 10.03.2010].

[7] N. Ferguson, J. Browne, "Issues in end-of-life product recovery and reverse logistics", Production Planning & Control, vol. 12, no. 5, pp. 534-547, 2001.

[8] A.K. Parlikad, McFarlane, D.C. "RFID-based product information in end-of-life decision making", Control Engineering Practice, vol.15, pp. 1348-1363, 2007.

[9] M. Simon, G. Bee, P. Moore, J. Pu and C. Xie, "Modelling of the life cycle of products with data acquisition features", Computers in Industry, vol. 45, pp. 111–122, 2001.

Ginis MvcGen – a Prototype of WebGIS Application Source Code Generator

Miloš Bogdanović[1], Aleksandar Stanimirović[2], Leonid Stoimenov[3]

***Abstract* –In this paper we will present our prototype WebGIS application source code generator called Ginis MvcGen. It can be used to create a WebGIS applications based on the usage of Model-View-Controller design pattern. This tool can contribute to faster development of high quality WebGIS software solutions. Ginis MvcGen can be easily used for creating WebGIS applications that integrate the framework for the development of WebGIS client with information obtained from heterogeneous information sources. It allows coupling of information retrieved from heterogeneous sources with map display of objects connected to obtained information using the Ginis MvcGen built-in JavaScript editor.**

***Keywords* – Web, GIS, source code generator**

I. INTRODUCTION

Development and usage of Web applications is not a novelty in the field of software engineering. Since the appearance of Web technology, a substantial growth has been registered worldwide in the use of this technology. Methodology of Web application development has changed in parallel with the development of Internet technologies. From application based on static content and design, followed by the usage of languages that enable dynamic page content and layout creation, Web applications have reached the level of collaborate systems in which the role of user is similar to application designer role. Due to its operating environment, Web applications have the ability to access data from heterogeneous sources of information. These applications can combine (integrate) obtained information. Therefore, they can be observed as access points to the different content that is displayed at user request.

The influence of Web-based solutions is also present in the field of Geo-Information Systems (GIS). Web-based GIS solutions are often referred to as Web Geo-Information Systems (WebGIS) [1]. WebGIS solutions represent the evolution of standard GIS system as a result of changing the basic operating and development environment. These solutions implement a standard set of GIS functionality, and make a step forward in terms of opportunities offered by the new operating environment. One of the basic advantages of using WebGIS solutions is the possibility of coupling various types of information with spatial context. Therefore, WebGIS solutions can be treated as access point for integration of information from distributed and heterogeneous sources.

The time devoted to the implementation of GIS systems has led to the emergence of a series of framework used for the development of different types of GIS systems (desktop, Web, mobile) [2][3][4][5]. The advantage that these solutions offer GIS system developers is effective implementation of the future solution's basic structure. The development of specific functionalities, usually information integration components, requires further implementation. Possible decrease of time spent in application development could be achieved through usage of frameworks that support information integration [6]. In order to be able to integrate retrieved data, GIS application frameworks use descriptions of information sources. Information source description is usually in the form of XML schema that provides information source access methods and characteristics of obtained information. Another possibility is the usage of semantic description of information sources in the form of intermediate semantic description layer created using special intermediate XML languages.

Although the existence of a framework significantly reduces the time required for system development, in most cases, usage of frameworks requires expert domain knowledge for development of basic application structure. The time that GIS system development experts initially spend could be further reduced if developers were using GIS system source code generator [7]. WebGIS source code generators produce most of the source code needed for application proper functioning.

This paper presents Ginis MvcGen prototype application. This tool is intended to contribute to faster development of high quality WebGIS software solutions. Generated solutions are often used as portals to spatial and geoinformations [8][9]. They can be also used as front-ends of spatial data infrastructures (SDI) [10].

II. SOURCE CODE GENERATOR ARCHITECTURE

Ginis MvcGen is our prototype WebGIS application source code generator (Fig. 1.). It can be used to create a WebGIS applications based on the usage of MVC (Model-View-Controller) design pattern. In addition to the basic GIS functionality, Ginis MvcGen generates source code required for the acquisition and visualization of information from heterogeneous information sources. Obtained information can

[1]Miloš D. Bogdanović is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: milos.bogdanovic@elfak.ni.ac.rs

[2]Aleksandar S. Stanimirović is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail:aleksandar.stanimirovic@elfak.ni.ac.rs

[3]Leonid V. Stoimenov is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: leonid.stoimenov@elfak.ni.ac.rs

be integrated and linked to specific functions of the generated application using the Ginis MvcGen built-in JavaScript editor.

Ginis MvcGen generates source code files needed for the development of WebGIS applications using ASP.NET MVC framework [11]. Generated files include source code files needed for application to be able to communicate with different information sources (RDBMS, SOAP services, REST services) and source code files for visualization of retrieved information. Each of the information sources is described through corresponding XML schema. XML schemas that describe the individual information sources contain data necessary for communication between generated application and described information sources. Source code generator performs parsing of XML schema that describes information source.

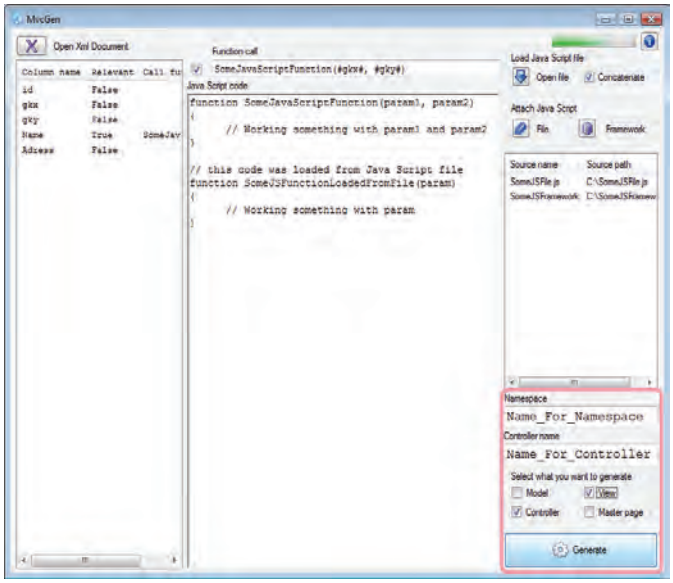


Fig. 1. Ginis MvcGen – WebGIS application source code generator interface

XML schema parsing provides generator with the description of communication methods used for retrieval of data from information source. Furthermore, it provides generator with the descriptions of data sets that will be retrieved from the information source. Ginis MvcGen provides users with opportunity to create custom JavaScript code and/or use existing JavaScript files. These files will be integrated into the generated application.

Descriptions of information sources in the form of XML schemas represent the starting point in the process of WebGIS application source code generation. This can be observed in Fig. 2. which represents the Ginis MvcGen source code generator architecture. Ginis MvcGen contains IO module which is responsible for obtaining the XML schemas of information sources. Information source descriptions can be obtained via the Internet or they may be obtained locally. It is necessary to transform obtained description into classes that represented the object representation of information sources characteristics. For this purpose, Ginis MvcGen uses a collection of utility classes. Utility classes are divided into two major groups: the classes that describe the types of data groups that can be obtained from the information source and classes that describe methods of access to particular information source. Instantiation of utility classes is performed by information source schema parser. This module receives information source description from IO module and performs its parsing. For each of the information sources, information source schema parser instantiates one object of utility class that describes the access method for a particular information source. Also, this module instantiates appropriate number of utility classes that describe data group types according to a number of different data groups obtained from the information sources.

In addition to the visualization of retrieved information, Ginis MvcGen allows the creation of WebGIS applications, which will integrate retrieved information, using one of the WebGIS client development frameworks. In the current development stage, Ginis MvcGen uses OpenLayers framework [5] for the implementation of the basic GIS functionalities. In order to generate an application that effectively integrates retrieved information with OpenLayers framework, Ginis MvcGen generator users have the ability to create custom JavaScript functions using JavaScript Editor module. The usage of JavaScript Editor module provides users with opportunity to invoke functions that belong to OpenLayers framework and combine them with information obtained from particular information source. Data obtained from specific information source can be transmitted into the custom function in the form of custom function input parameters. Groups of data that can be used for integration with OpenLayers framework are specially marked inside the XML schema of information source. If the information integration demands the usage of existing function defined in separate files, Ginis MvcGen offers the possibility of including these files into the generated application.

Generation of files that contain source code of generated application is implemented within the Code Generator Engine module. Code Generator Engine module performs integration of Information Source Schema Parser services and WebGIS framework Loader module services with custom functions created by Ginis MvcGen users. Code Generator Engine module contains static functions that can be divided into 4 groups according to their user interface capabilities and MVC component that they belong to:

1. Model generation methods
2. View generation methods
3. Master page generation methods
4. Controller generation methods

Code Generator Engine can use an arbitrary subset of static functions depending on the components that are expected as output (red frame in Fig. 1). Components expected as output are defined by Ginis MvcGen users through tools' interface. They can be combined with previously generated components simply by adding them into existing solution.

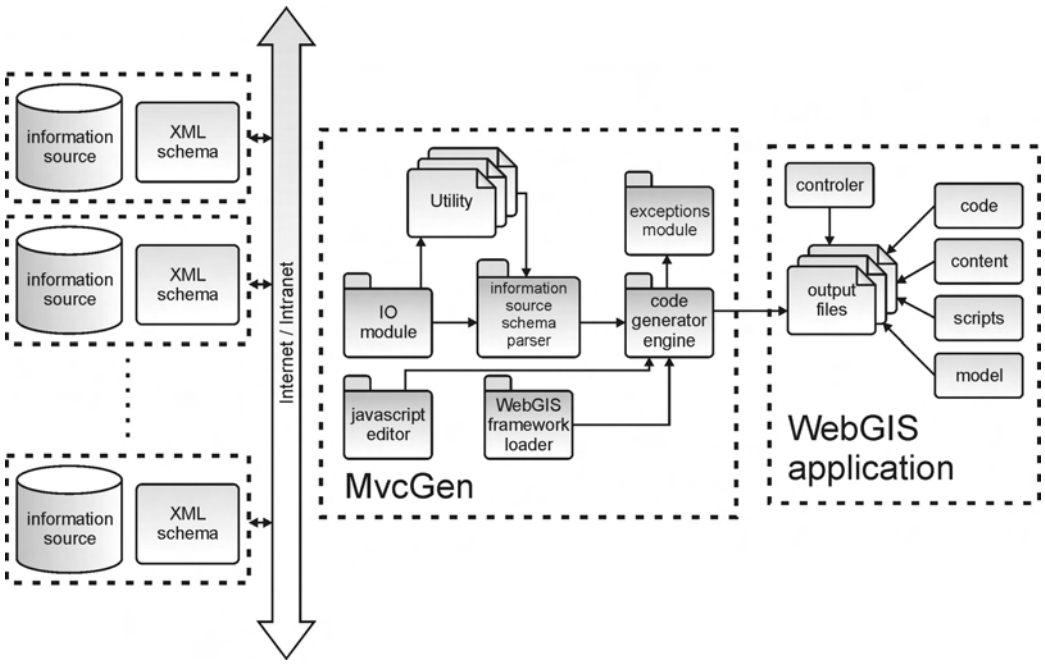


Fig. 2. Ginis MvcGen architecture

III. SOURCE CODE GENERATION PROCESS

WebGIS application source code generation process is initiated through loading of the description of one or more information sources as shown in Fig. 3.

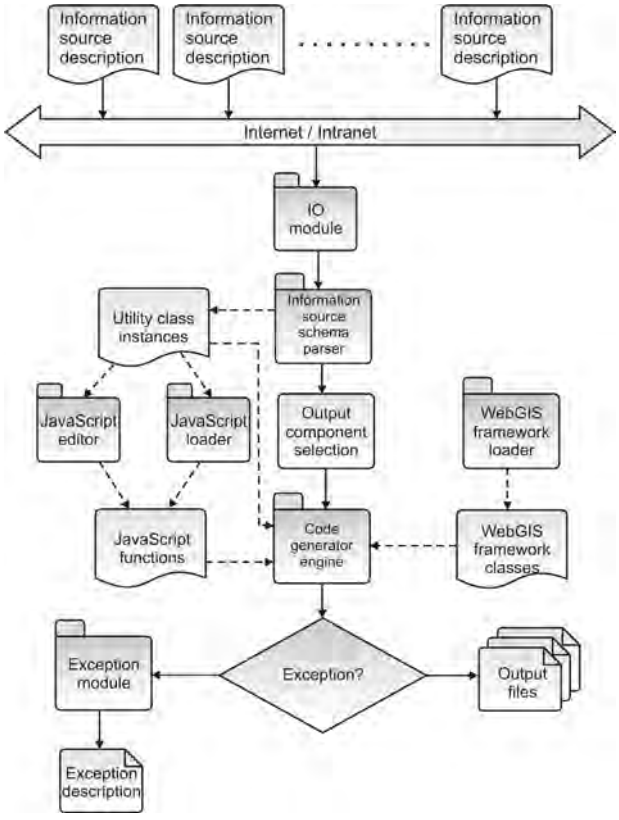


Fig. 3. Source code generation process

Information obtained from individual information sources will be integrated with the WebGIS client development framework. After selecting appropriate information source, Ginis MvcGen IO module obtains its XML schema description. IO module performs internal transformation of XML schema into two groups of utility classes: classes that describe method used for communication with the information source and classes that describe the data that can be retrieved from information sources. Generated utility class represents the input to the Information Source Schema Parser module. On the basis of information source description, an Information Source Schema Parser instantiates utility class e.g. creates an object representation of information source XML schema description. This fulfils all the prerequisites necessary for Ginis MvcGen user to determine how the integration of obtained information with WebGIS client development framework will be performed.

After parsing the description of information source, the integration of acquired data and WebGIS client development framework can be carried out. Ginis MvcGen gives users the ability to create custom functions using JavaScript language. As previously stated, inside the information source description it is possible to determine the group of data that can be used as input parameters of custom JavaScript functions. These functions integrate retrieved data with WebGIS client development framework. Based on the input parameters of custom JavaScript functions, it is possible to generate source code that invokes procedures from WebGIS client development framework.

Thus, the behavior of application that will be generated is conditioned by the information obtained, e.g. the integration of information from a geographic and non-geographic domain is performed. The development of custom functions is performed using the JavaScript editor that is embedded in

Ginis MvcGen. If Ginis MvcGen users have a pre-prepared functions, these functions can be integrated into generated application by JavaScript Loader module. Both functions created in the built-in JavaScript editor and functions loaded by the JavaScript Loader module are passed to Code Generator Engine module.

Code Generator Engine of Ginis MvcGen can generate a complete application or some parts of the application. Components selection is based on user preferences. As input, Code Generator Engine module receives the following components: WebGIS client development framework libraries, custom JavaScript functions (from built-in JavaScript editor and/or JavaScript Loaded module), object representation of information sources description and a list of components to be generated. Code Generator Engine module incorporates information retrieval procedure calls into the appropriate components of generated WebGIS application source code. Also, it mounts WebGIS development framework and binds it with retrieved informations by incorporating custom JavaScript functions.

IV. CONCLUSION AND FUTURE WORK

Ginis MvcGen belongs to the group of source code generators. It represents a good starting point when developing WebGIS software solutions. With the advantages and disadvantages that it brings, this tool can contribute to faster development of high quality WebGIS software solutions. Ginis MvcGen prototype source code generator can be easily used for making WebGIS applications that integrate the framework for the development of WebGIS client with information obtained from heterogeneous information sources. For the purposes of information integration Ginis MvcGen uses specially developed XML schemes for the description of information sources. It allows coupling of information retrieved from heterogeneous sources with map display of objects connected to obtained information. Ginis MvcGen requires user input in the form of functions that perform the integration of information e.g. connecting information obtained from a single information source with the framework for the development of WebGIS clients.

The downside of using Ginis MvcGen source code generator is reflected in the fact that it implies that users understand the semantics of information obtained from individual information sources. If this is not the case, Ginis MvcGen source code generator usage would be very difficult. For this reason, it is necessary to extend the architecture Ginis MvcGen. Instead of the previously used XML scheme for the description of information sources, a semantic description of information sources should be used. Such a description would be a priori known and easier to understand to Ginis MvcGen users. Namely, it would be enough to once get familiar with the characteristics of the semantics of information that can be obtained from heterogeneous information sources instead of having the obligation of knowing the characteristics of each of the information sources.

XML schema for the description of information sources that is currently being used would be replaced by the intermediate layer of mappings between groups of information that can be obtained from individual information sources and concepts that belong to semantic descriptions of information sources e.g. local ontology. For the purpose of creating intermediate mapping layer, it is possible to use existing techniques for semantic annotation of information sources. Since Ginis MvcGen uses RDBMS, SOAP and REST services as information sources, it is possible to use some of the techniques for deep Web annotation, Semantic Annotation for WSDL (SAWSDL) [12] and SA-REST [13]. In this way, users Ginis MvcGen would be given a unified view of heterogeneous information sources, and an opportunity to decide on information that needs to be integrated inside the application. The choice would be made at the concept level of local (domain) ontology.

ACKNOWLEDGEMENT

Research presented in this paper were partially funded by the Ministry of Science of the Republic of Serbia and PD Jugoistok Niš, within the project "Intelligent integration of geo-, business and technical information on the company level", ev. No 13003.

REFERENCES

[1] A. Shaig, *"An Overview of Web based Geographic Information Systems"*, SIRC, Dunedin, New Zealand, 2001

[2] MapTools, *"ka-Map"*, 2007, http://ka-map.maptools.org, last accessed April 2010

[3] MapTools, *"Chameleon"*, 2007, http://chameleon.maptools.org/index.phtml, last accessed April 2010

[4] A. Mitrović, D. Mitrović, S. Djordjević-Kajan, *"A Scalable, Object-Oriented GIS Framework"*, International Society for Photogrammetry and Remote Sensing, Workshop on New Developments in Geographic Information Systems, Milan, Italy, 1996

[5] Open Source Geospatial Foundation, *"OpenLayers"*, 2007, http://www.openlayers.org, last accessed April 2010

[6] L. Stoimenov, A. Stanimirović, S. Djordjević-Kajan, *"Realization of Component-Based GIS Application Framework"*, 7th AGILE Conference on Geographic Information Science, Heraklion, Greece, 2004

[7] N. Chalainanont, J. Sano, T. Minoura, *"Automatic Generation of Web-based GIS/Database Application"*, OSGeo Journal, Vol. 3, pp. 5-17, ISSN:1994-1897, 2007

[8] *"GEON project"*, 2002, http://www.geongrid.org/, last accessed April 2010

[9] *"NGDC Interactive Map Services"*, 2010, http://www.ngdc.noaa.gov/maps/interactivemaps.html, last accessed April 2010

[10] *"EIONET - European Environment Information and Observation Network"*, 1994, http://www.eionet.europa.eu/, last accessed April 2010

[11] S. Sanderson, *"Pro ASP.NET MVC Framework"*, ISBN: 978-1-4302-1008-5, Apress, 2009

[12] Semantic Annotation for WSDL Working Group, *"Semantic Annotations for WSDL and XML Schema"*, 2007, http://www.w3.org/2002/ws/sawsdl/, last accessed April 2010

[13] A. P. Sheth, K. Gomadam, J. Lathem, *"SA-REST: Semantically Interoperable and Easier-to-Use Services and Mashups"*, IEEE Internet Computing, vol. 11, no. 6, pp. 91-94, 2007

Bisection Method for DD Construction

Suzana Stojković, Radomir S. Stanković, Dragan Janković[1]

***Abstract* – Decision diagrams (DDs) are a data structure that allows compact representation of discrete functions. During DD construction of a decision diagram many Boolean operations over DD nodes are performed and many temporary nodes are created. To avoid repeating the same operation many times and creating identical nodes, the performed operation are stored in a compute table while all created nodes are stored in a unique node table. Because of that, DD construction is very memory intensive and reduction of the memory space used in DD construction is an often considered problem.**

We address this problem for the case when the functions to be represented by decision diagrams are specified in the PLA format. We propose a method for the construction of decision diagrams that performs recursively a partitioning of the input cube set. The efficiency of the proposed method is verified experimentally by a comparison with the classical Cube-by Cube algorithm and a recently proposed algorithm with non-recursive partitioning. Improvements in the memory management achieved by the algorithm proposed in this paper are on the average 32% to 35% compared to the classical algorithm and 18% in the case of the algorithm with non-recursive partitioning.

***Keywords* – decision diagrams, decision diagram construction, bisection method, Boolean functions representations**

I. INTRODUCTION

Decision diagrams (DDs) are a data structure used for compact representation of discrete functions. Due to their important applications in many areas of VLSI CAD, decision diagrams and programming packages for their construction and manipulation are nowadays a standard part of many related CAD systems.

Irrespective of the implementation details of a package used, during the construction of decision diagrams, many Boolean operations over DD nodes are performed with many of identical operations repeatedly performed. To avoid such situations, the basic principles for programming of DDs (defined in [1-2]) suggest using:

1. The unique node table (a hash table that contains all of the created nodes, which prevents creation of identical nodes many times), and
2. The compute table (a hash table that contains arguments and results of performed operations that prevents the performance of the same operation many times).

A consequence in practice is that a node table contains many more nodes than it is necessary to represent the function given. The compute table usually increases faster than the node table. Therefore, it may happen that for some functions it is impossible to generate a decision diagram within the allocated resources, although the size of the final DD is much smaller than the available memory space. Therefore, decreasing the number of temporary nodes and the number of performed operations in DD constructing is a very important task and has been considered in a number of publications; see, for instance, [3-9] and references therein.

In this paper, we present an improved method for construction of DD of a Boolean function that is given by in the PLA format, i.e., by cubes. In [9] it is presented an algorithm for DD construction with partitioning the input cube set. The method presented in this paper, also perform the partitioning, however, the partitioning strategy is different and consists in a recursive implementation of the partitioning procedure in many levels. Experimental results show that the presented method reduces the number of created nodes by 32.85% compared to the classical "Cube by cube" algorithm for DD construction from cubes, and for 18.51% compared to the algorithm with non-recursive partitioning the input cube set. The proposed algorithm reduces also the compute table for at about 35.9% and 17.96% for the discussed algorithms, respectively.

The paper is organized in the following way. Section 2 reviews two widely used representations of discrete functions: decision diagrams and the PLA format. In Section 3 we briefly discuss two methods for the construction of decision diagrams for functions specified by cubes: the classical "Cube by cube" method and a method with partitioning the input cube set presented in [9]. Section 5 proposes an improvement of the construction procedure by partitioning the input cube set in many levels named as the "Bisection method". Section 6 presents a set of experiments that compares the number of created nodes and the number of performed operations during DD construction in three algorithms discussed in this paper. Section 7 summarizes main features of the proposed method and presents the related conclusions about its efficiency and applicability.

II. REPRESENTATION OF SWITCHING FUNCTIONS BY CUBES AND BY DECISION DIAGRAMS

There are many ways to specify switching functions, as for example, truth-tables or truth-vectors, various functional expressions, as well as reduced representations such as cubes or decision diagrams.

A cube specifies the input assignment where a function takes the same value (1 or 0 or is unspecified). Thus, for an *n*-variable function, the cube is an *n*-bit string with entries in {0,1,-}, where 0 stands for the negated (complemented) input, 1 corresponds to the uncomplemented input, and "–" means the input is unspecified, i.e., can be either 0 or 1.

[1]Authors are with the Faculty of Electronic Engineering, A. Medvedeva 10, 18000 Niš, Serbia,

Contact author e-mail: suzana.stojkovic@elfak.ni.ac.rs

Berkeley PLA (*Espresso*) format that is used to specify two-level logic circuits, is widely exploited in many minimization and optimization algorithms as well as synthesis tools. A PLA specification consists of cubes for input assignments (the input part) and shows the corresponding output values in the output part.

Example 1. *Ful PLA specification of one Boolean function is:*

```
.i 4
.o 1
.c 4
1101 1
-110 1
-001 1
0-10 1
.e
```

.i, .o, and .c are the number of inputs, outputs, and cubes, respectively. The symbol .e denotes the end of the file.

Binary decision diagrams (BDDs) [2] are another form of reduced representations of two-level logic. A BDD is a graphical representation of the logic function *f* that is derived by the recursive application of the Shannon decomposition.

Definition 1. *Shannon decomposition*

Let f be a Boolean function and x_k be a variable in f. Then

$$f = \bar{x} f_{x_k=0} + x f_{x_k=1}$$

where $f_{x_k=0}$ and $f_{x_k=1}$ are cofactors of the functions f for x_k = 0 and x_k =1, respectively.

Formally, a BDD can be viewed as a directed acyclic graph G=(N,V,E) consisting of the set of non-terminal nodes (N), and terminal or constant nodes (leafs) (V) connected by edges (E). Non-terminal nodes are labeled by variables x_i in $f(x_1,\ldots,x_n)$, called the decision variables. Nodes to which the same variable is assigned form a level in the BDD. Among non-terminal nodes there is the root node representing the function *f*, while other non-terminal nodes represent subfunctions in *f* for different assignments of values to the subset of variables in *f*. Each non-terminal node (except the root node) has one or more input edges and two outgoing edges pointing to the cofactors of *f* for the corresponding assignments of input variables at the upper levels. The outgoing edges are labeled by the logic values of 0 and 1, which a decision variable can take.

Example 2. *BDD of the function f that is specified by the PLA format in Example 1 is shown in the Fig. 1.*

III. EXISTING ALGORITHMS FOR DD CONSTRUCTION FROM CUBES

The classical algorithm for DD construction is:

Algorithm 1. *"Cube by cube" for DD construction*

Given a function f by the cube set $S = \{c_1, c_2, \ldots, c_m\}$.

Step 1. For the first cube c_1 in S construct the DD (called masterDD).

Step 2. For other cubes c_i ($i \in [2,m]$) in S do

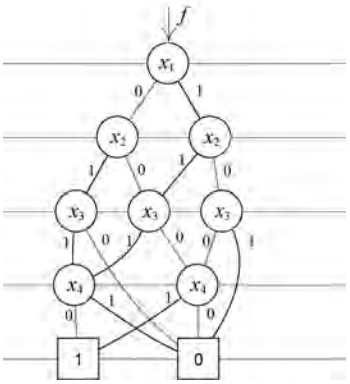


Fig. 1. BDD of the function from Example 1.

Step 2.1. Construct the DD representing the cube c_i and denote it as the temporaryDD.

Step 2.2. Add the temporaryDD to the masterDD representing all previously processed cubes by performing the operation OR (masterDD = masterDD <OR> temporaryDD).

As it is discussed above, and as it can be seen from the Algorithm 1, many operations are performed over the DD nodes during DD construction. It follows that the number of intermediately created nodes often exceeds considerably the number of nodes in the final DD. The following example illustrates disproportion between the number of nodes in the node table and the number of nodes in the current *masterDD* as a function of the number of processed cubes.

Example 3. *The diagram in Figure 2 shows the increase in the total number of created nodes (NN) and the number of nodes in the masterDD (size) as a function of the number of processed cubes in the constructing the decision diagram for the 25-variable benchmark function Apex2.*

In [9], it is proposed a method to reduce the number of nodes that are generated during DD construction by partitioning the input cube set.

Algorithm 2. *DD construction by partitioning the input cube*

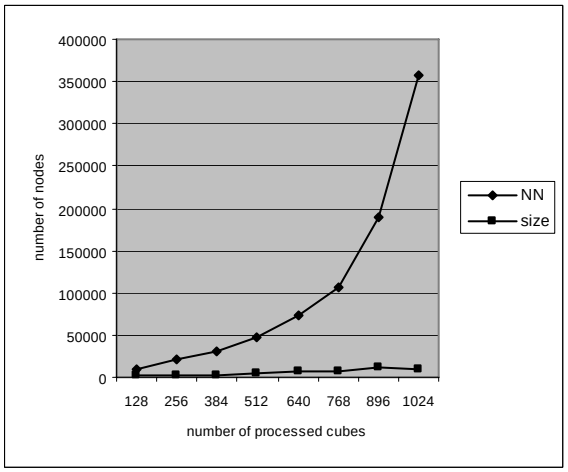


Fig. 2. The size of DD and the total number of created nodes (NN) in the generation of the decision diagram for Apex2.

set

Given a function f by the cube set $S = \{c_1, c_2, \text{K}, c_m\}$.

Partition S into partitions with size w.

Step 1. Construct the DD for the function that is defined by the first partition of the cubes by using the "Cube by cube" algorithm and name it masterDD.

Step 2. For each other partition p_j *(* $j \in [2, k]$*, k=m/w):*

Step 3.1. Construct the DD for the function defined by the partition p_j *by using the "Cube by cube" algorithm. Denote this DD as temporaryDD.*

Step 3.2. Add temporaryDD, created in the step 3.1, to the masterDD (i.e., perform the operation OR over masterDD and temporaryDD).

Experimentally it is determined that the optimal value for partitions size is $w = \sqrt{m}$, where *m* is the number of cubes.

IV. "BISECTION METHOD" FOR DD CONSTRUCTION

In this paper, we propose a modification of the algorithm for DD construction by partitioning the input cube set. The modification consists in a recursive application of the partitioning procedure and we used different size of subpartitions compared to the previous algorithm. In the proposed algorithm, the input cube is first split into two partitions of the equal size and then each subsequent subpartion is recursively partitioned in the same way until partitions of the size 2 are achieved. Formally, this algorithm can be described as:

Algorithm 3. *"Bisection method" for DD construction*

Given a function f by the cube set $S = \{c_1, c_2, \text{K}, c_m\}$.

If m>2:

Step 1a. Partition S into partitions with size m/2.

Step 2a. Construct the DD for the function that is defined by the first partition of cubes form S by using "bisection method". Denote this DD as lowDD.

Step 3a. Construct the DD for the function that is defined by the second partition of cubes form S by using "bisection method". Denote this DD as highDD.

Step 4a. Create result DD by merging lowDD and highDD (masterDD = lowDD <OR> highDD).

else:

Step 1b. Create DD by using "Cube by cube" method.

The following example compares the proposed "Bisection method" for DD construction with existing methods that are presented in the previous section.

Example 4. *Figure 3 shows the number of created nodes in construction of the DD for the benchmark function Apex2 when the DD is generated by the "Cube by cube" algorithm, the algorithm with non-recursive partitioning and by the "Bisection method".*

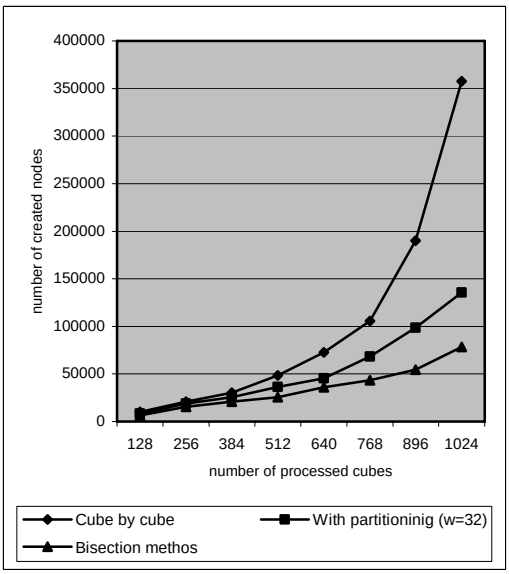


Fig. 3 The number of created nodes during generating the DD of the function Apex2 by different algorithms as a function of the number of processed nodes.

V. EXPERIMENTAL RESULTS

The proposed method is applied for the construction of DDs of some MCNC benchmark functions. Table I compares number of created nodes when the "Cube by cube" algorithm, the algorithm with non-recursive partitioning the input cube set and the "Bisection method" are used. In this table, the following labels are used:

- size – the number of the nodes in the final DD,
- N_1 – the number of created nodes in the "Cube by cube" algorithm,
- N_2 – the number of created nodes in the algorithm with non-recursive partitioning
- N_3 – the number of created nodes in the "Bisection algorithm"
- $\Delta N_{3\text{-}1}$ – the reduction ratio of the number of created nodes in the "Bisection algorithm" and the "Cube by cube" algorithm ($\Delta N_{3-1} = \frac{N_1 - N_3}{N_1} \cdot 100\%$).
- $\Delta N_{3\text{-}2}$ – the reduction ratio of the number of created nodes in the "Bisection algorithm" and the algorithm with non-recursive partitioning ($\Delta N_{3-2} = \frac{N_2 - N_3}{N_2} \cdot 100\%$).

To compare the total memory space that is used in executing each of the discussed algorithms, we also analyze the number of executed Boolean operations during the DDs constructions, i.e., the number of elements that are stored in the compute table. The results of these comparisons are shown in Table II. The meaning of labels is the same as in Table I, with the letter N for the number of nodes replaced by C for the number of computations. As it can be seen from these tables,

TABLE I
NUMBERS OF CREATED NODES WHEN DDS ARE CREATED BY DIFFERENT ALGORITHMS

Function	size	N_1	N_2	N_3	ΔN_{3-1}	ΔN_{3-2}
Alu4	1352	18036	16125	14589	19,11	9,53
Apex2	7102	372349	140671	78457	78,93	44,23
Apex1	28414	1147705	193739	56074	95,11	71,05
Apex4	1021	11308	10474	9755	13,73	6,86
Apex5	2705	16888	16165	13293	21,29	17,77
B12	91	316	312	273	13,61	12,5
Bw	118	460	340	293	36,30	13,82
Duke2	976	3262	2710	2283	30,01	15,76
Ex1010	1079	10357	10197	9446	8,80	7,36
Ex5	311	4841	2572	1882	61,12	26,83
In4	1109	8033	7004	6778	15,62	3,23
Misex3	1301	25504	25604	21003	17,65	17,97
Misex3c	810	8277	6002	5122	38,12	14,66
Pdc	696	3923	3157	2881	26,56	8,74
Spla	625	5454	5009	4734	13,20	5,49
Table3	941	4386	4147	3166	27,82	23,66
Vg2	1059	4923	3391	2878	41,54	15,13
Average					32,85	18,51

TABLE II
NUMBERS OF EXECUTED OPERATIONS WHEN DDS ARE CREATED BY DIFFERENT ALGORITHMS

Function	C_1	C_2	C_3	ΔC_{3-1}	ΔC_{3-2}
Alu4	30519	22350	17307	43,29	22,56
Apex2	647673	269503	166933	74,23	38,06
Apex1	1193198	224928	75843	93,64	66,28
Apex4	10814	10008	9350	13,54	6,57
Apex5	16582	15602	12193	26,47	21,85
B12	284	278	252	11,27	9,35
Bw	463	339	293	36,72	13,57
Duke2	3227	2637	2171	32,72	17,67
Ex1010	8563	8403	7649	10,67	8,97
Ex5	5609	3132	2347	58,16	25,06
In4	11143	9527	8894	20,18	6,64
Misex3	37733	32006	25598	32,16	20,02
Misex3c	12842	9091	7631	40,58	16,06
Pdc	3986	3152	2854	28,40	9,45
Spla	4412	3991	3785	14,21	5,16
Table3	3842	2818	2590	32,59	8,09
Vg2	6207	4038	3633	41,47	10,03
Average				35,9	17,96

the proposed “Bisection method” reduces the memory space needed for storing both the created DD nodes, and the performed computations. These two reduction ratios are approximately the same.

VI. CONCLUSION

The method for DD construction presented in this paper is based on the partitioning the input cube set recursively in multiple levels. The efficiency of the method is estimated experimentally over a set of MCNC benchmark functions. Experiments show that the presented method reduces memory space for two critical data structures in DD generation: the node table and the compute table.

As this approach does not require operations out of the set of standard operations usually implemented in decision diagram based packages, it can be used within the existing packages.

REFERENCES

[1] K. S. Brace, R. L. Rudell, and R. E. Bryant, “Efficient implementation of a BDD package” Proc. Design Automation Conference, pp. 40–45, 1990

[2] R. E. Bryant, “Graph-based algorithms for Boolean function manipulation”, IEEE Transactions on Computers, Vol. C-35, No. 8, pp. 677-691, 1986.

[3] B. Yang, Y. -A. Chen, R. E. Bryant and D. R. O'Hallaron, "Space- and Time-Efficient BDD Construction via Working set Control", Proc. Asian-South Pacific Design Automation Conference, pp. 423-432, 1998.

[4] S. Minato, "Streaming BDD Manipulation", IEEE Transactions on Computers, VOL. 51, pp. 474-485, 2002.

[5] S. Minato and S. Ishihara, "Streaming BDD Manipulation for Large-Scale Combinatorial Problems", In Proc. of ACM/IEEE Design, Automation and Test in Europe (DATE-2001), pp. 702-707, 2001.

[6] H. Ochi, N. Ishiura, S. Yajima, “Breadth-First Manipulation of SBDD of Boolean Functions for Vector Processing”, Proceedings of the Design Automation Conference, pp. 413-416, 1991.

[7] H. Ochi, N. Ishiura, S. Yajima, “Breadth-First Manipulation of Very Large Binary-Decision Diagrams”, *Proceedings of International Conference on Computer-Aided Design*, pp. 413-416, 1993.

[8] R. K. Ranjan, J. V. Sanghavi, R. K. Brayton, A. Sangiovanni-Vincentelli, “High Performance BDD Package Based on Exploiting Memory Hierarchy”, *Proceedings of ACM/IEEE Design Automation Conference*, pp. 635-640, 1996.

[9] S. Stojkovic, D. Jankovic, R. S. Stankovic, ”An Improved Algorithm for the Construction of Decision Diagrams by Rearranging and Partitioning the Input Cube Set”, accepted for publication in IEEE Transactions on Computers

Common Infrastructure for Interchange of Information in Power Supply Companies Based on Enterprise Service Bus

Leonid V. Stoimenov[1], Saša B. Tošić[2], Nikola P. Davidović[1], Nataša Ž. Veljković[1], Miloš D. Bogdanović[1], Aleksandar S. Stanimirović[1]

***Abstract* – In order to support their business processes, large electric power supply companies like ED Jugoistok Niš, use different information systems in almost every part of the company. However, these information systems are divided logically and physically and rely only on information available locally. These information systems in most cases generate only local subsets of data. The integration of information from local subsets would enable employees to perform complex data analysis related to company operations that would improve the company performance as a whole. This paper presents our solution for information integration in power supply companies. Our solution is based on Enterprise Bus Service and uses Web services for realization of interoperability between information systems.**

***Keywords* – Enterprise Service Bus, Service-Oriented Architecture, Web Services, Enterprise Application Integration**

I. INTRODUCTION

The rapid development of information and communication technology has provided an enormous amount of information to users of information systems that are available today [1]. Modern information systems used in utility companies have the ability to present information from a large number of distributed and heterogeneous information sources providing information about infrastructure, finances, employees etc. [2].

Electric power utility industry demands for efficient operations, high quality and reliable supply, low energy losses, customer satisfaction, along with expected deregulation and competition make it even more important to manage and operate the power supply network in an efficient and cost effective ways [3]. A typical Electric Utility system consists of Generation and Transmission/Distribution subsystems. The basic structure of Power Transmission/Distribution subsystem covers a huge network consisting of wide range of Equipments, Feeders & Facilities. In order to efficiently manage utility operations different disparate IT systems with different and unique roles exist. Typical systems used in the Electrical Utility are Customer Information Systems (CIS), Geographic Information System (GIS), Supervisory Control and Data Acquisition (SCADA), Distribution Management Systems (DMS), Automatic Meter Reading (AMR), Computerized Maintenance Management Systems (CMMS), Document Management System, etc. Such a large number of IT systems are intended to support business processes within Electric Utility system, such as maintenance and construction, records and asset management, network extension planning, customer care.

The main problem is that these IT systems have generally not been integrated, resulting in a complex and inefficient working environment for the users. Every IS vendor in ED Jugoistok (South Serbia Power Supply Company) developed each system only to comply with requirements imposed by particular department or set of users. In order to satisfy demands from various users, each vendor tried to cover broader set of functionalities for processing large scales of different data.

Cooperation between different information systems at ED Jugoistok is very low. Some information systems are integrated on database level which requires creation of additional stored procedures and views in database. This integration can be considered as point-to-point since information systems are using data stored in a database by other system. For example, in ED Jugoistok, this approach is used for integrating GIS with information from CIS and application for calculating technical losses is integrated with GIS, CIS and AMR.

The trend of adding new functionalities and data redundancy in order to cover ever-broader sets of users' demands [4] will ultimately result in the development of a single or several large-scale systems with complex maintenance issues. In addition, this could impose possible dependence on single vendor. In order to prevent this scenario, the process of integration of individual systems into a single entity needs to be introduced, with ability to preserve each system's independency.

Thus, there is a strong need for efficient integration of IT systems within Electrical Utility enterprise [3]. The new approach to integration would enable information sharing between information systems in a more effective way creating information infrastructure that will support migrating existing electric distribution network towards Smart Grid [5, 6].

This paper presents different approaches to enterprise application integration (EAI) in second chapter. Third chapter proposes our solution for information integration in Jugoistok.

[1]Leonid V. Stoimenov, Nikola P. Davidović, Nataša Ž. Veljković, Miloš D. Bogdanović and Aleksandar S. Stanimirović are with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mails: {leonid.stoimenov, nikola.davidovic, nataša.veljkovic, milos.bogdanovic, aleksandar.stanimirovic} @elfak.ni.ac.rs

[2] Saša B. Tošić is with the ED Jugoistok, Bulevar dr Zorana Đinđića 46a, 18000 Niš, Serbia, E-mail: sasa.tosic@jugoistok.com

Last chapter discusses benefits gained with information integration within Jugoistok.

II. Related Work

Previous requests have resulted in the need to integrate existing information systems and applications inside and outside of the Jugoistok company (intra and inter-organizational integration). Only solution that can fulfill these requirements is infrastructure for information exchange within company. This infrastructure has to provide a common model that can be used with various technologies and platforms and needs to be flexible and expandable in order to meet any future needs.

The solution for integration of heterogeneous applications of an enterprise emerged in form of EAI. EAI is defined in [7] as the unrestricted sharing of data and business processes among any connected applications and data sources in the enterprise. EAI can be accomplished using various architectures. Three most common ones are: point-to-point, broker based and message-bus based.

In point-to-point approach, applications are connected with each-other [8]. Since each application uses its own data format, in order to communicate with other applications, it needs to transform the data from local format to the format of the destination application. The total number of transformations, if the system has N applications, is N*(N-1)/2. Point-to point architecture is very suitable for small enterprises because it is quick and easy for implementation. Problem arises with increase of number of applications in the enterprise. E.g. for N=5 different applications we have to implement 6 different data transformation points. For complex system this problem can fully prevent implementation of information exchange infrastructure.

Broker-based is a better solution taking into account the number of transformations needed for the previous architecture. In the broker-based architecture, instead of connecting every application with each other, all applications are connected trough one central point called broker and use a common data model [8]. The data that an application sends needs to be transformed into a common model established between applications. In the broker-based architecture, transformations are done at the broker. In this way, the broker simplifies integration and enables the integration of a large number of systems, including existing and new ones that will be implemented later. The number of transformations is equal as the number of application. Besides the transformation, broker is also in charge of routing. Since all data that moves between the source and target applications is routed through the broker, it tends to become a bottleneck over time and a single point of failure. This often has a harsh impact on the performance of the integration solution.

Message-bus based architecture uses a central messaging backbone for routing and data propagation between applications [8]. This architecture is also known as publish/subscribe architecture. In this approach data transformations are distributed over the application adapters. Adapters perform data transformation to a common data model that all applications understand. A main advantage of this architecture is scalability, adding new and removing existing applications does not affect the system as a whole.

The rapid development of technology in the field of IS has enabled realization of platform for the integration, such that takes into account the uniqueness of the role of each system and its need for certain information, but also provides system with possibility to connect to the common data model. This concept is based on the message-bus approach, and is known as the Enterprise Service Bus (ESB) [8].

The data that an application sends to ESB needs to be transformed into a common model established between applications. After the transformation, data is forwarded to all interested applications and then re-transformed from common model to application-specific format. The role of ESB is to carry out the transformation and routing divisible on the adapters that connect the applications or systems to the bus.

In the ESB all applications are linked in a series along a common communication backbone. ESB serves as a router of data that needs to be exchanged between systems. It doesn't own the data, data is still stored in the IS that are responsible for them. It only uses their model and business logic during accessing, transforming and routing of data.

III. ESB as a Solution for Information Systems Integration in ED Jugoistok

Various information systems are used in Jugoistok. We have divided them into five logical groups based on their functionality. Systems that will be implemented in near future were also identified, considered, and classified. Logical groups are (Fig. 1):

- Systems for analysis and monitoring
- Technical information systems
- Business information system
- Systems using Data Warehouse
- Web portal and outside partners IS

Systems for analysis and monitoring are information systems used for telemetry of electric distribution network and consist of SCADA and Sensor Network. These systems generate a large amount of data about the current status of the network. Information systems that hold technical data considering network topology, objects connected to electric network and installations are classified as Technical information systems and consist of DMS, TIS, and GIS. These systems provide support for activities like repair and maintenance, planning, creating technical documentation etc. Enterprise Resource Planning, Document Management, Asset Management, and system for supporting Human Resources belong to Business Information Systems group. These systems primary support operations that are related to employees, documents, assets etc. and are outside of the scope of primary company's business activities. Data Warehouse is used by CIS and AMR systems. It adds a new value to information through analysis and data mining over already existing customer's data and consumption data. In order to facilitate easy information retrieval from various systems, Web portal with different privileges of access needs to be introduced. Also, it is necessary to provide outside partners with single access point to ED Jugoistok information infrastructure.

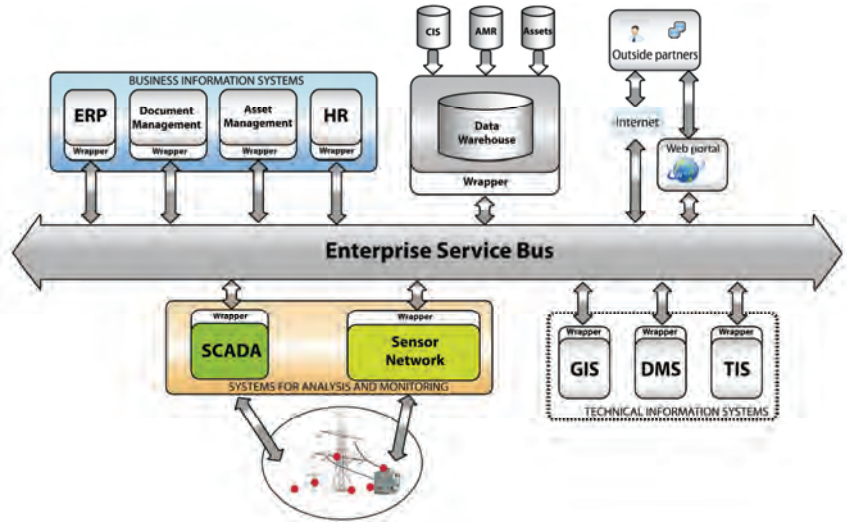


Fig. 1. Architecture of ESB in ED Jugoistok

For the purposes of integration of information systems in ED Jugoistok, in this chapter we suggest information integration architecture based on the Enterprise Service Bus (ESB). ESB provides a secure and reliable communication mechanism through the corporate network. ESB routes messages between sender and recipient. It has the ability to put messages on hold if the recipient is not available, which frees the sender's obligation of keeping and re-sending a message. Applications send messages to or receive messages from ESB through components called adapters. Adapters are application-specific interfaces to applications that provide access to other applications through the bus.

Because different applications represent data differently, transformation functions translate data from one application's data format to other, understandable to all other applications in the system [9]. After establishing communication through ESB, implementing adapters and transformation functions, and providing reliable data exchange between multiple applications, it is possible to define business processes that can use integrated information. Business process is a series of actions performed in a variety of applications.

Integration infrastructure based on ESB using special tools provides modeling, monitoring, and optimization of business processes. Common / shareable repository is used to store business rules and definitions of business objects. These common definitions allow loose coupling between applications and make the business process independent of application-specific data formats. Information related to the description of message formats with data, is called metadata.

One of the benefits gained with implementation of information integration and definition of business rules is the possibility of connecting with outside business partners. Adapters convert internal messages to Internet-compatible protocols like XML over HTTP. Additional encryption makes information traveling through Internet secure.

Information from different applications integrated using ESB can be published through unique Web Portal [8]. Web Portal integrates information from different IS, and displays them in a consistent, user-friendly way. In this scenario, ESB acts as buffer between the portal and other systems.

ESB is most frequently implemented using Service-Oriented Architecture (SOA). SOA is a paradigm that uses services as basic building blocks for developing complex software applications and systems. Services are independent, self-descriptive, and open components that support rapid and easy composition of shared applications. Within SOA, each program component can be considered as a service, usually available as Web services [9]. Web service is identified by Uniform Resource Identifier – URI, and uses open Internet standards for its description and data transfer. By implementing SOA concepts, development environment based on Web services exceed the other service-oriented paradigms through standardization and wide availability.

During application integration, it is necessary to take into account the environment in which the systems work, and besides presented logical division into groups, existing information systems are divided into real-time systems, back office systems and front office systems.

The core of the real-time systems integration is a real-time ESB, which is a proxy for the exchange of information between the various time-critical systems. In this part of the system, there are systems like SCADA and DMS. Architecture of Web services is not suitable for connecting applications that exchange considerable amount of data in real time. During the time of peak activity data is exchanged in such a frequency that the architecture based on Web services cannot meet the needs for that exchange. The problems of wrapping large amounts of data into XML format frequently brings the problems of increased data payload needed in order to communicate. Problems of parsing large XML files in the process of transforming information in formats used by other applications are also an issue. Therefore, integration is based on binary communication architectures of the middle layer (e.g., DCOM, Corba, Message Queue). In these specific cases, the application would have to be integrated into the ESB in such a way that does not compromise the basic concept of architecture based on services.

In order to achieve maximum flexibility during implementation of the adapters (interfaces) for each system it is convenient to use interface standard such as the Generic

Interface Definition - GID. An important role also plays Common Information Model (CIM) [10] database containing physical data according to the electric power system abstract data model. As such, it represents a unique point of exchange and storage of data on the model.

In the back-office segment, ESB represents the integration platform for connecting business applications that are not time-critical systems. It is based on the communication architecture of Web Services and represents the implementation of the SOA paradigm. ESB provides additional functionality such as message format transformation, routing and orchestration of services with the aim of performing various business processes. The messages exchanged are based on a common abstract model of CIM. Systems in the back-office segment are divided into two categories. One category is the service components so-called services, and other is consisted of graphical user applications, implemented on a single Web portal. Services implement the business processes' logic, and Web Portal implements a single point of access for users accessing to all applications.

Front-office segment is for users residing outside of the company. External users can access ESB via Intranet or Internet. In both cases, it is necessary to pay attention to the security of the system. Security mechanisms such as authentication, authorization, and encryption need to be implemented. In order to provide safe access, Public Key Infrastructure (PKI) should be applied.

For example, the business process that prepares the technical report needed for connecting new customers to electricity network uses information from following systems:

- GIS - provides locations of transformer stations and customers, estimation of technical losses
- AMR – load profile diagrams
- TIS - technical data considering connection points
- Web portal – user request and report generation

After the request is created on the Web portal, ESB uses repository in order to determine which systems can provide necessary information and which Web services need to be contacted. Request towards GIS's wrapper obtains the information about closest available transformer stations that are candidates for connection based on the location of the consumer. Next, it is necessary to obtain technical information considering transformer stations, feeders, and available connection sites from TIS. Load profile diagram of similar customers is added to the report obtained from AMR. Having potential consumer's location, analysis from GIS on how new connection will affect technical losses in that part of the network is added to the report. Resulting message in XML format is offered through the portal in the form of reports in the required format. Created report can be then submitted for further analysis in other departments.

IV. CONCLUSION

The proposed solution is based on the described principles of ESB, which provide significant advantages over other systems. ESB keeps up the existing functionality of the existing IS and minimize the work on their adaptation to new working environment. Any possible future expansion, as adding of the new IS, is maximally facilitated, and the only requirement that should be met is compatibility with the common data model and the ability to exchange data from a pre-defined methods.

The key advantages of ESB are:

1. Faster and cheaper adapting of the existing systems
2. More flexibility and easier changing according to the new requirements
3. Based on global and generally accepted standards
4. Scalability, i.e. easy extension of the use of the application in the whole business system compared to the initial use of application only on a single location
5. More configuration, less coding

The proposed infrastructure is a foundation for a future development of Smart Grid in ED Jugoistok. Since the implementation of the Smart Grid will affect many existing information systems in Electric Utility Companies, presented architecture will adapt information infrastructure in ED Jugoistok for future development of electric distribution infrastructure [5, 6].

ACKNOWLEDGEMENT

Research presented in this paper were partially funded by the Ministry of Science of the Republic of Serbia and ED Jugoistok Niš, within the project "Intelligent integration of geo-, business and technical information on the company level," ev. No 13003.

REFERENCES

[1] P. Johannesson, E. Perjons, "Design Principles for Application Integration", Springer Berlin / Heidelberg, pp.212-231 January 2000.

[2] L. Stoimenov, S. Đorđević-Kajan, "An Architecture for Interoperable GIS Use in a Local Community Environment", Computers & Geosicence, Elsevier, 2005, Vol. 31, No. 2, pp.211-220, March 2005

[3] A. Stanimirović, D. Stojanović, L. Stoimenov, S. Đorđević–Kajan, M. Kostić, A. Krstić, "Geographic Information System for Support of Control and Management of Electric Power Supply Network", IX Triennial International Conference on Systems, Automatic Control and Measurements SAUM, ISBN 86-85195-49-7, Niš, 2007.

[4] F. Menge, "Enterprise Service Bus", FROSCON, Sankt Augustin 2007.

[5] D. Coll-Mayor, M. Paget, E. Lightner "Future intelligent power grids: Analysis of the vision in the European Union and the United States", Energy Policy, 35(4):2453–2465, April 2007.

[6] A. Ipakchi, F. Albuyeh, "Grid of the Future", IEEE power & energy magazine, march/april 2009

[7] D. S.Linthicum, "Enterprise Application Integration", Addison-Wesley, ISBN 0-201-61583-5, 2000.

[8] Chappell, D.: Enterprise Service Bus. O'Reilly Media, Inc., ISBN 0-596-00675-6, June 2004

[9] M.-T. Schmidt, B. Hutchison, P. Lambros, R. Phippen, "The Enterprise Service Bus: Making service-oriented architecture real", IBM SYSTEMS JOURNAL, VOL 44, NO 4, 2005

[10] IEC - International Electrotechnical Commission: IEC 61970-301: Energy management system application program interface (EMS-API) – Part 301: Common Information Model (CIM) Base, 2003.

Development of E-government in Serbia and Bosnia and Herzegovina

Leonid V. Stoimenov[1], Nataša, Ž. Veljković[1], Sanja Bogdanović-Dinić[1], Srđan Č. Nogo[2] and Siniša S. Macan[2]

***Abstract* – Communication and information technologies as a means towards an information society impose the need to modernize public administration and to introduce electronic government as a state management tool. In this paper we will present the state of e-government in Serbia and Bosnia and Herzegovina at the central and at the municipal level covering the period of four years. In addition we will provide analysis of the present state of ICT infrastructure at the municipality level in both countries.**

***Keywords* – e-government, measurements of e-government, ICT infrastructure, legislation framework**

I. Introduction

The interpretations of e-Government are broad and divergent. Generally speaking, e-government can be defined as the use of ICT for delivering more effective and proficient government services to citizens, businesses or government agencies, by various communication means as the Internet, telephone, wireless devices or other communication systems. Effective e-government involves rethinking organizations and processes, and changing behavior so that public services are delivered more efficiently to the interested parties. As stated in [1] e-government contributes significantly to the process of transformation of the government towards a leaner, more cost-effective government. By intercepting appropriate technology, e-government can facilitate communication and improve the coordination of authorities within the different tiers of government.

To achieve the goal of e-government and improve the delivery of public services, ICT involvement is inevitable. ICT provides various ways for integrating information systems into one compact entity, offers relatively inexpensive infrastructure for data interchange and the possibility of accessing information and e-government services from any place.

The success of e-government implementation needs to be measured on all government levels, both state and municipality level. This can be done using various methodologies, as those listed in [2]. The most comprehensive methodologies are measuring the supply and the usage side of e-government services, as well as the state of ICT infrastructure. According to [3] the supply side of e-government can be evaluated through the sophistication and availability of 20 basic public services.

In the rest of this paper we will present the state of e-government in Serbia and Bosnia and Herzegovina (B&H), both on the state and on the municipality level. We will overlook the state of e-government from different viewpoints, including implementation of 20 basic public services, the legislative background and the state of ICT infrastructure.

[1]Leonid V. Stoimenov, Nataša Ž. Veljković and Sanja D. Bogdanović-Dinić are with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, e-mail: {leonid.stoimenov, natasa.veljkovic, sanja.bogdanovic.dinic }@elfak.ni.ac.rs

[2]Srđan Č. Nogo and Siniša S. Macan are with the Agency for Identification documents, register and data exchange B&H-IDDEEA, e-mails: srdjan@iddeea.gov, sinisa.macan@iddeea.gov.ba

II. Evaluation of E-Government On the State level

Evaluation of e-government on the state level can be done by measuring the development of 20 basic public services in terms of availability and sophistication, and percentage of individuals and enterprises that are using the Internet for interacting with public authorities. This evaluation is proposed in the i2010 benchmarking framework [3]. For the purpose of this paper we will also take into consideration the state of legislation and the ICT infrastructure, because they are important factors in implementing e-government services.

A. State of e-government in Serbia

E-government in Serbia in the past few years has significantly advanced, taking into account that the state has brought some of the necessary laws for its functioning [4]. In June 2009, the Ministry of Telecommunication and Information Society, has proposed a new Strategy of e-government development in the Republic of Serbia for the period 2009-2013. Special significance for e-government development in Serbia has the adoption and implementation of Law on Electronic Signature. Registration of certification bodies for issuing qualified electronic certificates enabled the use of qualified electronic signatures for signing electronic documents. In July 2009, the Law on Electronic Document was passed and it would have a major impact on the process of modernization, rationalization and the introduction of electronic government. The new Law on Public Procurement entered into force on January 2009 enabling, among other, the possibility of electronic procurement, including electronic auctions. One of the key innovations, brought by this Law, is the introduction of customer commitment to publish ads electronically on the public procurement portal, instead of in a

daily newspaper, as it was the case previously. In the next period, the Government of the Republic of Serbia plans to adopt a few more laws (such as Law on Electronic Government, Law on the Protection of Personal Data in Electronic Form and Law on Electronic Archives) that will significantly improve the implementation of e-government.

First evaluation of e-government in Serbia was done in 2004. Since then, benchmarking of e-government has been done yearly by measuring the sophistication of 20 basic public services and their online availability, according to methodology described in [5]. Sophistication of public services on the state level has grown from 15.25% (2004) to 47.42% (2008) [6] [7]. That is a total growth of 32% for the four years period. Services offered to businesses are more sophisticated in comparison with citizens' centric services [7].

When comes to the implementation of particular e-services, we can say that the most sophisticated are: Job search, Public procurement, Public libraries and Customs declaration.

Through Public Procurement Office portal it is possible to query existing procurements or electronically publish new and monitor their status. On the Web site of the National employment agency it is possible to search job database that covers the whole territory of Serbia as well as to register in order to apply for job. Through the state Web portal for customs declarations it is possible to submit the required documents electronically; in that way customs clearance takes much less time than usual.

Health related services are, on the contrary, least developed. Scheduling medical examination is not feasible through the Web site of the health care center or other state medical institutions. It can be done only by coming to the health care institution or in rear situations by phone calls.

E-Government internet portal was launched in March 2007, and it served as a source of information about basic public services and the corresponding authorities. In 2010 it was upgraded from information oriented to service-oriented Web portal. Its final function is to be a "one-stop-shop" by enabling users to easily and rapidly access information and electronic services of public administration from any place.

B. State of e-government in Bosnia and Herzegovina

In the last few years B&H made significant progress in bringing regulations in the field of telecommunications and ICT in the public government. Lack of systematic, organized and methodical approach to the application of e-technology in public administration contributes to the fact that B&H is one of the last countries in the region of Southeast Europe that has adopted IT strategy. In 2005, the B&H Council of Ministers adopted Policy, Strategy and Action plan of the information society that is focused on legal infrastructure, e-Education, e-Governance, ICT infrastructure and ICT industry development.

B&H has gradually begun adoption and application of the law on electronic communications infrastructure and related services in accordance with the guidelines provided by the relevant Directive of the European Union. In recent years B&H has made major progress in terms of regulations in the field of telecommunications, and therefore in the area of application of e-technology in public administration.

Official analysis of 20 public services in terms of sophistication and availability haven't been done so far. However there are some partial analysis prior to 2009 [8]. The progress in e-services implementation is evident for job search and library services. The Agency for Public Administration partially offers job search on-line through a portal. National University library is developing an integrated overview of individual library funds [8].

B&H is still a long way behind world-wide transactional and cross-organizational services as stated in [9], there is encouraging awareness of internet use as a tool for public institutions to disseminate information; and a majority of institutions, regardless of level, have their own Web sites.

Most of the authorized institutions offer on their Web sites basic information for citizens and other interested parties, but there aren't any interactive electronic services. Public administration of B&H is still missing official e-government Web portal with on-line services available to citizens and links towards sub-portals at state, entity and other levels. Portal for e-services should improve the effectiveness of B&H administration authorities, by providing the possibility to review and collect information and by enabling appropriate data exchange using e-services, as foreseen by Bosnia and Herzegovina laws. In 2009 a Web portal of public administration of Republic of Srpska, providing e-services to citizens, has being released at entity level and named *e-srpska*. These kinds of portals help B&H to fulfill required clauses for approaching European Union considering 12 e-services for citizens that each e-government, willing to become a member of EU, has to conform [10].

It is necessary to develop information systems with support for horizontal functions, that is, processes which are common to majority of institutions. This means that institutions can automate some of the most commonly used procedures and so produce larger budget savings, with only one software solution. Most of initiated and implemented projects since 2004 until now have been dealing with the need for resolving some critical issues of inherited situation in this area, and therefore were designed as solutions to one and only problem that would appear in certain public administration institutions. As a result of such actions, enormous differences in levels of applied e-technologies have appeared among public institutions, varying from very modern systems and well developed networks (such as the Agency for identification documents, recording and exchanging data B&H - IDDEEA) to very obsolete solutions [11].

Subject to document eSEE AGENDA+ for development of information society in countries in the South East Europe 2007-2012, which is in accordance with the document i2010 European information society for e-Government [3], IDDEEA has a responsibility to develop through this Project three e-services for citizens:

- Personal documents (ID card, Driver License and Travel documents)
- Registration of vehicles (new, used or imported vehicles)
- Reporting on change of residence (address change)

III. Evaluation of E-Government On the Municipality Level

A. Serbia

Large number of cities and municipalities offer trough their Web sites, download forms and information of interest for citizens and businesses. Information that local governments most commonly post on their Web sites often includes description of government departments and officials, contact information, economic development data, the local government law book or special announcements. Some local government Web sites offer the ability to print forms that citizens can fill out and mail or bring to the government office. More elaborated sites offer interactive forms that can be filled out and submitted online.

The most common electronic services in Serbian local self-governments are: Virtual registry office, Voter's lists and e-Notary. The service "Virtual registry office" is used for ordering birth certificates (which are rarely supported by transactions operations - payment of certificates fees at the time of ordering). "Voter's list" is an online service used by citizens for checking the list before elections, in order to see if they are on the list and contact authorities if opposite. Some municipalities provide a complete monitoring of citizens' cases via the Internet through the e-Notary. This service is based on internal municipal service that allows movement and processing of case files in electronic form, without using the internal delivery book. Using the code obtained after submitting the request at the municipality in person, citizens are able to track and receive information regarding the status of their cases over the Internet.

Some municipalities have so-called citizens assistance centers that operate according to the principle of "one stop shop" – all services are provided in one place. Service centers have counters for work with citizens and/or legal persons. Counters are mostly organized in a way that every counter is dedicated to one municipal service. In some cases assistance centers provide even more - submissions are electronically distributed to appropriate municipal services due to the unique information system and the specialized "back office" software. All the electronic services to citizens and those that are available over the Internet in these municipalities are based on this "back office" software.

Majority of municipalities in Serbia so far have no information system upgrade with applications that allow the digitalization of paper documents and complete electronic processing. Majority of municipalities, which is certainly a good example of introducing e-government within the municipality, have electronic registries. The existence of information about citizens in electronic form provides the possibility of issuing birth certificates quickly and easily, thus reducing congestion in the citizens' assistance center. Many cities/municipalities provide issuing of birth certificates in local government branch offices, cited in remote locations that belong to the territory of the city/municipality. In order to implement this type of service it is necessary to have a unique database of citizens at the state level, or to provide access to independent databases of cities/municipalities based on digital certificates.

When comes to the state of ICT infrastructure, as stated in [12], there is still 15% of municipalities with almost no infrastructure. It is evident that all municipalities without ICT infrastructure are among the group of weakly developed or underdeveloped municipalities. For example, 13% of municipalities have no servers, 15% of municipalities have no network printer, in 6% of municipalities network infrastructure doesn't exist, in 8% of municipalities less than 10% of employees have no Internet access, in 7% of municipalities less than 30% of employees have a computer in their workplace. On the other hand, 75% of municipalities have from 1 to 5 servers, while 13% of municipalities owns more than 5 servers, 40% of municipalities have an email server, in 60% of municipalities over 50% of employees have computers in their workplace, in 47% of municipalities over 50% of employees have an Internet access. The encouraging fact is that all municipalities have Internet access and that a large number of municipalities have a sufficient number of computers, networking equipment and servers for basic operational needs.

B. Bosnia and Herzegovina

Public administration of B&H provides services to citizens in 146 municipalities, 2 entities, in Brčko District and at the state level of B&H. At each level there are various problems involved in proper functioning of public administration. Principles and practice at the municipality level or a canton are different from principles adopted in other municipality, canton or B&H. The existing state at all public administration levels in Bosnia and Herzegovina shows there is lack in understanding the importance of e-technologies that could be involved in its functioning. Introducing e-government into public administration of Bosnia and Herzegovina is mostly considered as computerization of certain business processes, and then as a means for reforming public administration. There is a lack of transparency regarding work and finance matters. There is neither horizontal nor vertical electronic communication. Public administration offices use different operating systems, applications and database systems due to lack of standards and no global plan for introducing information technologies in the state administration. Web sites of public administrations provide citizens with not many services, not frequently updated information, and rarely include forms that can be printed locally [8].

Technical capabilities of administration units are somewhat satisfying. Large number of local legal and administrational organizations is properly technically prepared and has quality information equipment. However, that equipment isn't often used in appropriate and satisfying way and in accordance with regulations and procedures, which are established by competent institutions. 55% of employees in administration know how to use a computer, but only 5% of professional IT staff is employed in IT or equivalent sectors. Internet connection lacks in usage. There is formal presence of e-government online in the form of only several municipality, entity or state Web pages, basically used as official sources of

information. These pages provide users with static data about government institutions, sectors, ministries, agencies etc, as well as with contact information, but largely lack at actuality of data. One particular project that has been successfully implemented is the Project for municipalities' needs for issuing Birth and Citizenship certificates for the purposes of issuing Biometrical Passport. Usage of this Web solution has significantly shortened the lines in municipalities and lowered the pressure on referents.

The problem of verification of breeder documents in the process of identity chain (issuance of travel documents) is recognized as the weakest link of the system. Related experiences worldwide are equal to those of B&H. The Agency is trying to raise the work quality level in this field as high as possible in order to harmonize the process of issuing documents with the European standards [13]. The Law on Travel Documents does not impose the obligation that an individual who possess an identification document must submit an application for issuance of travel document accompanied with a birth certificate and certificate of citizenship. It is assumed that such check will be conducted during the process for issuance of identity card. Unfortunately, it is evident that the process for issuance of identity cards had a range of weaknesses, mainly in the field of breeder documents. Therefore, data currently contained in the civil register cannot be taken as absolutely accurate.

Thanks to the Governance Accountability Project - GAP, 70 municipalities at the B&H territory will be a part of the uniform system. Through implementation of this system, besides fulfilling of the requirements to introduce the system of biometric travel documents, municipalities will obtain appropriate network infrastructure as a basic condition for establishment of municipality information system: electronic registers, data exchange between municipalities, e-government, PKI, electronic signature-verification, etc.

Due to unconcern of the institutions for local administration, being uninformed and because of weak awareness related to process for meeting requirements for the purpose of accession to the EU, financing and realization of this project by local administrations themselves is not possible in a reasonable period of time. Certainly, there are municipalities which will meet the required conditions as specified by EU standards and procedures, but it is even more certain that there are municipalities where such process will last unreasonably long.

Responses to the questions on possession of Electronic Registers, existence of local networks, position of municipality election commissions (OIK), etc, from all municipalities in B&H were provided in cooperation with GAP. The obtained data confirms that 67% of municipalities already have electronic registers and that 83% municipalities have already met the technical requirements for the access to the information system of the Agency.

III. CONCLUSION

E-government has been a very important field of exploration in past few years, especially for the developing countries of the Balkan region. If successfully implemented e-government will bring benefits as better efficiency, accountability and accessibility to public services.

In this paper we have presented the current state of e-government in Serbia and B&H, from different aspects covering the period of four years. We have identified the critical factors related to e-government introduction, on the state and on the municipality level. They are focused around the following key drivers: legislation framework, ICT usage, systematization and education of personnel, municipal leaderships support and appropriate funding. With the proper usage of ICT, countries have opportunity to make a significant progress in a relatively short time period and to place themselves among efficient, citizen oriented e-governments.

REFERENCES

[1] M. Datar, A. Panikar and K. Farooqui, "Emerging Trends in E-Government", Critical Thinking in E-Governance, Editor J. Bhattacharya, GIFT Publishing, 2008, pp.37-46, available online: www.iceg.net/2008/books/1/4_37-46.pdf

[2] M. Sakowiz, "How to Evaluate E-Government? Different Methodologies and Methods", UNPAN, unpan1.un.org/intradoc/groups/public/documents/un/unpan020076.pdf

[3] i2010, 2011–2015 benchmarking framework, http://ec.europa.eu/information_society/eeurope/i2010/benchmarking/index_en.htm, Accessed 6 April 2010.

[4] Stoimenov, L., Markovic, N. et.al, "Best ICT practices in Serbian municipalities", Faculty of Electronic Engineering, Nis, http://gislab.elfak.ni.ac.rs/egovernment/Best_ICT_Practices.pdf, 2009.

[5] Prepared by Capgemini, "Benchmarking the supply of online public services", 2007, ec.europa.eu/information_society/eeurope/i2010/docs/benchmarking/egov_benchmark_2007.pdf

[6] M. Gusev, G. Armenski, "Gap Analysis of eGovernment in Western Balkans", www.metamorphosis.org.mk, 2006.

[7] L. Stoimenov, N.Markovic, et al., " E-government evaluation in serbia based on implementation accessment of 20 basic public services", Faculty of Electronic Engineering, Nis, http://gislabweb.elfak.ni.ac.rs/egovernment/e-government_study_2007.pdf, 2008. (In Serbian)

[8] S. Gerin, B. Vujičić, "eGovernment Services in Bosnia and Herzegovina", Informatica, 31, 1,pp.373-377., 2007.

[9] Office of the PAR Coordinator, "Systematic Review of Public Administration in B&H", Final Report, Book 5, 2005.

[10] Council of Ministries of B&H, Directorate for European Integration B&H, Strategic of integration Bosnia and Herzegovina into Europion union, Sarajevo,B&H, 2006.

[11] S.Nogo, S. Macan, "E-services", SMART E-GOVERNMENT conference, Belgrade, 2009.

[12] National Information Technology and Internet Agency, "Analysis of ICT infrastructure in Serbian cities and municipalities", http://www.rzii.gov.rs, 2009. (in Serbian)

[13] Final report, Project regarding connection of municipalities in the process for implementation of project for biometric travel documents - e-passports., IDDEEA- SIDA-Sweden, 2009.

Software Module for Clinics of Neurology as a Part of Medical Information System Medis.NET

Aleksandar M. Milenković[1], Petar J. Rajković[2], Dragan S. Janković[3], Tatjana N. Stanković[4] and Miroslava D. Živković[5]

***Abstract* – This paper presents software module that provides all necessary functionalities for clinics of neurology. Module Medis.Neuro has been developed at Faculty of Electronic engineering in Niš as a part of goverment project which is funded by Ministry of Science and Technological Development of Serbia, for needs of Clinic of Neurology in Niš. Medical information system MEDIS.NET has been developed under mentioned goverment project that covered most of funcionality which one modern medical information system should provide for healthcare institutions in Serbia. One of developed moduls under MEDIS.NET is modul Medis.Neuro. Collaboration with primary healthcare institutions is also presented in this paper. Data exchanging between healthcare institutions is realized using web services on both sides.**

***Keywords* – software module, information system, clinic of neurology, health care, Medis.Neuro, Medis.NET, web services.**

I. INTRODUCTION

Neurology is a medical specialty concerned with diagnosis and treatment of all categories of disease involving the central, peripheral, and autonomic nervous systems, including their covering, blood vessels, and all effectors tissue, such as muscle [1]. Now it is common that each clinical center has special department for neurology. Modern department for neurology requires modern medical equipment which includes electronic devices such as contemporary scanners, MRI (magnetic resonance imaging), modern surgery room with precision medical instruments and sophisticated information system for tracking patient disease and treatment history (anamnesis).

The Clinic center of Niš is one of the bigest health care clinic centers in Serbia which also includes Clinic of Neurology. Number of patients that visited Clinic of neurology in Niš from 2000 to 2007 is shown on Figure 1.

[1]Aleksandar M. Milenković is with the Faculty of Electonic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, e-mail: aleksandar.milenkovic@elfak.ni.ac.rs

[2]Petar J. Rajković is with the Faculty of Electonic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, e-mail: petar.rajkovic@elfak.ni.ac.rs

[3]Dragan S. Janković is with the Faculty of Electonic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, e-mail: dragan.jankovic@elfak.ni.ac.rs

[4]Tatjana N. Stanković is with the Faculty of Electonic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, e-mail: tatjana.stankovic@elfak.ni.ac.rs

[5]Miroslava D. Živković is with the Faculty of Medicine, Bulevar Dr Zorana Đinđića 81, 18000 Niš, Serbia, e-mail: miraz@medianis.net

Patient at health center receives referral for hospital treatment. Referral usually contains information of location where patient is referred, date of referral validity, referral diagnose and referral text.

At the moment patient arrives at clinic of neurology, it is necessary to establish his anamneses. That includes examination of large number of parameters in order to determine correctness of discussion about current condition of the patient. After reception of patient at clinic of neurology, his condition is constantly monitored: different kind of analysis and tests are conducted. Large numbers of parameters are recorded. Some of the parameters need to be measured with great frequency and to be compared with previous parameters for observation of the course of treatment. All the facts mentioned above indicate on existence of a large amount of data stream that must be stored in efficient way. Data access must be efficient, as well. Chronic diseases as well as some specific (vital) diseases that patient had can be of great importance for further therapy. All that is stored in patient's health record which can be found in Health Center Niš. Data searching can be difficult and exhausting task with great possibility that such data is nonexistent or lost. There are some situations where patient has no possibility to bring necessary data from health center as well as to properly answer on relevant questions at clinic of neurology. After hospital treatments, patient receives discharge letter that need to be forwarded to chosen general practitioner at health center.

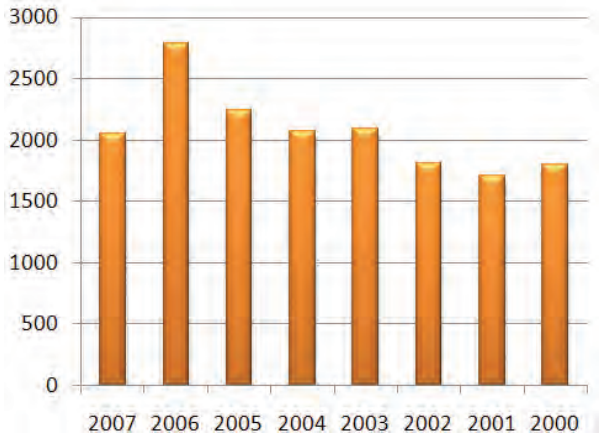


Fig. 1. Number of patients for period 2000 to 2007 at Clinic of Neurology Niš

All the facts mentioned above refer to conclusions that application of information system at clinic of neurology is essential as well as necessity in its modern conduct.

Information system should provide efficient input, review and searching of vast amount of data, to represent comfortable and user-friendly environment for the end users [2]. When applied, information system should provide and accelerate necessary data transfer. The fact mantionend above is refered to the use of information systems at clinic of neurology and at health center, as well as collaboration between them. At the same time it should provide access only to data (from electronic health record) which are significant for the clinic of neurology. During data exchange, it is necessary to encrypt sensitive data for the reasons of preventing abuse in case of their interception. At the end, OLAP cubes can be created against database to extract required statistics.

In the design and development of Medical information system MEDIS.NET (realized at Faculty of Electronic Engineering in Niš, as part of the project of Ministry of Science and Technological Development of Serbia and Health Center Niš as one of the project participants) there was a requirement for developing clinic of neurology module [3]. Proposed solution demonstrates most of functionalities that one modern and efficient module for clinics of neurology should have. Part of mentioned functionalities will be described in the following section.

After introduction, same functionalities of module for clinics of neurology are presented in this paper. Architecture view of module is given after the section which describes module functionalities. At the end of the paper, after architecture view, there is a short conclusion.

II. Overview of Module Medis.Neuro Functionality

After correct logging on the system, user enters application Medis.Prijemna. Application interface appearance is different for different users. It depends on user privileges which are defined by the system administrator [4]. If active user is an employer at clinic of neurology, and if he has privileges which are defined by module for user privileges configuration, module Medis.Neuro will appear. Medical nurses and doctors should not have access to the same patient's medical data (some of them should have access to read data, other to create new records and to update existing). All data entered and stored in database are without possibility of physical removing. Instead of deleting record, flag invalid data is set in database tables for all rows which became invalid. Most of the data are sensitive so it is recommended to add status attributes for each row: who added record, date and time of adding record, who set status flag for specific record (record is valid, record is invalid, new version of it exists, record is deleted) and timestamp when this action was executed. This is necessity for tracking all changes in done against the data source.

When receiving patient at clinic of neurology, the action of searching and finding patient's data need to be performed by the system. Searching the patient can be done using following search criteria: unique personal number (JMBG), insurance number (LBO), first name, last name. Different combinations of the mentioned filters for searching can be used. Usage of barcode reader is also available. When searching for patient through database does not give results, adding the new one is possible. Updating all demographic patient information found in the database is also allowed. After finding (or adding) the patient, user can access all functionalities of the module of neurology Medis.Neuro by clicking to the button "Active history of diseases". After that, main form of mentioned module appears on the screen. Appearance of the main form is shown in Figure 2. Functionalities of this module are organized by tabs (*Reception*, *Clinical findings*, *Diagnostic methods*, *CTM*, *Scale for patient evaluation*, *Temperature list*, *Drugs and materials*, *Discharge*, *Exitus*). All forms show basic information about selected patient for the reason of notifying user about the patient constantly. Software tool for rapid development and customization of medical information systems, developed at Faculty of Electronic Engineering Niš (in mentioned government project), is used to generate some of module forms [5].

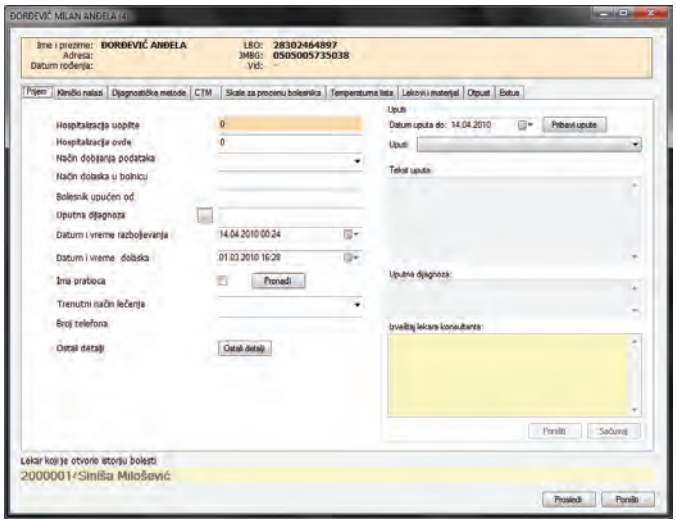


Fig 2. View of the module Medis.Neuro main form

Reception tab refers to patient's reception at clinic of neurology. It is necessary to fill all basic information about the way of patient's arrival, who brought him, number of patient's arrivals at clinic of neurology (it is assumed that he was there before, but not necessary), date when patient came, date when disease begun, patient's phone number. By mouse click on the button *Other Details*, special form will appear for more details. The right side of the main form contains details about the patient's referrals. The referral that was written by doctor at health center can be retrieve to the clinic of neurology by using web service which is the property of health center. This action can be done from the module Medis.Neuro main form, but the assumption is that health center uses our information system MEDIS.NET also. On this way doctor at clinic of neurology is immediately able to see referral from the health center with details: date when referral was created, referral expiring date, referral's text and reception diagnoses. Also, doctor can give his opinion that will be saved at the health center database over available web service (right side of the main form).

Data obtained by patient examinations can be entered at form tab *Clinical findings* (Figure 3) (*headaches*, *current*

disease, personal anamneses, family anamneses, psychological status, status praesens, status neurologicus, internist examination, decursus). Each form has numerous parameters. That is the reason that clear and good design of the forms with numerous parameters in exactly defined order is required, so the end users could have an easy way of navigation through the forms. There are possibilities to add new data record for each examination. Any new entries are always stored with a new date. Depending on available privileges, logged user can update examination records for selected date.

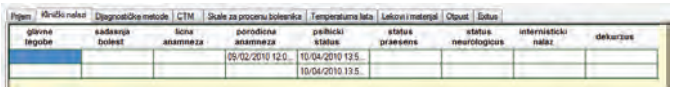


Fig 3. Example of some *Clinical findings* for selected patient

The patient's examinations, using diagnostic devices and diagnostic skills, are provided by *Diagnostic findings* and *CTM* tabs on the main form. These tabs support next examinations: *fundus, liquor, biochemical tests, haemostatic status, urine, acid-base status, Doppler, angiography, CTM infractus cerebri, CTM haemorrhagio, CTM of reduction changes*. Module can be integrated with biochemical laboratory information system (LIS) for accepting results after performed analysis [6].

Scales for evaluating patient's condition are *Glasgow coma scale, European stroke scale* and *Barthel's index*. All scales are on the tab *Scale for patient evaluation.* Each scale is presented as special form with parameters which can be chose from combo box list. Each parameter's value has specific weight. After filing parameters, the sum of them all represents evaluating factor.

merenja	16/02/2010 11:31:27 AM	16/02/2010 11:31:11 AM	16/02/2010 11:30:55 AM
temperatura	40.000	38.900	37.500
sistolni_pritisak	70.000	100.000	120.000
dijastolni_pritisak	40.000	60.000	89.000
puls	40.000	60.000	75.000
napomene			

Fig 4. Graphical representation for measured *temperature*, *blood pressure* and *pulls* as part of tab *Temperature list*

Tab *Temperature list* is graphical representation for measured *temperature, blood pressure* and *pulls* on the specific date (Figure 4). This way monitoring of mentioned parameters at selected date is very easy.

Tab *Drugs and materials* is referred to applied therapy and given therapy. Monitoring of prescribed and used medication can be done with this tab.

If patient dies at clinic of neurology, it is necessary to write some note about that at tab *Exitus* at the main form. Tab *Discharge* covers all necessary data that must be entered when patient ends his treatment(s) at clinic of neurology. Discharge date, in what condition patient was discharged from hospital as well as discharge letter, is necessary to be filled out. When patient comes at health center, doctor, using the web service from clinic of neurology can get discharge letter at the patient's electronic health record (EHR).

Finally, OLAP cubes can be created on database to extract required statistics for improving quality of given medical care and service as well as improving methods of patient treatments.

III. Architecture View

For module developing we have utilized a classic three-tier architecture which includes three logical layers (data layer, business layer and presentation layer). The logical layers are shown in Figure 5.

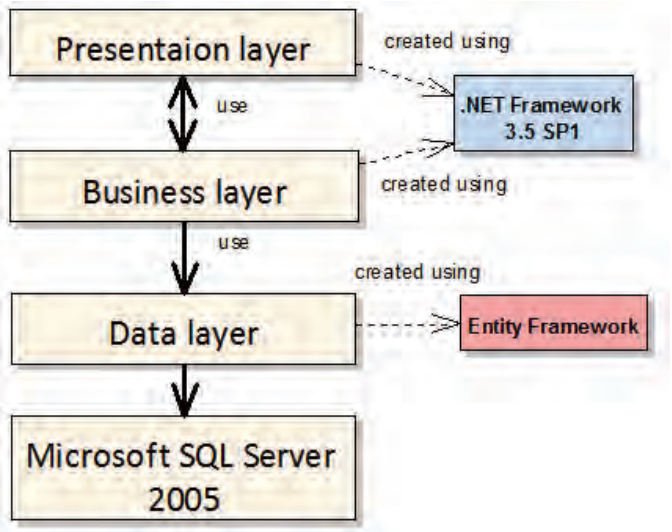


Fig. 5. Three-tier architecture of module for clinics of neurology

Data layer is implemented under business logic layer. The function of this layer is to access directly the database, i.e. to accept data from database and forwards them to a higher layer. On the other hand, it is necessary to accept data from business logic layer and store them to the database. Data layer is implemented by using Microsoft Entity Framework [7]. Microsoft SQL Server 2005 is used for database platform. With small changes in this layer there is possibility to connect module with different database platforms.

Business logic layer contains all the necessary functionality of this module.

Presentation logic layer was implemented by using .NET Framework 3.5 SP1. This layer should provide easy and simple way of using application by end users. Primary objective of this layer is fast and easy data entry.

The doctor from health center can write referral for patient's hospital treatment by using electronic health record (EHR). When patient arrives at the hospital for treatment, (clinic of neurology in this case), doctor can receive written referral from health center by using web service which is the property of health center where referral is originally made. Web service at clinic of neurology provides to doctors in health center possibility of receiving discharge letter over the internet. Received data from clinic of neurology (discharge letter) is available to doctor at health center as any data in electronic health record. All sensitive data (method parameters) are encrypted within the request sent to web service. Web service decrypts data in calling method which was called by client application. After that, web service processes request and try to find searching data in the database. Required data are sent as a response to the request of client application. Before sending data, sensitive data are encrypted in web service method and then sent over the internet to the client application in health center. Client application must decrypt received data. Finally, data can be shown to doctor through client application (EHR) in health center. At the end of the process, health center doctor can see patient's discharge letter. DES algorithm is used for data encrypting. In this way interception of sensitive data and their abuse is prevented. Figure 6 shows entities which participate in data exchange and their interconnection.

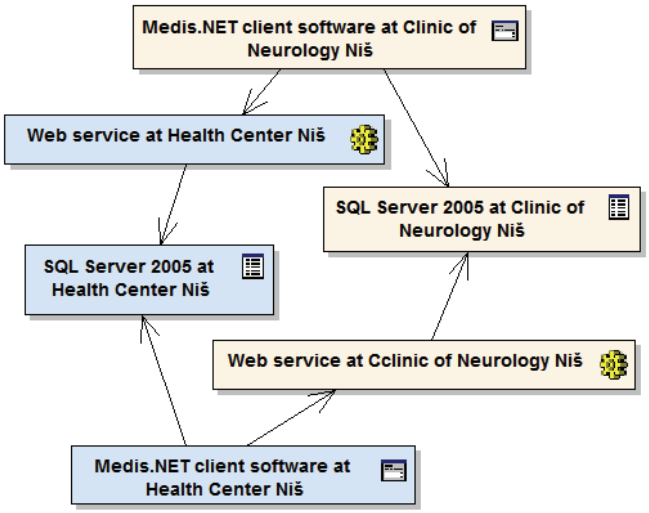


Fig. 6. Data exchange between information system at Health Center of Niš on one side and information system at Clinic of Neurology of Niš on the other side

IV. Conclusion

The existence of modern working processes at clinics of neurology demands using of modern information system. This paper describes software module Medis.Neuro as a part of medical information system MEDIS.NET that provides most of the required functionalities of clinic of neurology. Application of module Medis.Neuro should improve business of clinic of neurology trough accelerating working process (avoiding multiple data input), easy monitoring patients' condition, easy access to the large amount of important data for patient treatments to doctor, extracting statistics for improving quality of given medical care and service. The daily routine of work is established, so this module should not come upon to resistant of the medical personal (end system users) [2]. It is in coordinated with natural input of data as well as with natural order of necessary examination. Medis.Neuro is reliable for all clinics of neurology in secondary health care. A complete adjustment of this module has been done for concrete necessities at Clinic of Neurology in the Clinical Center Niš.

Created module is in the test phase and experimental work, in order of showing advantages and disadvantages. After the optimization phase, it will be implemented at the Clinic of Neurology Nis.

Thanking Note

The research presented in this paper were funded by the Ministry of Science and Technological Development of Serbia within the project in the field of technological development, "Improvement, integration and collaboration of information systems in health care institutions" No TR1305.

References

[1] http://en.wikipedia.org/wiki/Neurology – Neurology. Last access time 10 April 2010.

[2] Tatjana Stanković, Aleksandar Milenković, Dragan Janković, "From optional talk to medical information system's user interface", INFOTEH 2010, Jahorina, March 2010, CD Proceedings, ISBN: 99938-624-2-8.

[3] http://medisnet.elfak.ni.ac.rs – Laboratory for Medical Informatics at Faculty of Electronic engineering in Niš. Last access time 10 April 2010.

[4] Aleksandar Milenković, Petar Rajković, Dragan Janković, "Module for configuration of user privileges in medical information system", INFOTEH 2010, Jahorina, March 2010, CD Proceedings, ISBN: 99938-624-2-8.

[5] http://medisnet.elfak.ni.ac.rs/node/61 – Software tools for rapid development and customization of medical information systems. Last access time 10 April 2010.

[6] Aleksandar Milenković, Dragan Janković, Dušan Vučković, Tatjana Stanković, "LabIS - Information system for biochemical laboratories", YU INFO 2010, Kopaonik, Republic of Serbia, March 2010, CD-Proceedings, ISBN: 978-86-85525-05-6.

[7] http://msdn.microsoft.com/en-us/ – The ADO.NET Entity Framework Overview. Last access time 10 April 2010.

An Implementation of a Scheduling Tool in a Medical Information System

Ivica M. Marković[1], Stevica S. Cvetković[2] and Dragan S. Janković[3]

***Abstract* – In order to achieve high quality in providing healthcare services with limited medical personnel resources an appropriate scheduling strategy for both staff and patients is required. Therefore a carefully designed IT solution for this purpose which is described in this paper can significantly improve scheduling process in a healthcare institution and make it more efficient.**

***Keywords* – Medical information system (MIS), Staff and patients scheduling, Appointment scheduling.**

I. Introduction

High demands in providing healthcare services together with limited medical personnel resources require a proper scheduling of staff and patients. It is known that efficient utilization of manpower has always been a key concern in any organization and one of the most important means to achieve productivity gain. For this reason, the staff and patient scheduling is an area that has become increasingly important in healthcare services. Also the working conditions of the staff in the health services have an important effect on quality of provided services and overall health of the patients [1], [2].

Scheduling process in a healthcare institution is an activity that consumes a large amount of administrative resources and, above all, time of the users themselves. Also appointments for certain diagnostic examinations are planned for 2-3 months in advance. Therefore primary contribution of this study is to focus on real world problems in medical scheduling services and to suggest a complete scheduling solution that can be configured and understood easily. Such a solution should have the following properties:

- Simple, clear and intuitive user interface suitable for use by medical personnel with a limited IT knowledge;
- System configurability based on the parameters stored in the database;
- Preventing conflicts during appointment scheduling and taking care of matching patients with their chosen doctor;
- Centralized scheduling system that provides insight into the status of all participants within the system;
- Generic reporting mechanism.

Due to complexity of the problem, entire model is broken down into two manageable tasks: (1) creating work time schedule for medical personnel and (2) scheduling of appointments for patients.

Our scheduling solution, Medis.NET Scheduler, is a part of a medical information system (MIS) which is being developed at The Faculty of Electronic Engineering in Niš (http://medisnet.elfak.ni.ac.rs) in cooperation with The Health Care Center in Niš [3]. Development of medical information system is a central part of a project supported and funded by The Ministry of Science and Technological Development of Republic of Serbia.

In the rest of the paper we first describe main use cases of the application and after that we present an overview of our system architecture.

II. Application's Main Use Cases

As already mentioned above we split scheduling problem into two parts and therefore two main groups of use cases can be analyzed – medical personnel work time scheduling and patient appointment scheduling as shown on Fig. 1.

A. Medical personnel work time scheduling

Work time scheduling tool in Medis.NET Scheduler is a tool which could be used for creating work time schedule for employees in any organization or company although our focus here is on its application within health care institutions. It is suitable for creating schedule for employees who do not have fixed work hours but instead change their work time by shifts. Moreover this tool supports changing of shifts by different complex patterns which we call work schemes (or work patterns). Main feature of work time scheduling tool is automated generating of work time schedule based on preconfigured data. An effort to enter configuration data should be made only once and after that schedule for employees is generated automatically at certain time periods. Later user actions should only include modifications of configuration data and entering exceptions from configured pattern such as absences from work.

Initial configuration of the tool should include the following actions:

- Connecting employees to the offices in which they work;
- Entering data about shifts (for example first shift from 7 AM to 2 PM, second shift from 1 PM to 8 PM);
- Entering data about breaks (break times are not fixed for a shift but depend on organizational unit also);
- Creating work pattern which defines the sequence of shift changing (for example first shift during one week and

[1]Ivica M. Marković is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: ivica.markovic@elfak.ni.ac.rs

[2]Stevica S. Cvetković is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: stevica.cvetkovic@elfak.ni.ac.rs

[3]Dragan S. Janković is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: dragan.jankovic@elfak.ni.ac.rs

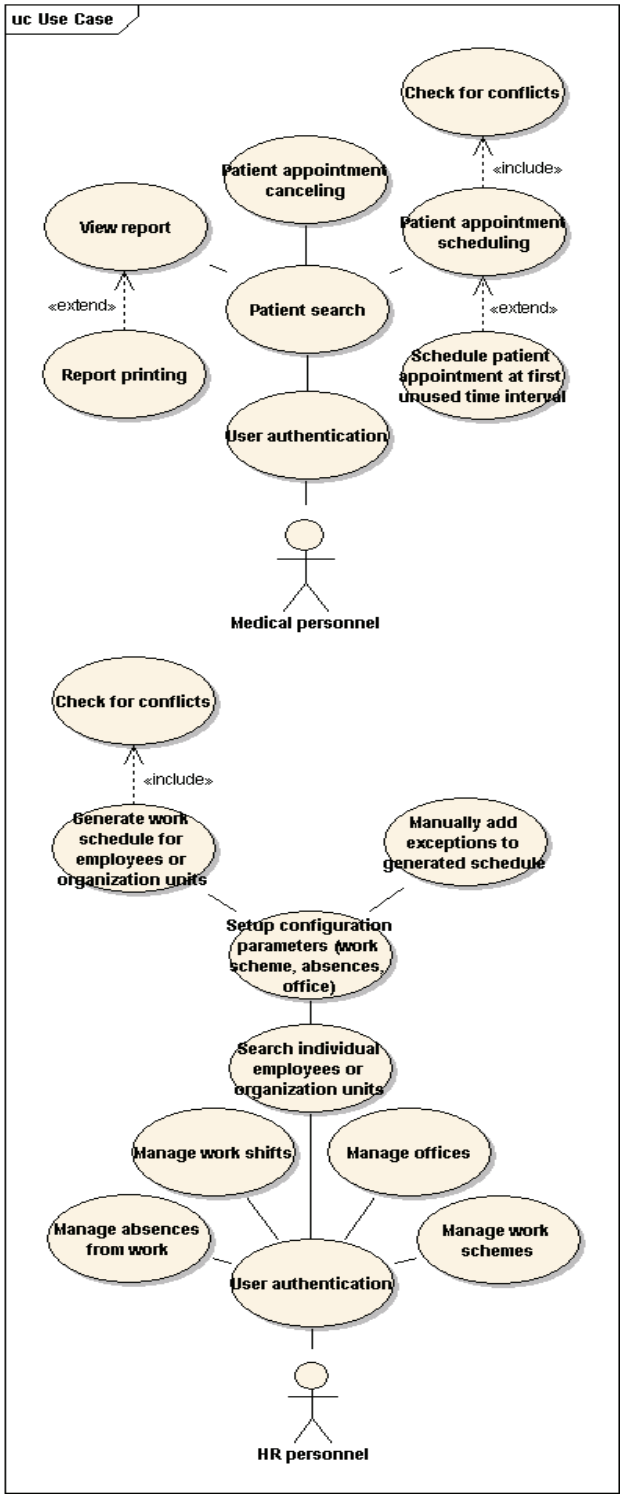


Fig. 1. Use case diagram

second shift during the second week, or maybe always work in first shift);

- Choosing an appropriate work pattern for each employee.

After initial configuration of the tool operator might only need to perform the following actions:

- Handle eventual changes in initial configuration (this should be very rare);
- Insert data about planned holidays of the employees;
- Change generated work schedule in case of employee's sudden absence from work.

This configuration tool, besides managing work time for employees, can also be used for managing configuration parameters required for patient appointment scheduling. One such parameter is average estimated examination time. This parameter is provided by Serbian Ministry of Health, and is defined for each medical specialization. For example, average estimated examination time of gynaecologist is 15 minutes, while for general practitioner is 10 minutes.

B. Patient appointment scheduling

After logging to the system, a medical personnel performs appointment scheduling in two successive stages. The initial stage is performed in the main GUI window that is used to find the patient in the database. At the second stage, doctor's appointment scheduling for the selected patient is carried out. This stage is conducted on a separate GUI window (AppointmentSchedulingWindow).

In order to enable efficient and visually intuitive scheduling tool, we developed custom TimeGrid visual control. It occupies central part of the AppointmentSchedulingWindow and offers to the user ability to assign the patient to doctor by simple mouse click (Fig. 2). The TimeGrid control displays all available appointment time intervals as grid cells, whereby already appointed time intervals are pink colored, while free intervals have white color. During software design phase, system configurability was considered as an important issue. As a consequence of this, all available time intervals are generated during application run-time, based on the doctor's working time stored in the database and maintained by HR personnel. Default layout displays TimeGrid for selected doctor in the next seven days, but it is possible to display time intervals for arbitrary selected period. A mouse click on a white cell results in doctor's appointment in the time interval specified by the cell. Appointment cancelling could be done by double click on already appointed time interval cell. As additional functionality, automatic appointment of the first not scheduled time interval is offered to the user. Generic report mechanism offers wide range of reports to be generated and printed, including a report which contains list of appointed patients for the next working day.

It is important to emphasize that our software is able to successfully prevent different scheduling conflict situations. There are two conflict scenarios, which must be prevented:

- Scheduling of a patient at one time interval for the two or more doctors;
- Scheduling of a doctor at one time interval for the two or more patients.

First step in the conflict prevention is implemented through GUI (TimeGrid). Actually, all time intervals which are already appointed (pink colored cells) are disabled for further scheduling. However, since the system is used in demanding multi-user environment, it is necessary to perform additional conflict checks immediately before database inserts. Therefore, our approach is to check for conflicts on temporarily locked tables and perform insert into database, all in a single transaction.

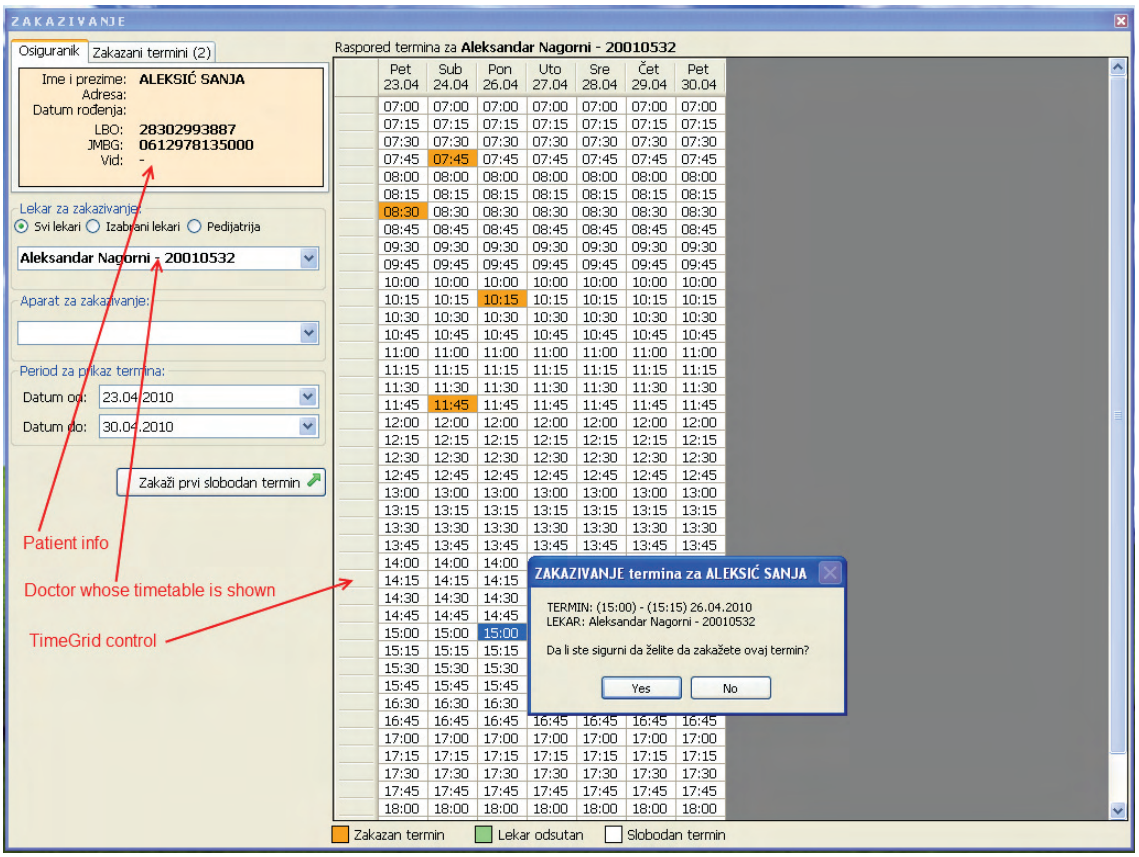


Fig. 2. Appointment scheduling window

III. System Architecture

Application Medis.NET Scheduler, as well as whole Medis.NET system, is a three-layer application which consists of the following three layers: data access layer, business logic layer and presentation layer. It is implemented using Microsoft .NET Framework 3.5 and integrated development environment Microsoft Visual Studio 2008 SP1. Application's database is operated on Microsoft SQL Server 2005.

For implementation of **data access layer** we used ADO.NET Entity Framework [4], a solution for object-relational mapping (ORM) within .NET Framework. ORM resolves a well known "impedance mismatch" problem between object oriented data layer classes and relational database. Microsoft Visual Studio 2008 SP1 contains auxiliary tools for automatic generating .NET classes which represent object model for target database. Mapping between relational database and automatically generated object model is defined through Entity Data Model (EDM) which uses XML syntax. In addition, Entity Framework introduces an extension (Language Integrated Queries - LINQ) to .NET programming languages which enables defining database queries as a part of standard language constructs.

Business logic layer consists of set of classes implemented in C# programming language. These classes contain application specific logic:

- User authentication and checking user rights level;
- Managing configuration parameters for generating work time schedule;
- Correcting generated schedule;
- Creating appointments for patients;
- Creating appointments reports etc.

Presentation layer defines end user's view of the application. It was carefully designed in order to achieve uniform look and simplicity of use because it is intended to be used by medical personnel who, in general, have minimal knowledge of information technology. Because of this we combined well known mechanism of Windows forms and new libraries from Windows Presentation Foundation (WPF). More details on using WPF in Medis.NET information system can be found in [5].

Database scheme was designed keeping in mind configurability of the solution as a very important request in Medis.NET Scheduler. This request was fulfilled by storing all configuration parameters in the database. Tables relevant for Medis.NET Scheduler application are shown in Fig. 3. Central table in the diagram is *employee* and it contains data about personnel (both medical and non-medical) in a healthcare institution. The rest of the tables can be divided into two groups – tables related to patient appointments (*appointment*, *specialization*, *employee_specialization*) and tables related to work time scheduling (*work_scheme*, *employee_work_scheme*, *work_scheme_shift*, *shift*, *employee_shift*, *break*, *absence_type*, *absence*). Some more explanations of these tables are listed here:

- *appointment* – contains data about appointed time periods and data about eventually canceled appointments;
- *specialization* – contains data about doctors' specializations and planned values of patient examination times for each specialization;
- *employee_specialization* – contains relations between tables *employee* and *specialization*;
- *work_scheme* – contains data about work schemes (work patterns);
- *employee_work_scheme* – connects employees with their work schemes;
- *work_scheme_shift* – connects work scheme with shifts which are included in the work scheme and contains frequency of changing the shift with other one;

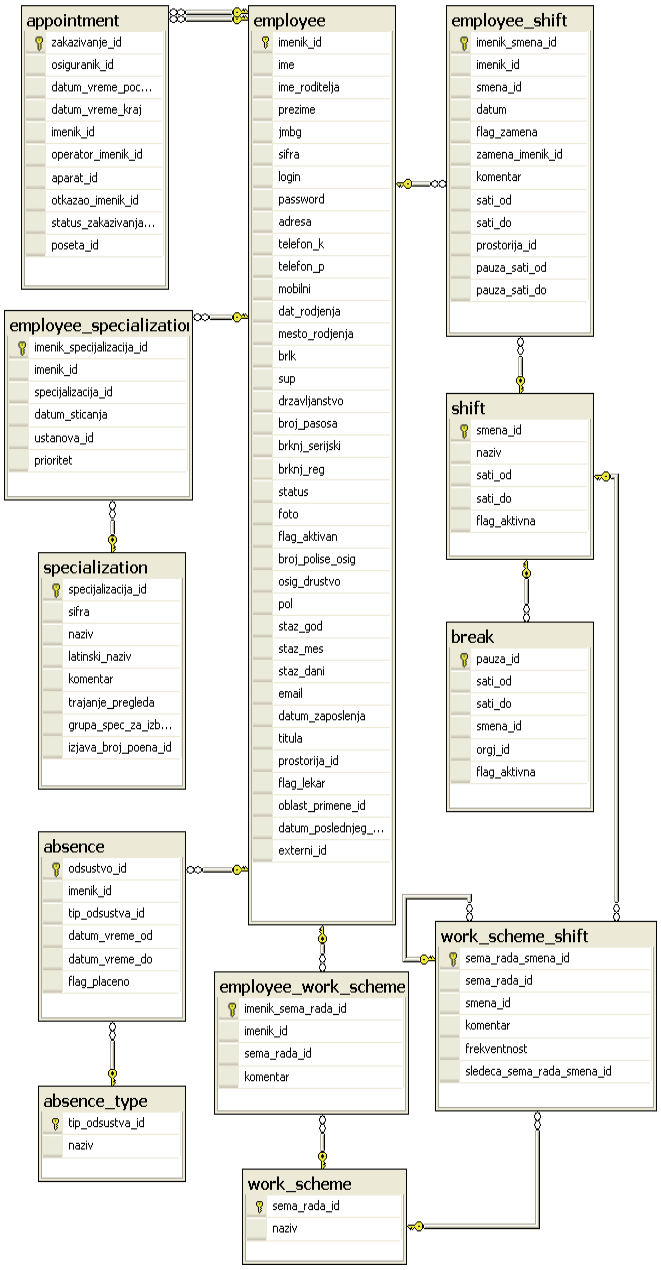


Fig. 3. Database diagram

- *shift* – contains data about work shifts such as start and end time;
- *employee_shift* – connects employees with their shifts, redundant table containing one record for employee per day which enables making exceptions from generated schedule;
- *break* – contains data about breaks during work time for each shift and organization unit;
- *absence_type* – contains data about different absence types such as holidays, sick leaves etc;
- *absence* – connects employees with corresponding absence types for certain time periods.

IV. Conclusion

In this paper we presented a solution for scheduling employees work time and patients appointments which is a part of a medical information system. By using latest technology achievements it aims to make efficient utilization of medical personnel resources and comprehensive scheduling of patients' appointments at The Health Care Center in Niš. Nevertheless there is still significant room for improvements in this area. Our future research will be targeted towards optimizing utilization of available healthcare equipment by involving various optimization techniques [6] in our software and improvement of the set of offered search criteria requirements for empty terms

Acknowledgement

Presented solutions for medical information systems are results of projects supported and funded by the Ministry of Science and Technological Development of the Republic of Serbia (Project Nr TR13015).

References

[1] D. Gupta, B. Denton, "Appointment scheduling in health care: Challenges and opportunities", IEEE Transactions, vol. 40, no. 9, pp. 800-819, 2008.

[2] S. N. Ogulata, M. Koyuncu, E. Karakas, "Personnel and Patient Scheduling in the High Demanded Hospital Services: A Case Study in the Physiotherapy Service", Journal of Medical Systems, vol. 32, no. 3, pp. 221-228, 2008.

[3] P. Rajković, D. Janković, V. Tošić, "A Software Solution for Ambulatory Healthcare Facilities in the Republic of Serbia", Healthcom 2009, Conference Proceedings, pp.161-168, Sydney, Australia, 2009.

[4] A. Adya, J. Blakeley, S. Melnik, S. Muralidhar, "Anatomy of the ADO.NET Entity Framework", ACM SIGMOD International Conference on Management of Data, Conference Proceedings, pp. 877-888, Beijing, China, 2007.

[5] I. Marković, S. Pešić, D. Janković, "Using XAML in Representation of Dental Charts in Electronic Health Record", ICT Innovations 2009, Conference Proceedings, pp. 247-256, Ohrid, Macedonia, 2009.

[6] A.T. Ernst, H. Jiang, M. Krishnamoorthy, D. Sier, "Staff scheduling and rostering: A review of applications, methods and models", European Journal of Operational Research, vol. 153, no. 1, pp. 3-27, 2004.

SESSION CSIT III

Computer Systems and Internet Technologies III

Software Tool for Random-Based Generation of Switching Function Benchmarks

Miloš M. Radmanović[1]

***Abstract* – Traditional approach to the measurement of performance for CAD algorithms involve the use of sets of so-called "benchmark functions". This paper describes a system (software tool), that generates and manipulates random constraint two-level representation of a Boolean function corresponding to the so called "real-world" functions. The software tool satisfies both the requirements to build some common benchmarks useful to compare different research results, and to create a tool for supporting intensive test of new algorithms. The paper gives an overview of the functionalities of the software tool and describes parameters that characterize the layout of a function. Tool is made publicity available in an attempt to extend standard sets of benchmark functions.**

***Keywords* – Synthetic random benchmark generator, switching functions, Berkeley PLA format, software tool.**

I. INTRODUCTION

Benchmark functions design is the basis for the performance evaluation of today's CAD algorithms. The most commonly method to verify the competitiveness of a new algorithm consists of applying this algorithm to a set of benchmark functions design in a given experimental settings. The experimental results are then compared to those obtained by applying a comparable algorithm to the same set of benchmarks design. This implies that the quality of CAD algorithm's evaluation is only as good as the functions design is used for benchmarking.

In recent years, there have been several initiatives for assembling benchmark design suits consisting of two-level representation of Boolean functions to improve the quality of CAD algorithms. An extensive research on benchmarks was conducted in both academia and industry. Industry use real customer designs as benchmarks to demonstrate performance of their products over their competition. However, customer designs are usually confidential and provided under non-disclosure agreements. Most of the benchmarks in academia originate from conferences and workshops. For instance MCNC.91 [5] and IWLS.05 [6] were published for workshops on Logic Synthesis. The MCNC.91 benchmark suite has standardized libraries with representative circuit designs ranging from simple circuits to advanced circuits obtained from industry. MCNC benchmarks are very popular in academic research. The IWLS.05 contains diverse circuit designs derived from past conference benchmarks, open source community of hardware designers, and industry to represent a variety of applications. The benchmarks were synthesized and organized into a standardized library with a common timing infrastructure, standard interfaces and reporting formats to promote easy exchange of benchmarks and experimental results in the community. Conference benchmarks are widely circulated because they are freely available in public domains. Open source communities allow members to share benchmarks, methodologies, and results. For instance OpenCores [7] is an open source community that deals with semiconductor intellectual property cores. Academia and industrial corporations use freely available OpenCores designs to benchmark their products.

In some cases, benchmark designers turn to automatic generation of synthetic benchmarks to evaluate new architectures which they can't efficiently test using existing benchmarks [10]. The paper [8] presents an approach to generate synthetic benchmarks for evaluating new architectures and tools, which don't have representative evaluation benchmark sets. For instance, the paper [9] implemented an algorithm for generating synthetic benchmarks and used them to study optimality and scalability of placer tools.

The problem of a correct experimental evaluation is crucial in several areas of Logic synthesis, in particular when a computationally intensive test of specific algorithms is needed. This paper describes a software tool (synthetic benchmark generator), that generates and manipulates random constraint two-level representation of a Boolean function corresponding to the so called "real-world" functions. My attention to the problem generates from previous work on the evaluation of the specialized algorithms for calculation over decision diagrams [1], [2]. This previous work has shown that evaluation of some algorithms needs for reliable benchmarks, like a library of similar problems having different characteristics. In particular, I address the problem of defining a set of random-based two-level representation of a Boolean function according to a number of well designed parameters and their management by a software tool. Software tool generates files in widely used Berkeley PLA (Espresso) format [13] for two-level representation of Boolean function. Tool is made publicity available [15] in an attempt to extend standardized libraries of benchmark functions.

There are many different approaches to generate random-based benchmarks [3], [4], [8], [11], [12]. When in need of random-based two-level representation of a Boolean function benchmarks some approach can be followed. The first approach consists of random generation of two-level representation of a Boolean function with the probability of appearance for each element of the representation. The second approach consists of random appearance of the basic logic

[1]Miloš M. Radmanović is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: milos.radmanovic@gmail.com

operation sequences and the third approach consists of random appearance of the arithmetic logic unit operation sequences.

This paper is organized as follows: Section 2 shortly introduces the Berkeley PLA (Espresso) format for two-level representation of Boolean function; Section 3 presents the technical details of the random-based benchmarks generator proposed in this paper; Section 4 describes the software tool capabilities to offer an experimental workbench and gives some examples of using generated benchmarks. Some concluding remarks end the paper.

II. BERKLEY PLA FORMAT

The Berkley PLA (Espresso) format is a logical representation of a set of Boolean equation. The format has been expanded to allow for multiple-valued logic functions, and to allow for the specification of the don't-care set. Programs exist to translate a set of equations into this format (e.g., eqntott, bdsyn, eqntopla). PLA format is described as a character matrix with keywords embedded in the file to specify the size of the matrix and the logical format of the function.

The minimum required set of keywords is: *.i* (specifies the number of input variables) and *.o* (specifies the number of output variables) for binary-valued functions, or *.mv* for multiple-valued functions. A complete list of the keywords is given in [13].

The default PLA file formats are compatible with the Berkeley standard format for the physical description of a PLA (Programmable logic array). It is generally assumed that the PLA format is specified such that each row of the PLA fits on a single line in the file. A term is represented by a "cube" which can be considered either a compact representation of an algebraic product term which implies the function value is a 1, or as a representation of a row in a PLA which implements the term. A cube has an input part which corresponds to the input plane of a PLA, and an output part which corresponds to the output plane of a PLA. Each position in the input plane corresponds to an input variable where a *'0'* implies the corresponding input literal appears complemented in the product term, a *'1'* implies the input literal appears uncomplemented in the product term, and '-' implies the input literal does not appear in the product term. For each output, a literal *'1'* means that this product term belongs to the ON-set, a *'0'* means that this product term belongs to the OFF-set, a '−' means that this product term belongs to the DC-set, and a '~' implies this product term has no meaning for the value of this function. The ON-set of a Boolean function is defined as the set of minterms for which function value is a 1. The OFF-set of a Boolean function is defined as the set of minterms for which function value is a 0. The DC-set (don't care set) is defined as the set of minterms for which the function value is unspecified. A function is completely described by providing its ON-set, OFF-set and DC-set. A complete description of PLA format is given in [13].

The Berkley PLA format representation is illustrated in Figure 1 by a two-input adder.

```
.i 4
.o 3
.p 11
100- 010
0101 010
001- 010
1-00 010
0-10 010
1111 010
-111 100
11-1 100
-1-0 001
-0-1 001
1-1- 100
.e
```

Fig. 1. Example of Berkley PLA format

III. SOFTWARE TOOL

In this section I described the basic ideas followed to generate random two-level representation of Boolean function based on Berkley PLA format. The procedures build two-level representation by using a set of parameters to control the structure of the cubes.

The first generation algorithms controls nine two-level representation's characteristics: (1) number of input variables, (2) number of output variables, (3) number of cubes, (4) probability of appearance for input literal '0', (5) probability of appearance for input literal '1', (6) probability of appearance for input literal '-', (7) probability of appearance for output literal '0', (5) probability of appearance for output literal '1', (6) probability of appearance for output literal '~'. I believe that changing these characteristics it is possible to generate a significantly representative subset of similar two-level representation having different characteristics. The proposed generation algorithm is able to generate either a large benchmark sets with several levels of difficulty to test average performance of a reasoning algorithm or to create a particular instance to test such an algorithm in extreme situations.

The basic idea is to randomly map set of input literal '0' on set of input part of cube. The random mapping is controlled by literal '0' density (probability of appearance). The other aspects to control random mapping are input literal '1', input literal '-', output literal '0', output literal '1' and output literal '~' density.

The actual generating algorithm can produce only synthetic benchmark functions not corresponding to the so called "real-world" functions. I am now modifying it for producing also benchmark functions corresponding to the "real-world" functions.

The second and the third generation algorithms controls six two-level representation's characteristics: (1) number of input variables, (2) number of output variables, (3) number of cubes, (4) probability of appearance for input literal '0', (5) probability of appearance for input literal '1', (6) layout of output ON-set and OFF-set.

The second generation algorithm proceeds similarly to the first generation algorithm with two notable differences. First, I don't use random mapping of input literal '-'. Second, ON-set and OFF-set of a Boolean function is defined as a basic logic

operation (AND, OR, NOT, NAND, NOR, EXOR and EXNOR) sequence.

In order to guarantee the "real-world" benchmark functions, the third generation algorithm extends the second generation algorithms by defining ON-set and OFF-set as a arithmetic-logical operation (adder, multiplier) sequence.

To exploit the capabilities of generating random two-level representation of Boolean function, I have inserted it in a software tool (synthetic benchmark generator) that allows a flexible interaction between a user and generation algorithms.

Software tool is written in MS Visual C++ and use MFC technology [14]. It consists of two basic modules: (1) Random two-level representation generator and (2) Interaction module that allow user to interact with two-level representation in Berkley PLA format and to control representation's characteristics as previously described. It is to be noted constant possibility of naming and saving a current generated function, of choosing particular generating algorithm to be used and of additional editing a two-level representation in PLA format.

IV. Using the Generated Benchmarks

Binary Decision Diagrams (BDDs) are data structures convenient for representation of discrete functions. BDDs are derived by the reduction of the corresponding binary decision trees (BDTs). The reduction is performed by sharing the isomorphic subtrees and deleting the redundant information in the BDT using the suitably defined reduction rules. BDDs are often substantially more compact than traditional normal forms such as conjunctive normal form and disjunctive normal form. They can also be manipulated very efficiently. Hence, BDDs have become widely used for a variety of CAD applications, including symbolic simulation, verification of combinational logic and verification of sequential circuits.

Multiple-output switching functions are represented by shared BDDs (SBDDs) [16] having a separate root node for each output. Thus, SBDDs are obtained by sharing isomorphic subtrees in BDDs for outputs of function, considered as separate particular switching functions.

I now show examples of ability of creating subset of similar two-level representation having different characteristics. Below I give tables of different time and space SBDD testing statistics with similar generated two-level representations of Boolean functions. I performed the testing on a PC Pentium IV on 2,66 GHz with 4 GB of RAM (MS Windows 7 Ultimate). The memory usage for all tests was limited to 2 GB.

It is now possible to show the effectiveness of the work perform by the generation algorithms.

Table 1 describes SBDD time and space statistics using generated benchmark based on the first algorithm controlled by changing the number of inputs and the density of input literal '0'. In most cases it is shown that low density of input literal '0' in two level representation of Boolean function produce more SBDD nodes. Differences in performances between functions with different number of inputs can be evaluated from generated benchmarks. Several classes of functions can be compared under control of some parameters.

Table 2 describes SBDD time and space statistics using generated benchmark based on the first algorithm controlled by changing the number of inputs and the density of output literal '1'. In most cases it is shown that medium density of output literal '1' in two level representation of Boolean function produce more SBDD nodes. Also, it is shown that extremely low and high density of literal '1' increase the size of SBDD.

Table 3 describes SBDD time and space statistics using generated benchmark based on the third algorithm controlled by changing number of inputs, outputs and cubes. In most cases it is shown that large number of cubes produces more SBDD nodes.

I encountered a number of factors that requires repetitions of experiments. The major factor is non trivial, unexpected and often unexplainable variability of results under slightly different parameters condition.

I just report the fact that first experimentation I am performing is quite satisfactory. It opens the possibility to create a number of random benchmarks with similar parameters and all different on random basis.

TABLE I
SBDD TIME AND SPACE STATISTIC USING GENERATED BENCHMARKS BASED ON FIRST ALGORITHM (DENSITY OF INPUT LITERAL '0')

Function name	inputs/outputs/cubes /density of inp. literal '0'	time [s]	nodes
rnd0_01	75/100/150/5	53.24	1635332
rnd0_02	75/100/150/15	2.94	183136
rnd0_03	75/100/150/25	1.11	71960
rnd0_04	75/100/150/35	0.68	39766
rnd0_05	75/100/150/45	0.54	37683
rnd0_06	75/100/150/55	0.5	32185
rnd0_07	75/100/150/65	0.42	26346
rnd0_08	75/100/150/75	0.4	26755
rnd0_09	75/100/150/85	0.53	28517
rnd0_10	75/100/150/95	0.61	24397
rnd0_11	100/100/150/5	104.02	2352778
rnd0_12	100/100/150/15	3.66	214117
rnd0_13	100/100/150/25	1.24	73326
rnd0_14	100/100/150/35	0.74	46874
rnd0_15	100/100/150/45	0.48	33324
rnd0_16	100/100/150/55	0.46	30385
rnd0_17	100/100/150/65	0.48	32200
rnd0_18	100/100/150/75	0.46	32505
rnd0_19	100/100/150/85	0.56	31984
rnd0_20	100/100/150/95	0.98	38543
rnd0_21	150/100/150/5	54.05	1785260
rnd0_22	150/100/150/15	3.08	181595
rnd0_23	150/100/150/25	1.32	76929
rnd0_24	150/100/150/35	0.67	44976
rnd0_25	150/100/150/45	0.65	43895
rnd0_26	150/100/150/55	0.49	37768
rnd0_27	150/100/150/65	0.64	41911
rnd0_28	150/100/150/75	0.54	38098
rnd0_30	150/100/150/85	0.62	42794
rnd0_31	150/100/150/95	1.35	62956

TABLE II
SBDD TIME AND SPACE STATISTIC USING GENERATED BENCHMARKS BASED ON FIRST ALGORITHM (DENSITY OF OUTPUT LITERAL "1")

Function name	inputs/outputs/cubes /density of inp. literal '0'	time [s]	nodes
rnd1_01	75/100/150/5	0.09	11282
rnd1_02	75/100/150/15	0.25	22808
rnd1_03	75/100/150/25	0.53	39326
rnd1_04	75/100/150/35	0.89	49733
rnd1_05	75/100/150/45	0.96	56972
rnd1_06	75/100/150/55	1.34	57661
rnd1_07	75/100/150/65	0.95	43908
rnd1_08	75/100/150/75	1.51	59644
rnd1_09	75/100/150/85	1.17	50242
rnd1_10	75/100/150/95	0.62	28826
rnd1_11	100/100/150/5	0.11	13891
rnd1_12	100/100/150/15	0.28	26655
rnd1_13	100/100/150/25	0.53	41679
rnd1_14	100/100/150/35	0.68	44775
rnd1_15	100/100/150/45	1.03	57787
rnd1_16	100/100/150/55	0.91	51122
rnd1_17	100/100/150/65	1.26	57370
rnd1_18	100/100/150/75	1.41	59082
rnd1_19	100/100/150/85	0.96	40281
rnd1_20	100/100/150/95	0.48	23023
rnd1_21	150/100/150/5	0.14	18338
rnd1_22	150/100/150/15	0.31	32156
rnd1_23	150/100/150/25	0.51	44581
rnd1_24	150/100/150/35	0.84	58650
rnd1_25	150/100/150/45	0.84	56385
rnd1_26	150/100/150/55	1.21	67633
rnd1_27	150/100/150/65	1.32	62448
rnd1_28	150/100/150/75	1.23	55816
rnd1_30	150/100/150/85	1.38	57329
rnd1_31	150/100/150/95	0.62	34542

TABLE III
SBDD TIME AND SPACE STATISTIC USING GENERATED BENCHMARKS BASED ON THIRD ALGORITHM

Function name	inputs/outputs/cubes	time [s]	nodes
rnd_mul32_01	64/64/250	0.29	17675
rnd_mul32_02	64/64/500	0.81	34727
rnd_mul32_03	64/64/750	1.59	50998
rnd_mul32_04	64/64/1000	2.74	67391
rnd_mul48_01	96/96/250	0.43	27642
rnd_mul48_02	96/96/500	1.27	54327
rnd_mul48_03	96/96/750	2.59	80574
rnd_mul48_04	96/96/1000	4.38	106761
rnd_mul60_01	120/120/250	0.57	35098
rnd_mul60_02	120/120/500	1.63	69059
rnd_mul60_03	120/120/750	3.35	102815
rnd_mul60_04	120/120/1000	5.85	135944

.

V. CONCLUSION

This paper describes a software tool that generates and manipulates random constraint two-level representation of a Boolean function in Berkley PLA format corresponding to "real-world" functions. The tool includes three approaches to build random two-level representation by using a set of parameters to control the structure of the cubes.

The tool satisfies both the requirements to build some common benchmarks useful to compare different research results, and to create a tool for supporting intensive test of new algorithms.

REFERENCES

[1] M. Radmanović, "Tools for Calculating Autocorrelation Spectrum by Using The Wiener-Khinchin Theorem", 39th Int. Conf . ICEST 2006, Sofia, 2006.

[2] M. Radmanović, R. Stanković, C. Moraga, "Analysis of Decision Diagram based Methods for the Calculation of the Dyadic Autocorrelation", International Journal of Systemics, Cybernetics and Informatics, Pentagram Research Publication, July 2007.

[3] S. Bhawmik, V.K. Narang and P. Pal Chaudhuri, "Selecting Test Methodologies for PLAs and Random Logic Modules in VLSI Circuits - An Expert System Approach", The VLSI Journal, Volume 7, Issue 3, Pages 267-281, September 1989.

[4] K. Iwama, S. Sawada, K. Hino, H. Kurokawa, "Random Benchmark Circuits with Controlled Attributes", In. Proc. of European Design and Test Conference 97, 90 – 97, 1997.

[5] F. Brglez, "ACM/SIGDA Benchmarks Electronic Newsletter", DAC 93 Edition. June 1993, 1-22, http://www.cbl.ncsu.edu/benchmarks.

[6] C. Albrecht, "IWLS 2005 Benchmarks," 2005. http://iwls.org/iwls2005/benchmark_presentation.pdf

[7] "OpenCores", www.opencores.org/.

[8] P. Verplaetse, J. Campenhout, and D. Stroobandt, "On Synthetic Bencmark Generation Methods," In. Proc. of IEEE In. Symposium on Circuits and Systems, vol. 4, pp. 213-6, 2000.

[9] C. Chang, J. Cong, M. Romesis, and M. Xie, "Optimality and Scalability Study of Existing Placement Algorithms," IEEE Transactions on CAD of Integrated Circuits and Systems, vol. 23, no. 4, pp. 537-49, 2004.

[10] R Njuguna, "A Survey of FPGA Benchmarks", Project Report, http://www.cs.wustl.edu/~jain/cse567-08/ftp/fpga/

[11] J. Darnauer and W. Dai, "A Method for Generating Random Circuits and Its Application to Routability Measurement", In. Proc. of the 1996 ACM 4. In. Symposium on FPGA, 66 – 72, 1996.

[12] J. Harlow and F. Brglez, "Design of Experiments for Evaluation of BDD Packages Using Controlled Circuit Mutations", In Proc. of the In. Conference on Formal Methods in CAD, 64-81, 1998.

[13] R. Rudell, "Espresso Misc. Reference Manual Pages", University of California, Berkeley, California, USA, http://embedded.eecs.berkeley.edu/pubs/downloads/espresso/index.htm

[14] Microsof Visual Studio 6.0, Microsoft Corporation, http://msdn.microsoft.com/en-us/library/ms950417.aspx

[15] Stgroup CIITLAB Software http://stgroup.elfak.edu.rs/home/node/4

[16] T. Sasao T, M. Fujita M, "Representations of discrete functions. Boston", Kluwer Academic Publishers, 1996.

Metadata Models for Technology Enhanced Learning in SINUS Project*

Kamenka Staykova[1] and Ivo Hristov[2]

***Abstract* – This paper presents the basic idea, functionalities and information structures of Technology Enhanced Learning Environment of the SINUS Project. The emphasis is put on the decisions taken to form the suitable metadata constructions supporting the processes of search, extraction and annotation of the used multimedia objects.**

***Keywords* – Technology Enhanced Learning, Metadata, Ontologies**

I. Introduction

Current approaches to Technology Enhanced Learning (TEL) are fundamentally based on providing a learner with appropriate learning content. In the current Learning Management Systems the allocation of learning resources is done predominantly at design-time and there are not many capabilities to reuse existing blocks of information objects. Authoring of adaptive content, realized partially in the so-called Learning Content Management Systems, is one of the most important and labor-intensive activities in the modern TEL practice. Traditionally authoring of adaptive content relies on the design of a fine-grained domain model and careful indexing of various Learning Objects (LO) with multiple domain concepts targeting re-usability and re-purposing of information object in different learning situations or learning environments.

SINUS Project is a 3-years long research project of the National Science Fund of Bulgaria, which mission is to demonstrate how the learning process could be supported and facilitated by the modern technologies of Semantic Web Services (SWS). The major approach of the SINUS Project is based on analyzing and exploiting the advantages of SWS technology in the automation of Learning Objects discovery, selection and composition within one distributed service architecture seamlessly integrated through ontologies [1]. The project architecture defines number of information objects stored into set of repositories that are accessed via http protocols as XML Web services. The information objects have additional semantic annotation that enables development of the system in order to compose learning environment personalized in some aspects.

The paper is organized as follows: The next section presents SINUS Project giving short overview of project goals, description of Use-Case scenarios, terminology and some architectural views. The third section presents the abstract annotation levels and metadata models in SINUS Project; each level is described in respect to decisions taken about metadata organization and its connection to some standards taken into account. The last section is a short conclusion.

II. Technology Enhanced Learning Environment of the SINUS Project

A. Basic Idea

SINUS project follows the idea to increase the effectiveness of preparation and use of adaptable learning content and to change the data and metadata-based paradigm to a dynamic service-oriented approach. The learner is provided with a dynamic supply of appropriate functionalities, in order to enable a dynamic adaptation to the learning context at runtime of a learning process, including dynamic access to different multimedia sources. The learning content creators are supported during the creation of learning materials and procedures giving assistance to learners. In SINUS Project Learning Objects are based on multimedia objects. The needed metadata to organize the work of SWS and access to Learning Objects is based on international metadata standards and domain ontologies. SINUS Project goes further with the lessons learnt from two IST FP-6 European research projects: INFRAWEBS and LOGOS, and the ambition of the team is to be developed more mass-user-centered and user-friendly approach and tools.

B. Use-Cases

The learning domain chosen for demonstrations of SINUS platform functionalities is Bulgarian Iconography. It presents a fruitful field to demonstrate how the digital photos of iconographic works (icons, wall-paintings, etc.) could be used during the learning process. The multimedia resources of SINUS Use-Cases are digital photos. The Multimedia Digital Library "Virtual Encyclopedia of East-Christian Art" [2] is the source of multimedia content for learning scenarios.

[1]Kamenka Staykova is with the Institute of Information Technologies, Bulgarian Academy of Sciences, Acad. Bonchev str. bl.2, 1311 Sofia, Bulgaria, E-mail: staykova@iinf.bas.bg

[2]Ivo Hristov is with the Institute of Information Technologies, Bulgarian Academy of Sciences, Acad. Bonchev str. bl.2, 1311 Sofia, Bulgaria, E-mail: collin@iinf.bas.bg

*The paper is supported by the Research Project SINUS: Semantic Technologies for Web Services and Technology Enhanced Learning, funded by the National Science Fund of Bulgaria No. D-002-189.

The learning domain of Bulgarian Iconography is potential field of interest for people with wide range of learning demands: from formal and specialized professional education to self-training or personal cultural investigations. The SINUS Use-Cases are directed primary to formally organized students of Art disciplines, History, Theology. There are two general user roles round which the Use Cases are build: The Educationalist and the Learner.

The Educationalist prepares learning tasks for her/his students by means of the environment. The Educationalist creates sets if instructions associated to each particular learning task and prepares or recommends the multimedia resources to be used in "solving" the learning task. Usually the learning task contains two tightly connected parts: 1) to be gathered a collection of multimedia resources (digital pictures of iconographic works) and 2) the collection to be analyzed or commented in some specific aspects, which concern the learning (art specifics, historical aspects, theological aspects, etc.). The solution of such a learning task is a multimedia document containing the digital images of prepared collection and text, comments of the author-learner. The created multimedia document is called project in SINUS platform terminology.

The Lerner receives a particular learning task and some instructions. S/he is expected to collect appropriate multimedia resources and to prepare analyses in multimedia document: project. At each single step from the receiving of the learning task, through the access to different multimedia sources of information and visual resources, to the preparation of the project the Learner is supported by the functionalities of SINUS platform.

C. Overview of Conceptual Architecture of SINUS Platform

The conceptual architecture of SINUS platform consists of two main elements: Design-time Environment and Run-time Environment.

The Design-time Environment proposes:

- Information structures for storing and retrieving semantic and non-semantic data:
  - Semantic Web Service Repository enables efficient storage and retrieval of all elements of the Semantic Web: goals, ontologies and Semantic Web Services.
  - Learning Metadata Repository contains annotated metadata about Learning Objects, used for grounding of SWS.
  - Information Indexer and Retriever contain a special representation of both Semantic Objects and Learning Objects.
- Tools for creation and maintenance of resources, metadata and supporting ontologies:
  - Semantic Web Service Creator aims at designing Semantic Web Services by reusing already existing semantic and non-semantic descriptions stored in the Semantic and Learning Metadata Repositories.
  - Ontology Creator aims at creating ontologies in a user-friendly manner.
  - Semantic Goal Creator provides means for creation of SWS-based reusable goals for designing TEL applications.
  - Learning Content Description Tool aims at creating metadata annotations and consists of Digital Objects Annotation Package, Semantic Annotations Package and Learning Objects Annotation Package (see Fig. 1 below). Multimedia and Learning Objects are annotated according to certain TEL standards and some formal ontologies.
- Methods used for creating and maintaining Semantic and Learning Objects:
  - Combination of TEL-specific and logic-based methods for object discovery
  - TEL-specific decision-support methods for dynamic service composition
  - Several methods for calculating similarity and object retrieval – structural, linguistic, statistical, fuzzy, etc.

The Run-time Environment is responsible for discovery, dynamic composition, execution and monitoring of SWS. The Run-time Environment consists of the SWS Discoverer, SWS Composer and SWS Executor components.

SINUS conceptual architecture is an adaptation of the INFRAWEBS SWS-architecture towards Technology Enhanced Learning applications and presents a novel approach for creating and maintaining Semantic Web services and SWS applications. It is based on tight integration of similarity-based and logic-based reasoning. Similarity-based reasoning is used for fast finding of approximate solutions, which are further concretized by the logic-based reasoning.

One of the objectives of SINUS Project is the development of new methods and tools for creation and semantic annotation of Learning Objects compatible with Semantic Web Services methodology. The investigation aims at design of activity scheme and information models for dynamic creation and adaptation of learning objects (multimedia objects, annotated with content- and context-oriented metadata), facilitating their reusability and addressing active authoring as learning activity. The research addresses the limitations of standard LO metadata through the usage of ontologies to represent the knowledge encoded in the metadata in machine readable forms. The information models and methods for learning content organization are based on pragmatic restrictions and modifications of the project LOGOS information models (see III.B below). This is a step to orienting the information models and tools to mass authors of learning materials without need of specific knowledge and skills concerning information models and basics of ontology engineering.

III. METADATA MODELS OF MULTIMEDIA OBJECTS WHITHIN SINUS

SINUS environment provides its users with effective methods to operate with different multimedia objects during the working process. Presented above Use-Case of working within the area of Bulgarian Iconography illustrates the processes of search, extraction and annotation of multimedia objects. The architecture aims the following annotation levels

(see Fig. 1): Digital Libraries level, Semantic Objects level and Learning Objects level. Each one of the levels has three major elements:

- Repository – The repository works as storage space for different types of objects. A repository is accessible via http protocol and presents the functionality as web services.
- Specific annotation package – The annotation package is a set of software tools for managing descriptions specific for the level and according to particular standard.
- User role (annotators) – a group of users dedicated to work with the objects on each level. The groups of annotators are defined on the basis of functional requirements for given set of logically related operations on the objects [3]. These sets of operations specify the annotation levels that have the similar professional qualification requirements from the user roles that perform them.

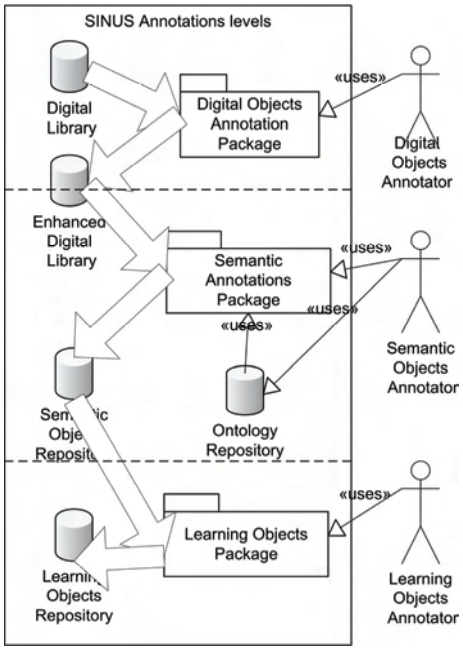


Fig. 1 Annotation levels

A. Level of Digital Libraries

The digital libraries level is dedicated for storage, management and editing of digital objects. The objects of digital library are not intended to have any educational purposes. The digital library is pre-existing feature with specific formats and architecture. The only requirement for the library is to expose its objects as well known XML Web Services. The role of digital object annotator on the Fig. 1 is to insert additional metadata in the cases when descriptions provided by the library are not sufficient. Annotations are stored into extended digital library. The digital object annotation package has to be able to insert annotation at least according to Dublin Core Standard [4]. The approach of inserting an annotation according only to Dublin Core Standard has limited capacity of description of video clips. For sustainable and yet flexible description of digital objects it is reasonable the usage of public standard for metadata description like Moving Picture Experts Group (MPEG 7). According to [5] the MPEG standard allows different operations on images and video like describing objects' movement, camera motion, relations between objects, returning a list of videos with similar or dissimilar temporal and spatial elations, describing actions of given video content and preparing a list of videos where similar actions happen.

Concerning the metadata model for described Use-Case in SINUS Project we intend to use the fields of Dublin Core Standard at this level.

B. Level of Semantic Objects

Semantic Objects' level is the essential layer in SINUS metadata architecture. It gives flexibility in defining possible usage and re-use of multimedia objects, which could be described in different contexts according to different ontologies. Currently ontologies are the constructions which represent the semantics of human realm in form usable for software programs. The definition given in [6] is widely accepted and says that ontologies are "formal specifications of a shared conceptualization of a certain domain". This understanding is the ground of the rapidly growing interest to implemented ontologies. Creating new ontology starts from knowledge elicitation from domain experts, this is knowledge level; then working on symbol level the knowledge engineers formalize this knowledge; then at the implementation level the ontological construction is implemented in particular formalism or language (Frame Logics, Conceptual Graphs, OIL, OWL, etc.).

With the increasing number of applications using ontologies the problem with ontologies interoperability arises. It is in fact a problem of knowledge exchange among different *worlds*, which are introduces by different ontologies. In [7] four different general types of ontologies are defined. A *Top-level Ontology* describes very general concepts like space, time, event, which are independent of a particular problem or domain. A *Domain Ontology* describes the vocabulary related to a generic domain by specializing the concepts introduced in the *Top-level Ontology*. A *Task Ontology* describes the vocabulary related to a generic task or activity by specializing the *Top-level Ontologies*. Concepts in an *Application Ontology* often correspond to roles played by domain actors while performing a certain activity. Usually all the four types of ontologies are involved in real complicated tasks and platforms like in support of Technology Enhanced Learning proposed in SINUS Project.

Here we are concentrated only on Domain Ontology and metadata constructions supported by it at the level of Sematic Objects in SINUS platform. Semantic metadata of multimedia object in terms of SINUS Use-Case is tightly connected to the learning domain. Terminology of the learning domain Bulgarian Iconography is accessible trough the ontology "Bulgarian Iconographical Objects" [8], OBIO for short. OBIO is first created under the project LOGOS **"**Knowledge-on-Demand for Ubiquitous Learning" [11] and now enhanced.

In all its variations OBIO could be asserted to be a specialization and uses as Top-Level Ontology the CIDOC Conceptual Reference Model [9], which is an ICO standard since 2005 for organizing documentation of libraries preserving art collections and artifacts of world historical heritage. For the purposes of SINUS Project the ontology is implemented in OWL, it consists of 48 classes with 39 different properties. Semantic metadata of Use-Case multimedia objects in SINUS platform is expressed by ontological constructs of OBIO.

There are two cases with semantic metadata when the user of SINUS platform is searching for appropriate multimedia object. The first case is all the semantic features of the object to be supported by the Digital Library where this object is looked for. The second case is when the user needs to access some features of the multimedia object, which are not operating features of objects for the searching machine of the Digital Library. In the first case the record of semantic metadata of multimedia object is completely fulfilled with information received by SWS from the Enhanced Digital Library. In the second case the SINUS platform supports additional metadata of user interest in the Semantic Object Repository. In fact, for the user is not recognizable how the metadata of particular multimedia object is accessed and stored, the user works on the level of SINUS platform and uses all the functionalities there: access to the Digital Library, Semantic Annotations Package, access to the Domain Ontology, etc.

C. Level of Learning Objects

The learning objects layer is dedicated to handle the educational aspects of previously created semantic object. Educational aspects of semantic objects are made by enrichment of the semantic objects with Learning Object Metadata (LOM) according to [10] standard. The LOM enable search, educational annotation, and classification of semantic objects different by seize, purpose and granularity. The standard is widely used and accepted. The learning objects annotator is an Educationalist. This is user role in SINUS platform which is responsible to manage educational usage of the objects. The learning objects package has to be able to reference an object on the semantic level and to support the educational annotations with LOM. The educational annotations according to LOM standard are stored into the Learning Objects Repository.

IV. CONCLUSION

SINUS Project and TEL in general provide opportunity for ambitious and interesting research, especially concerning metadata models, which support annotations of multimedia and Learning Objects. In this paper were presented the decisions in organizing multilevel architecture for effective operating with metadata in SINUS Project. Further challenges in front of the team are grounding Semantic Web Services, interoperability of different ontologies, problems concerning annotation, access and extraction of multimedia objects.

REFERENCES

[1] D. Dochev, G. Agre, "Towards Semantic Web Enhanced Learning", KMIS 2009, Conference Proceeding, pp. 212-217, Madeira, Portugal, 2009.

[2] L. Pavlova-Draganova, V. Georgiev, L. Draganov, "Virtual Encyclopaedia of Bulgarian Iconography", Information Technologies and Knowledge, vol.1, №3, pp. 267-271, 2007.

[3] I. Hristov, "Information Objects in E-learning Authoring Studio Architecture - an overview", IIT/WP-254, 2009

[4] D. Hillmann, "Dublin Core Metadata Initiative- Using Dublin Core", 2005, http://dublincore.org/documents/usageguide/, (Last accessed April 2010)

[5] T. Sikora, "The MPEG-7 visual standard for content description—An overview", IEEE Trans. Circuits Syst. Video Technol., vol. 11, pp.696–700, 2001.

[6] T.R. Gruber, "Toward Principles for the Design of Ontologies Used for Knowledge Sharing", Formal Analysis in Conceptual Analysis and Knowledge Representation, Kluwer, 1993.

[7] N. Guarino, C. Masolo, G. Vetere, "OntoSeek: Content-Based Access to the Web", IEEE Intelligent Systems, 14(3), pp.70--80, 1999.

[8] K. Staykova, D. Dochev, "Ontology Bulgarian Iconographical Objects- Creation and Experimental Use", Cybernetics and Information Technologies, Vol.9, №1, pp. 25-37, 2009, http://www.cit.iit.bas.bg/CIT_09/v9-1/25-36.pdf, (Last accessed April 2010)

[9] N. Crofts, M. Doerr, T. Gill, S. Stead, M. Stiff, "Definition of the CIDOC Conceptual Reference Model, version 5.0.2", 2010, http://cidoc.ics.forth.gr/official_release_cidoc.html

[10] IEEE LOM, Draft Standard for Learning Object Metadata, http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf, 2002, (Last accessed April 2010)

[11] Deliverable D3 "LOGOS Project Platform", Project LOGOS Knowledge-on-demand for Ubiquitous Learning, http://logosproject.com, (Last accessed April 2010)

Towards Estimation of Reliability for Embedded Software Systems

Aleksandar Dimov[1]

***Abstract* – Software reliability is a critical concern to be taken into account when reasoning about embedded software systems. As in many other areas in software engineering, architectural approach towards estimation of reliability has many advantages, especially when applied in early development phases. This paper presents a review of state of the art in architectural based reliability models applicable for embedded software systems and points out the unresolved issues with respect to modelling of such systems.**

***Keywords* – Software reliability, Architecture-based software reliability models, Uncertainty**

I. INTRODUCTION

In recent years there is a trend towards increasing usage of embedded software systems in all areas of human life. Additionally, some of these systems are responsible for the control of different safety-critical processes, for instance in space, nuclear or transportation systems, which have high requirement toward their non-functional characteristics. Non-functional characteristics are also known as quality parameters and define additional constraints and requirements on how software should perform its functionality. In such conditions of paramount importance is to provide methods modeling and reasoning about non-functional properties in order to be able to adequately design safety-critical software systems.

One significant quality parameter is *dependability* [[2]], which is defined as the ability of a computing system is to deliver services that can justifiably be trusted. Dependability is characterized by several attributes, such as reliability, availability, safety[1], confidentiality, integrity and maintainability. In this paper we are going to look at reliability only and the methods for estimation of software reliability. Reliability is defined as the continuity of correct service, i.e. the belief that a software system will behave as per specification over a given period of time and is usually modeled as a stochastic value. It may have different measures like: probability of failure; mean time between system failures or failure rate.

Basically, there exist two broad categories of reliability assessment models: black-box and white-box models. Black-box models are used to reason about reliability of software systems, without taking into account their internal processes or structure. On the contrary, white box-models consider some internal information about architecture of the system. White box models are also called Architecture-Based Reliability Models (ABRMs). Usually architecture-based software reliability estimation takes the following main steps [[8]]: (1) Identification of computational modules (components)[2] within software architecture; (2) Description of the actual architectural model – this includes how components are interconnected and interact with each other (3) Definition of components failure behaviour – at this step the reliability parameters of components and their measures are identified and (4) Combination of the failure behaviour with the architectural model.

Application of white box models has a lot of advantages, among them are: ability to reuse information about reliability parameters of both the system and the components that constitute it; ability to find these modules that influence systems reliability the most, i.e; possibility to isolate and remove reliability "bottlenecks" within the system and etc. For these reasons we focus our research work on white box models.

Although significant amount of research has been undertaken in recent years in the area of architectural software reliability models, there exists a lot of work to be done in the area [[4]]. One very important such issue which exists in the field is *uncertainty* and more than 30 years of research failed to find universally accepted solution on how to model uncertainty.

Our previous research work [[3]] assessed white-box software reliability models upon several important issues that should be taken into account when analyzing embedded and safety-critical software system. In this paper we are continuing this research by adding one more issue – the uncertainty, inherent in reliability parameters. Uncertainty may exist either on individual components reliabilities or in their operational profile and this way it propagates to the calculated reliability estimation of the overall system. Uncertainties are very common mainly due to limitations and assumptions taken by different techniques adopted for reliability estimation.

A reliability model for embedded software systems should be able to take into account the following issues:

- Modeling of dependency between component failures – most of the models assume that a failure in one component never result from failures in another components.

[1]Aleksandar Dimov, PhD is an assistant professor at the Department of Software Technologies, Faculty of Mathematics and Informatics, Sofia University "St. Kliment Ohridski"; 125 Tsarigradsko Shosse Blvd., bl. 2, room 222, 1113, Sofia, Bulgaria e-mail: aldi@fmi.uni-sofia.bg

[2] We use the terms *module* and *component* as synonyms within this paper.

- Ability to estimate reliability parameters of components into the system – most models currently assume that reliabilities of modules are already known in advance.
- Ability to model uncertainty in system reliability.

The rest of the paper is organized as follows: Section two surveys the current state of the art in white-box reliability modeling with respect to applicability for embedded systems; Section 3 makes an analysis of these models and finally Section 4 concludes the paper and streamlines the directions for further research in the area.

II. How do we take into account uncertainty in reliability estimations?

In this section we are going to survey some the architecture based models that take into an account uncertainty in reliability parameter. According to the broad classification three main groups of white-box models are known: state based models, path-based models and additive models. State-based models use Finite State Machine (usually Markov chain) representation of systems' architecture. They attempt to take into account, all of the possible traces of components execution within the system. Path-based models are similar to state-based models, but consider only finite number of component executions traces. The latter usually correspond to system test cases. Additive models do not concern actual architectural configuration of the software system. Instead, they assume a specific distribution process for components failure behaviour and on that basis infer formulae for calculation of reliability.

Basically, there exist two main sources of uncertainty in reliability estimates calculated using architecture-based reliability models:

- Uncertainty in operational profile – operational profile is a frequency distribution that gives the relative probability that a specific function of the program will be executed. In other words, uncertainty in the operational profile represents uncertainty in the environment, in which a component is integrated and further executed.
- Uncertainty in reliability values – what is the level of confidence we have in the correctness of particular values of reliability

The most traditional approach with respect to the second step in white-box reliability modeling (i.e. creation of architectural model) is to present the architecture with the system as a Markov chain, where states of the chain represent system components, and the edges – transitions between the components. The first model to employ this approach is described in [17] and many of subsequent models are based on it.

Next subsections of the paper make a brief review of different models taking into account uncertainty. We also analyze ability of the models to deal with the other two issues for reliability estimation of embedded software systems as pointed out in the introduction. For ease of reference, we have used the name of the first author as the name of the model, even though these models are collective works of several researchers as indicated by the references. Models are presented in ascending alphabetical order, according to the name of the first author.

A. Adams model

Tom Adams actually presented an approach towards calculation of confidence intervals of reliability with respect to uncertainty caused by unknown operational profile [1]. This model does not take into account uncertainty in reliability parameters of modules within the system. Adams model does not take into account estimation of component reliability parameters nor dependency between component failures.

B. Gokhale model

This model presents reliabilities of components and operational profile not as point estimates but as stochastic values themselves. The key of the model is generation of analytical functions that give the mean and variance of component reliabilities and operational profile and consequent combination of these functions into calculation of the mean and variance of reliability of the whole system. This model also does not take into account estimation of component reliability parameters and dependency between component failures.

C. Goseva-Popstoyanova models [6], [7], [9]

Very similar to Gokhale's is a group of models resulting of research, carried out in West Virginia University, USA. Models presented in [7], [8] calculate system reliability as a random variable is proposed. Authors propose to estimate its value given that the parameters of the model (i.e. component reliabilities and operational profile transition probabilities) are also random variables. First of the two models is based on moments of random variables and the second – on Monte-Carlo simulation. The first one is based on presentation of the reliability in a Taylor series and approximating it by taking into account the first one or two terms in it. However, this approach does not allow taking into account uncertainty in the operational profile of components. Both these facts make it an approximate method for assessing system uncertainty. The method based on Monte-Carlo Simulation allows taking into account both uncertainty in component reliabilities and operational profile. Nevertheless, it assumes that the reliability and transition probabilities follow a given a priori (for instance Beta) distribution. Assumption about such distributions is also a possible source of uncertainty.

In *Kamavaram* model [9] entropy is used to quantify the uncertainty of system reliability, based on uncertainty in operational profile and uncertainty in component reliabilities. However, this method estimates only uncertainty and does not allow for estimation of a value for the system reliability itself. This is the reason that we do not distinguish it in our analysis.

Similarly to previous cases this group of model also takes component reliability parameters as given and does not take into account dependency between component failures.

D. Popic model

Reference [11] presents a model that extends Singh model (cf. subsection F.) and makes it capable to take into account error propagation from one component to another. This way it is able to model dependency between components. For this purpose this model takes into account the error propagation probability in system architectures. It represents the probability that an erroneous state, generated at one component will not be detected but will propagate to other components during system execution. As this is an extension of Singh model it also assumes existence of information about failure rates of components in the architecture.

E. Roshandel models

One of the earliest works taking into account uncertainty in operational profile of software components that build the software system is presented in [12].

Its approach towards the estimation of individual components reliability when both the implementation and the data for the operational profile are unknown. Component architecture is modeled with an extension of classical Markov chain, called Hidden Markov Model (HMM) [16]. This is the main point of ability of this to model uncertainty. In fact it models presents uncertainty of a transition between states of the chain. Transition probabilities of HMM are estimated by an iterative algorithm, which upon convergence provides optimal component reliabilities.

This model is further elaborated in [18], [19], where different views of software architecture, presenting systems structure and behavior are employed. Further, a so-called *Global Behavioral Model* is constructed. It represents the behavior of the software system as a function of the collective behavior of its constituent components. Behaviour of components is modeled using a set of concurrent state machines, which means that the current state of the system is presented with a set of component states. In the next step, a *Dynamic Bayesian Network* is used to estimate reliability of the system, combining the behavioural model with individual components' reliabilities. The approach is able to model a rich variety of uncertainties: uncertainty of individual component reliability values; uncertainty of system startup process, and uncertainties of human-system interactions. However, the model does not give any guidelines to quantitatively estimate uncertainty. Authors only make a sensitivity analysis of the model towards variations in reliability parameters of components and operational profile.

F. Singh Model

Singh model [13] is not specifically aimed at uncertainty, however similarly to Roshandel model it uses a Bayesian approach towards reliability calculation. Goal of this model is to be applicable early in the development process where reliability values of system modules and operational profile are not known. Nevertheless, the model assumes existence of information about failure rates of components in the architecture. This way, it estimates the mean and variance of software failure probability. However for this purpose authors also assume some given distribution (Beta distribution) of reliability parameters of individual components within the system, which as said above may also be a source of uncertainty. Singh model does not explicitly consider component failure dependence.

G. Zhang model [9]

This model upgrades the one presented in [20], which is essentially a path-based model, with ability to model component reliability as a function of transition probability to and from other components within the system. This way it becomes possible to take into account uncertainty due the environment in which a component is integrated and the systems operational profile. In fact, the empirical validation of the model shows that even when it is known in advance, operational profile of the system itself may introduce some uncertainty in the reliability of the components.

III. DISCUSSION AND ANALYSIS

In this section we analyze architecture-based software reliability models with respect to their suitability for embedded software systems. Table 1 shows a comparison of the models with respect to the three issues that should be addressed when estimating reliability of embedded software systems.

TABLE I

COMPARISON OF ARCHITECTURE-BASED RELIABILITY MODELS

	Modeling of uncertainty		**Dependency of component failures**	**Estimate reliability parameters of components**
	Op. profile	**Comp. reliability**		
Adams model	Yes	No	No	No
Gokhale model	Yes	Yes	No	No
Goseva moments model	No	Yes	No	No
Goseva-Monte-Carlo model	Yes	Yes	No	No
Popic model	Yes	Yes	Yes	No
Roshandel model	Yes	Yes	No	Yes
Singh Model	Yes	Yes	No	No
Zhang model	No	Yes	No	No

As seen from table one, none of the known models take into account all the identified issues for embedded systems. Dependency between component failures and estimation of individual component reliabilities are addressed by only one model each. Although uncertainty is a matter of broad research only one models (Adams) actually assume it is a

variance of the reliability estimation within some confidence interval. We believe that reliability should not be presented as a point value, but together with the variance in this value, i.e. with its confidence interval.

Additional issue is that empirical study of some of the presented models have shown that uncertainty in reliability estimation of the overall system is mostly influenced by the reliability of components that build the system reliability [21]. Usage profile, i.e. probability of transition of execution from one component to another has significantly less impact on system reliability. This way it is more important to be able to take into account uncertainty due to inaccurate component reliabilities than inaccurate usage profile.

IV. CONCLUSION

Reliability is a major concern to be taken into account when designing and implementing embedded software systems. Currently, there exist a lot of unresolved issues with architecture-based software reliability models and uncertainty in estimation is one of them. This paper presents a review of current state of the art in solving the issue of uncertainty.

Analysis of the models has shown that all of the important issues when modeling reliability of embedded systems need further research. In this respect our plans for further research include development of modeling framework that takes into account most of the issues listed in this paper.

ACKNOWLEDGEMENT

The work presented in this paper was partially supported by grants from the National Science Fund, part of the Ministry of Education and Science in Bulgaria under the PD01-0106 (ARECS) project.

REFERENCES

[1] Adams, T., Total Variance Approach to Software Reliability Estimation, IEEE Transactions on Software Engineering, vol. 22, no. 9, Sept. 1996, pp. 687-688.

[2] Avižienis, A., Laprie, J-C., Randell, B., Basic concepts and Taxonomy of dependable and secure computing, IEEE Trans on Dependable and Secure computing, Vol. 1, Issue 1, Jan -March 2004.

[3] Dimov, A. and Punnekkat, S.: "On the Estimation of Software Reliability of Component-Based Dependable Distributed Systems", In Proc. of the 1st International Conference on the Quality of Software Architectures (QoSA 2005), LNCS 3712, Springer-Verlag, September 2005, Erfurt, Germany, pp. 171-187.

[4] Gokhale, S., Architecture-Based Software Reliability Analysis: Overview and Limitations, In IEEE Transactions on Dependable Security Computing 4(1): 32-40 (2007).

[5] Gokhale, S., Quantifying the Variance in Application Reliability, in Proceedings of Pacific Rim Dependability Conf., Mar. 2004, pp. 113-121.

[6] Goseva-Popstojanova, K. and Kamavaram, S., Software reliability estimation under certainty: generalization of the method of moments, In Proc. of the 8th IEEE International Symposium on High Assurance Systems Engineering, 2004, March 2004, pp. 209-218.

[7] Goseva-Popstojanova, K. and Kamavaram, S., Assesing Uncertainty in Reliability of Component-Base Software Systems, 14th International Symposium on Software Reliability Engineering (ISSRE'03), 2003.

[8] Goseva-Popstojanova, K., Trivedi, K.S.: Architecture Based Approach to Reliability As-sessment of Software Systems, Performance Evaluation, Vol.45/2-3, June 2001

[9] Kamavaram, S., and Gosseva-Popstojanova, K., Entropy as a Measure of Uncertainty in Software Reliability, In Proceedings of the 13th International Symposium on Software Reliability Engineering, 2002 (ISSRE 2002), pp. 209-210.

[10] May, J., Testing the reliability of component-based safety critical software, In Proceedings of 20th International System Safety Conference (S. Thomason, editor), System Safety Society, August 2002, pp. 214–224.

[11] Popic, P., D. Desovski, W. Abdelmoez and B. Cukic, Error Propagation in the Reliability Analysis of Component Based Systems, In Proceedings of the 16th IEEE international Symposium on Software Reliability Engineering (ISSRE), November 2005, Washington, DC, pp. 53-62.

[12] Roshandel, R., Medvidovic, N.: Toward Architecture-based Reliability Estimation, In Proc. of the Workshop on Architecting Dependable Systems, International Conference on Software Engineering (ICSE 26), Edinburgh, UK, May 2004

[13] Singh, H., Cortellessa, V., Cukic, B., Gunel, E., Bharadwaj, V.: A Bayesian approach to reliability prediction and assessment of component based systems. In Proc. of 12th International Symposium on Software Reliability Engineering (ISSRE'01), 2001

[14] Zhang, F., Zhou, X., Chen, J. and Dong, Y., A Novel Model for Component-Based Software Reliability Analysis, In Proceedings of the 11th IEEE High Assurance Systems Engineering Symposium, (HASE) 2008, pp.303-309, 3-5 Dec. 2008.

[15] Zhang, X and Pham, H., An analysis of factors affecting software reliability, In Journal of Systems and Software, 50(1), 2000, pp. 43-56.

[16] Rabiner, L.: A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, In Proceedings of the IEEE, Vol. 77(2), pp 257-286, 1989.

[17] Cheung, R. C.: A user-oriented software reliability model, IEEE Transactions on Software Engineering, 6(2): 118-125, 1980.

[18] Roshandel R. et al. A Bayesian Model for Predicting Reliability of Software Systems at the Architectural Level. In Proceedings of 3rd International Conference of Quality of Software Architecture (QoSA 2007), Boston, MA, July 2007.

[19] Cheung, L., Roshandel, R., Medvidovic, N., and Golubchik, L. 2008. Early prediction of software component reliability. In Proceedings of the 30th international Conference on Software Engineering (Leipzig, Germany, May 10 - 18, 2008). ICSE '08. ACM, New York, NY, 111-120.

[20] S. Yacoub, B. Cukic, H. Ammar, "A scenario-based reliability analysis approach for component-based software", IEEE Transactions on Reliability, Vol. 53, No. 4, pp. 465-480, December 2004.

[21] Goseva-Popstojanova K. and Hamill, M., Architecture-Based Software Reliability: Why Only a Few Parameters Matter?, In Proceedings of the 31st Annual International Conference on Computer Software and Applications Conference, 2007, vol.1, pp.423-430, 24-27 July 2007.

A New Approach in Tracking Efficiency of Anti-Spam Filter

Slobodan Mitrović[1], Valentina Radojičić[2]

***Abstract* – With the growing number of spam attacks, monitoring of Mail Transfer Agent (MTA) and anti-spam filter is required. Time series analysis and forecasting models for *received messages/identified spam* ratio could be used for performance assessment of some modules of the system. In this paper a simple approach in tracking efficiency of anti-spam filter is considered. Presented experimental results are based on the log data of the MTA and anti-spam filter of The Faculty of Transport and Traffic Engineering Computer Center in Belgrade.**

***Keywords* – monitoring, anti-spam filter, time series, efficiency.**

I. INTRODUCTION

During the last several years anti-spam protection has became equally important as protection against viruses or hacking, because of emerging growth of unsolicited email. Various techniques for filtering spam have been developed and implemented in software solutions. However, spam emails are frequently changing their forms in order to decrease efficiency of filtering, which can be assessed by monitoring process.

Based on this observation, a simple approach in tracking efficiency of anti-spam filter is proposed. MTA log files can be used for calculation of *received messages/identified spam* ratio which can be presented in time series form. This form is suitable for analysis and forecasting techniques can be applied as well. By assessment of real and predicted data, efficiency of given filter can be determined.

In Section II the concept of used MTA and anti-spam filter monitoring is described. Section III briefly presents prediction techniques applied to a given time series. Section IV presents our experimental results followed by corresponding conclusions.

II. MONITORING EMAIL STATISTICS

Email servers based on LINUX/BSD platforms are suitable for integration with different kinds of anti-spam and anti-virus filters, as well as various types of logging and monitoring tools. One of the most popular MTAs is *Postfix* which can be successfully integrated with syslog engines such as *Syslog-NG*. In this way all activities related to email processing can be recorded in log files. The Faculty of Transport and Traffic Engineering Computer Center in Belgrade uses combination of *RRDTool* engine and *Mailgraph* script in order to collect statistics and presents them in graphical manner, Fig. 1.

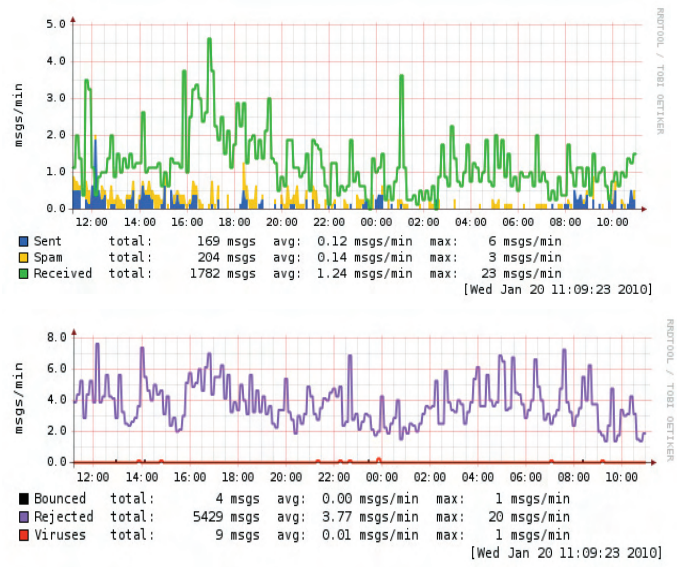


Fig. 1. Collected statistics presented by Mailgraph

Mailgraph is a very simple mail statistics RRDtool frontend for Postfix that produces daily, weekly, monthly and yearly graphs of received/sent and bounced/rejected mail. RRDtool has the ability to store and display data that changes over time in a *Round Robin Databases (*RRDs) that stays fixed in size over time.

In our case, *Postfix – Syslog-NG – RRDtool – Mailgraph* (PSRM) integration collects statistics which is based on the following rules:

1. Source server initiates session of sending email to our PSRM. In case of matching some SMTP restrictions, email will be rejected and session will be closed;
2. If there is no SMTP restrictions matching email will be accepted and sent to anti-spam and anti-virus filters;
3. In spam case or virus detection email will be marked, otherwise, it will be sent to recipient mailbox.

Hence, our RRDs contain statistics related to number of rejected emails, received emails, as well as spam and infected messages identified in bulk of received emails. RRD data can be extracted to XML file. Furthermore, it could be used in any other tool for time series analysis. In this case, time series are related to levels of received emails and level of identified spam in bulk of received emails. In this paper, other statistics are not significant for further analysis.

[1]Slobodan D. Mitrović is with the Faculty of Transport and Traffic Engineering, Vojvode Stepe 305, 11000 Belgrade, Serbia, E-mail: s.mitrovic@sf.bg.ac.rs

[2]Valentina Radojičić is with the Faculty of Transport and Traffic Engineering, Vojvode Stepe 305, 11000 Belgrade, Serbia, E-mail: valentin@sf.bg.ac.rs

III. TIME SERIES AND PREDICTION METHODS

A time series is a collection of time ordered observations $\{y_1, y_2, ..., y_n\}$, each one being recorded at a specific time t, appearing in a wide set of domains such as Finance, Production and Control [1]. In this case, time is independent variable used in analysis of trends (T), seasonal (S), cyclical (C) and irregular/randomized (I) variations that influence the demand data. General time series model can be described as [2]:

$$Y = T + S + C + I \tag{1}$$

Trends can be described by appropriate development, such as linear, exponential, logistic, polynomial, etc. While seasonal factors represent variations that repeating themselves in regular shorter intervals, up to one year, cyclical factors represent variations which interval of repeating is significantly longer.

Smoothing models is another group of time series models that can be considered, relays the assumption that behavior of relatively near future will be same or similar to relatively close history. Moving averages models as well as exponential smoothing models belong to this group [6].

For now, only models which are used in this paper are briefly presented.

A. Linear trend

Linear trend can be described as follows:

$$y = a + bt \tag{2}$$

$$b = \frac{\sum ty - n\bar{t}\,\bar{y}}{\sum t^2 - n\bar{t}^2} \qquad a = \bar{y} - b\bar{t} \qquad \bar{t} = \frac{\sum t}{n} \qquad \bar{y} = \frac{\sum y}{n}$$

Where are: y - item to be forecast (dependent variable); a, b - parameters to be calculated from historical data; t - point of time (independent variable) and n - number of pairs of values (t,y).

B. Moving average

The new value is calculated as the mean of a number of observed values (m):

$$F_{n+1} = \frac{1}{m}\sum_{i=n-m+1}^{n} X_i \tag{3}$$

Value m can be determined experimentally [3].

The purpose of this approach is to reduce irregularities caused by, for instance, seasonal variations [2].

C. Exponential smoothing

Whereas in the simple moving average the past observations are weighted equally, exponential smoothing assigns exponentially decreasing weights over time [2].

$$F_{n+1} = F_n + \alpha(X_n - F_n) = \alpha X_n + \alpha(1-\alpha)X_{n-1} + \alpha(1-\alpha)^2 X_{n-2} + ... \tag{4}$$

Value of α can be between 0 and 1 ($0< \alpha<1$). Recommended values are between 0.2 and 0.3 [3].

D. Examining systematic errors

Examining systematic errors is possible to check by the following procedure; denote the values on the fitted line corresponding to the recorded values of y by $\hat{y}$, and further denote the first readings by y_1, $\hat{y}$, the second by y_2, $\hat{y}$ and so on; calculate

$$w = \sum_{i=1}^{n-1}(y_i - \hat{y}_i)(y_{i+1} - \hat{y}_{i+1}) \qquad v = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{5}$$

and then check the value

$$2 - 2\frac{w}{v}$$

This statistic, known as the *Durbin-Watson statistic*, should lie between 1.5 and 2.5 and preferably between 1.7 and 2.3. If the value is outside this range, the data should be examined for indications of systematic deviation from the fitted curve [2].

IV. TRACKING ANTI-SPAM FILTER EFFICIENCY

This approach is based on the idea that we could predict PSRM integration behavior in the near future. By comparing prediction results and fresh statistics we can track deviation of filter efficiency and do some corrections and tweaks.

Log data of presented PSRM integration has been collected for received email and detected spam from 12.07.2008. 10:00AM, to 15.10.2009. 02:00AM, in 56 hours snapshots. The collected data has been presented by Fig. 2.

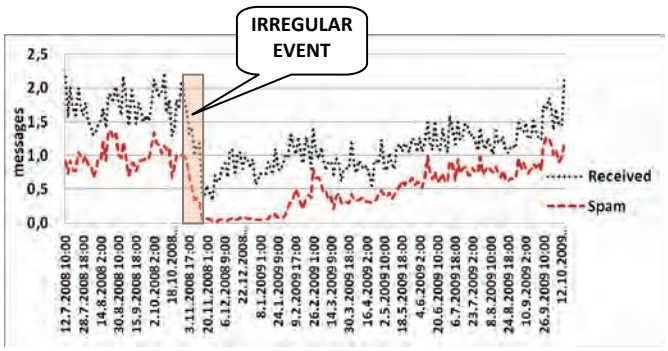


Fig. 2. The total received email and detected spam

From 1st to the 17th day of November, 2008, there have been some system changes. That event must be treated as *the irregular event*. In order to apply some forecasting methods, that event must be excluded, as it shown by Fig.3.

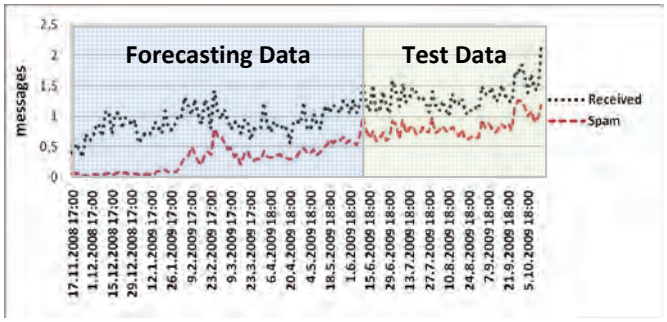


Fig. 3. The total received email and detected spam without the irregular event

It could be notice that each time series (*received* and *spam*) is divided in two parts – for creating forecasting model (from 17.11.2008. 17:00 to 6.7.2009. 18:00) and for purpose of testing (from 6.7.2009. 18:00:00 to 12.10.2009. 18:00).

For the purpose of forecasting procedure, three models have been chosen: moving average, exponential smoothing and linear regression, as shown in Fig. 4:

1. two moving average models have been created:
   - with step of 3 snapshots per 56 hours, which means 1 week (week*R(x3)* and week*S(x3)*)
   - with step of 12 snapshots per 56 hours, which means 1 month (month*R(x12)* and month*S(x12)*)
2. exponential smoothing with $\alpha = 0.2$ (in Fig 4. by legend *expo(R)* and expo*(S)*)
3. linear trend has been calculated by equation (2):
   - for received email - *linear(R)*:
     $$y_R = 0.70411 + 0.00487 \cdot t$$
   - for detected spam - *linear(S):*
     $$y_S = -0.03134 + 0.00743 \cdot t$$

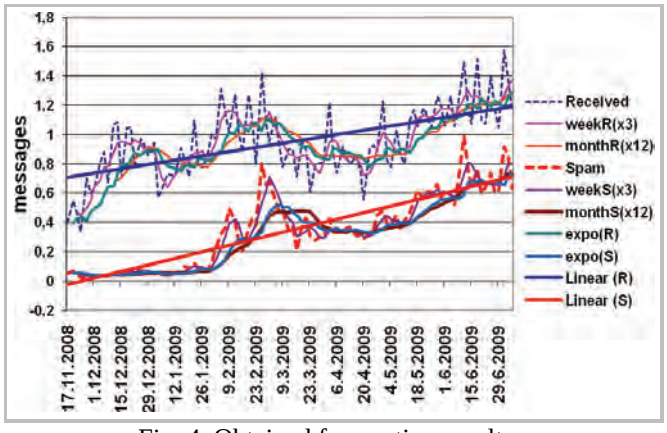


Fig. 4. Obtained forecasting results

Examination of systematic errors has been made by Durbin-Watson statistic. There were also calculated square errors sums and results are shown on Table I.

Durbin-Watson statistic shows no systematic errors in *received* series. The same statistic shows some systematic errors in *spam* series that could be explained by behavior of filter. After system change, new untrained anti-spam filter has been applied and level of detected spam has been decreased in the first 60 days.

TABLE I
DURBIN-WATSON STATISTIC AND SQUARE ERROR SUMS

Prediction method	Durbin-Watson statistic	Square error sums
weekR(x3)	2.547421012	1.81007311
weekR(x12)	1.654315532	2.77121651
expo(R)	1.779870971	3.36099602
linear(R)	1.344073907	3.74682626
weekS(x3)	1.724419852	0.44067631
weekS(x12)	0.709907847	1.21867089
expo(S)	0.87492502	1.19914342
linear(S)	0.642861566	1.47628343

The training process has been also followed by increased frequency of spam attacks which resulted intensive training and that can be visually noticed – after a number of days *spam* curve started to "follow" *received* curve. This fact is confirmed in Fig. 5. by values of deviations (*devR* and *devS*), as well as correlation coefficient – 0.73465.

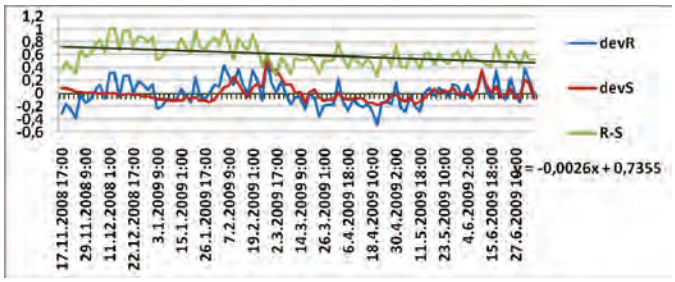


Fig. 5. Deviations of *received* and *spam* series

One of the most important characteristics is related to calculation of linear trend for each series. The slope of linear development of detected spam emails is greater than slope of linear development of received messages, which means that anti-spam filter is becoming more successful as time goes by. This characteristic is confirmed with difference between amounts of received messages and detected spam, as we can see in Fig 5. Also, it could be confirmed by levels of square errors in Fig. 6.

It is possible to conclude that liner trend implementation is the most convenient way for this kind of monitoring purposes and it will be used in the following consideration, only.

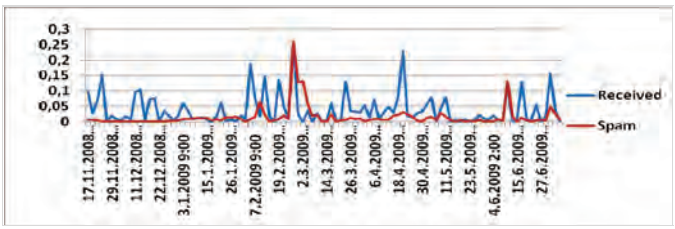


Fig. 6. Square error sums

Calculated linear trends have been tested with *Test data* series. Results are presented in Fig. 7.

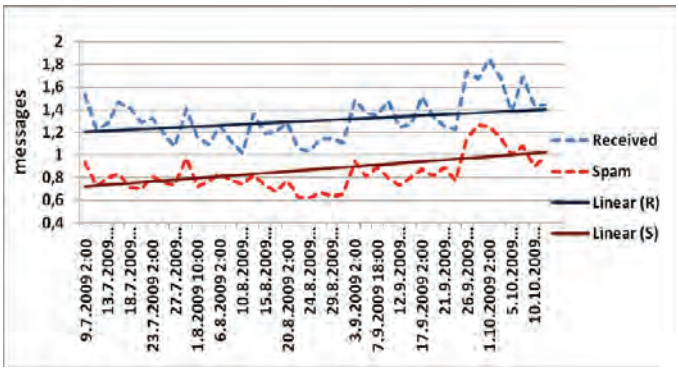


Fig. 7. Forecasted linear trends and *Test data* series

Testing has shown that forecasted linear trends successfully "followed" by curves of received messages and detected spam which is confirmed by correlation coefficient of 0.880009.

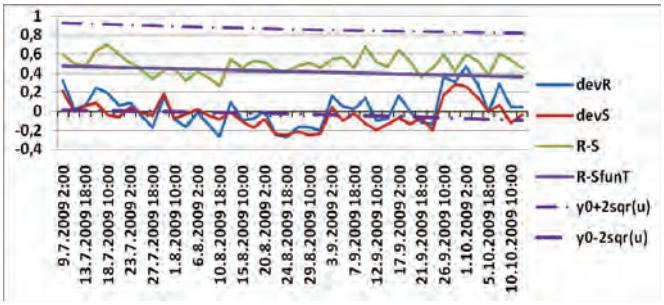


Fig. 8. *Test data* series - deviation of *received* and *spam* series, difference between received messages and detected spam and confidence interval of 95% (*y0+2sqr(u); y0-2sqr(u)*)

Distance between received messages and detected spam continue to decrease which is presented by curve *R-S* in Fig. 8. This curve remains deeply inside of the corresponding confidence interval (95%).

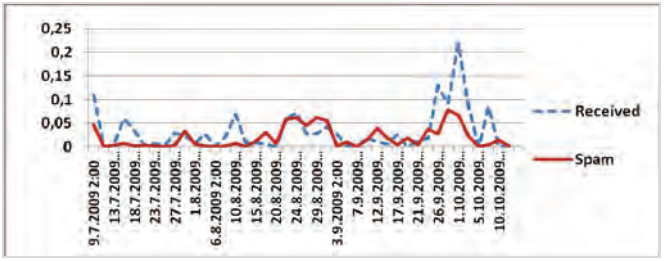


Fig 9. Square error sums

Also, it could be noticed that these square error sums related to difference of time series and corresponding linear trends remained on the same level, Fig 9.

V. Conclusion

Forecasting methods could be used for improvement of anti-spam filter efficiency tracking process, because of their simplicity and variety. Selection of suitable forecasting model depends on e-mail system time series characteristics.

In this paper linear trend has been selected because of advantages in process of visual monitoring of generated graphs, especially with presence of confidence interval.

Although there was no a huge spam attack or detection of anti-spam filter issues during test phase it is reasonably to suppose two different cases. First, higher level of spam attack could show higher level of difference between received messages and detected spam that could stay inside confidence interval. In this case regular tweaks of filter could be done in order to improve detection performance. On the other hand, important issues, such as *Bayesian poisoning*, could lead that level of difference between received messages and detected spam could arise out of confidence interval. In this case some important interventions are required.

Authors believe that for long-term forecasting, in the case of longer time series, linear trend have to be replaced with some more appropriate model such as logistic (Gompertz) trend. There is also possibility for presence of cyclical factors, caused by various factors, such as software replacement, etc.

Acknowledgement

This paper presents some of results in research project TR-11013, funded by Ministry of Science and Technological Development of Republic of Serbia.

References

[1] S. Makridakis, S. Weelwright, R. Hyndman, *Forecasting: Methods and Applications*. John Wiley & Sons, New York, USA, 1998.
[2] Leijon H., *Forecasting Theories*, PLANITU Doc. 61-E, ITU
[3] V. Radojičić, *Prognoziranje u telekomunikacijama*, *GND-Produkt*, Belgrade, Serbia, 2003.
[4] G. Sakkis, I. Androutsopoulos, G. Paliouras, V. Karkaletsis, Constantine D. Spyropoulos and P. Stamatopoulos, Stacking classifiers for anti-spam filtering of e-mail, Proceedings of the 6th Conference on Empirical Methods in Natural Language Processing (EMNLP 2001), Pittsburgh, USA, pp. 44–50, 2001.
[5] J. R. Evans, *Statistics, Data Analysis, and Decision Modeling*, James Robert Evans, Pearson/Prentice Hall, 2007.
[6] S. Mitrović, Neke tehnike detekcije spam poruka, Zbornik radova: SYM-OP-IS 2009, Ivanjica, 2009., str. 55-59
[7] S. Mladenović, S. Mitrović, S.Janković, Unapređenje implementacije modela anti-spam zaštite, Zbornik radova: PosTel 2008, Beograd, 2008., str. 279-288
[8] http://wiki.apache.org/spamassassin, (28.3.2010.)
[9] http://www.balabit.com/network-security/syslog-ng, (28.3.2010.)
[10] http://mailgraph.schweikert.ch, (28.3.2010.)
[11] http://oss.oetiker.ch/rrdtool/doc/index.en.html, (28.3.2010.)
[12] http://www.postfix.org, (28.3.2010.)

Simulation Modelling of Technology and Capacity Design in Border Railway Station

Kire Dimanoski[1*], Gordan Stojić[2], Slavko Vesković[3], Miloš Ivić[3], Sanjin Milinković[3]

Abstract: **Simulation and modelling are the crucial methods which are recently used for infrastructural capacity dimensioning. This is not applied just for infrastructure, it's also applied on determination of the optimal number of workers on matched positions, as the number of shunting locomotives, squads, station regions etc. In this work you can see simulation model that enables dimensioning the work technology and the capacities of border railway station.**

Keywords **– modelling, simulation, capacity, technology, border railway station**

I. INTRODUCTION

The capacities of one railway station, in specified time period and terms, are enabling receipt, processing and dispatch trains. Defining capacity is needed to define the time tables, the traffic organization and technological processes, their optimization, planning of investments etc. The basic problem which arises is how to dimension capacities, so the train service can be carried out without problems. It's necessary to have in mind that, infrastructural facilities and resources are extraordinarily expensive, as at the building and procurement point also at the maintaining. And as a result of this, there are large expenses for working force. This means that their improper dimensioning can affect on the profitability of the work of the railway because railway capacity is not static. It is extremely dependent on how it is used [9].

In this paper it is shown how to dimension the technology and capacity in the border stations. Border stations are points in which there is a significant trains standing time on the way to the final destination. The trains in the border station are hold for passenger's exchange, the police and customs formalities, locomotives change, adding or removing wagons, technical and commercial review, receiving documentation from and off the trains, making documentation and so on. Often, the border station represents point of turnover for the passenger trains in internal traffic. All those standings directly affect on the capacities of the border station and its all on the grounds of the existing technology in processing of the trains.

In order to dimension capacities and technology of the border station are developed two models: analytical and simulation model. The models are tested on the example of the border station Kremenica.

Nowadays the most used ways to analyze and simulate the station traffic is to use some of the commercial expensive software's where the possibilities are limited. Considering that, my research showed that it is not needed to buy such software and to simulate traffics on the following kind of way. In other systems it is necessary to use parallel processes to reduce the computing time and therefore simulators are related for solving with parallel computing based on a multiple-CPU server and they concentrate, in particular, on the speed-up performance [6] [13], but in this simulation, the processes are obtained by low performance computer.

The technology and railway station capacity modelling are presented in the following papers [12], [4][4], [5] and [10]. The paper [12] presents a simulation model for technology and capacities optimization for interim stations (transit stations) with usage of the Non-Markovian systems queueing theory. To simulate the railway traffic at the stations (into the railway transit stations), in [10] *Cellular Automata* is used, and in the [4] *hybrid Petri nets-based simulation model.* In the paper [5], marshaling yard station model is presented, where the station optimization is the main question, and is based on the simulation modelling of the technological operations such as train formation and unformation. The analytical modelling of the technological operations in the marshaling yards it is made in the paper [14].

In the work [7] for traffic congestion controls uses a queue thresholds. For the queueing theory is used the system GE/GE/1/N approximation. Developed to study the spread of traffic congestion in complex networks.

Modelling the spread of traffic congestion in complex rail networks in [8] is used a Weight-evolving traffic network model which is based on Barrat–Barthelemy–Vespignani (BBV) model. This paper simulates and analyzes the process of the emergence and spreading of congestion, which is triggered by adjusting of data generating speed and data sending ability of the network.

In the literature I have found no papers which are dealing with optimization of the work in border stations using simulation modelling, but there are presented simulation models which are concerning the movement of trains to advance adopted technology in the transit railway stations.

The border railway stations shall perform customs work which could affect on the retention of the trains, especially in freight traffic. Retention time of trains is not neglected in passenger traffic. Moreover, there are other railway technical operations, such as: change of locomotive and train personnel, inventory of train, etc. which also affect the maintenance of trains at border stations. Technology and time of execution of

[1] Kire Dimanoski, Secondary Technical School "Vlado Tasevski", 3 Makedonska brigada bb, 1000 Skopje, Macedonia, email: kdimanoski@yahoo.com (corresponding author)

[2] Department of Traffic Engineering, Faculty of Technical Sciences, University of Novi Sad, Serbia

[3] Department of Railway Transport and Traffic, Faculty of Traffic and Transport Engineering, University of Belgrade, Serbia

customs operations and technology directly affect to infrastructure facilities dimensioning in the border station facilities.

For example among EU members, the trains are kept in border station in order to performance the customs affairs there, between certain countries there is some retention of trains for carrying out the railway technology operations.

This paper presents a simulation model for technology and infrastructure facilities optimization in the border railway stations based on discrete stochastic systems simulation such as the technological processes in the border stations and in other train stations. Results are compared with results obtained by analytical way.

The model is tested on example of Kremenica border railway station in Macedonia, which is planned in all strategies for reconstruction, and it is a boundary between a country that borders the EU (Macedonia) and EU Member State (Greece).

II. TYPE AND PURPOSE OF THE KREMENICA STATION

Kremenica station is located on the main railway line Veles-Bitola-Kremenica-Greece border, which is a part of the Corridor's X arm "D", Kremenica also is a frontier station between the Macedonian Railways (MZ) and the Greek Railways (CH). Kremenica has competence for receipt, processing and dispatching passenger trains and goods trains in internal and international traffic. Station has two tracks for receipt and dispatch trains, where trains are stopping to exchange passengers, locomotive change and customs procedures, while the loading and unloading of goods wagons is envisaged to be made on the reversing triangle. Neighboring station are station Bitola (MZ) and station Mesonision (CH).

III. SIMULATION MODEL

For station Kremenica capacity and technology modelling is used GPSS/FON V.4.0 simulation program. According to model needs, input and output train speed are defined.

Train retention time in the model devices it is calculated by the following equation:

$$t_{zd} = \frac{3{,}6 \cdot (L + \ell_{voz})}{V} \qquad (1)$$

where:

L - length of the elements in the real system expressed in meters;

ℓ_{voz} - length of train expressed in meters;

V - train speed when the train is passing the observed device, expressed in km/h.

For the purposes of the model, length of the elements (devices) of the real system are calculated, and the retention time on them. However, the real time, when device is busy and the train is passing, is far greater. The train retention at the first element in the formation of driving route[1] is calculated according to the equation:

$$T_{ze} = t_{fp} + t_{us} + \frac{3{,}6 \cdot (L + \ell_{voz})}{V_{ul}} \qquad (2)$$

where:

t_{fp} – driving route formatting time; in this work is 0,3 *min* or 18 *s;*

t_{us} - signal perception time when the trains are moving with classical speed (9 *s*);

V_{ul} – average train speed when the train is entering the station (*km/h*).

Occupation time of the "n" element in the driving route (T_{ze_n}), is calculated by the following equation:

$$T_{ze_n} = t_{fp} + t_{us} + \sum_{i=1}^{n-1} t_{zd_i} + t_{zd_n} \qquad (3)$$

where:

$\sum_{i=1}^{n-1} t_{zd_i}$ - train retention time sum on the previous devices on the driving route;

t_{zd_n} - train retention time on the "n" device.

Similar logic applies when it comes to the output driving. The occupation time of the "i" track device in the shunt-dispatching yard is calculated by the equation:

$$T_{ze_i} = t_{zd_i} + t_{fp} + t_{ot} + \frac{3{,}6 \cdot \ell_{voz}}{V_{izl}} \qquad (4)$$

where:

V_{izl} – average speed when the train leave the station (*km/h*);

t_{ot} - train dispatch time, which is taken to be 1 min.;

Occupation time of the "*n*" output device in the shunt-dispatching yard is:

$$T_{ze_n} = t_{fp} + t_{ot} + \sum_{i=1}^{n-1} t_{zd_i} + t_{zd_n} \qquad (5)$$

Model boundaries are: from station Bitola side is the dispatch main signal of that station and from Mesonision station side boundary is the country border line. Input device from Bitola side is OTSEK1 and from Mesonision side is OTSEK2 device. These two devices are first devices in which the trains are entering in the simulation system.

Although the input devices in the model are first devices which are taking transactions, also they are last devices in the simulation output drive from Kremenica station.

Trains entering in the system is done after they leave the station Kremenica or Mesonision (border station in Greece), and that is performed after the trains leave the last station switch or after the dispatch of the train. And the trains exit from the system is performed after they leave the last device of the system, i.e. after the main exit signal of the Kremenica station on both sides. All devices are shown in (Fig. 1.).

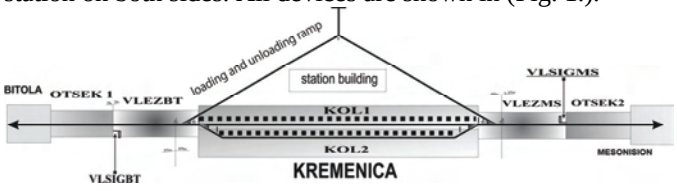


Fig. 1. Scheme of the devices used in the model simulation

[1] In the modelling (according to the train movement dependence table for this model), it is taken that one driving route includes more devices (example: one or more group's switches, tracks etc.)

The movement of trains in the model as in the real system, as same in the model, it's subject of certain patterns. In the movement of trains through the station are used many different types of routes: entry, exit and transit routes and overlap routes. Due the consideration of all necessary and possible driving routes for this model routes table is developed, in which we enable the realization of all simultaneous driving. In the model for routes regulating and security the following rules are used:

1. When we are forming a route (entry route means that a overlap route is included), another route can be formed if only these two do not overlap, intersect or touch,
2. All relevant signals whose implementation could jeopardize the route (front, ends, or jeopardizing with overcoming), should show a signal of prohibited driving,
3. The route in any of its parts must be unoccupied by other vehicles or bands.

In the model are processed 3 types of trains such as: passenger, goods transport (classical technology) and goods transport (combined transport technology). Based on the statistical examination it is established that 25% of the trains are passenger trains, and to 75% are freight trains, from them, the 50 % are classic cargo and 50% are intermodal.

Testing of models is made for the following terms of train's traffic with the alternatives included:

- ALTERNATIVE1 (for 1 pair of trains in 4 hours),
- ALTERNATIVE2 (for 1 pair of trains in 3 hours),
- ALTERNATIVE3 (for 1 pair of trains in 2 hours),
- ALTERNATIVE4 (for 1 pair of trains in 1 hour).

Generation of trains is performed by equal distribution in minutes, like this: ALT[240,30], ALT2[180,15], ALT3[120,10] and ALT4[120,10].

The testing results are shown in (Figures: 2, 3, 4,).

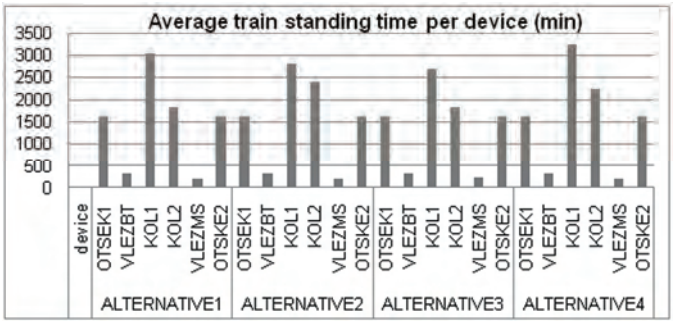


Fig. 2. Average train standing time per device (min)

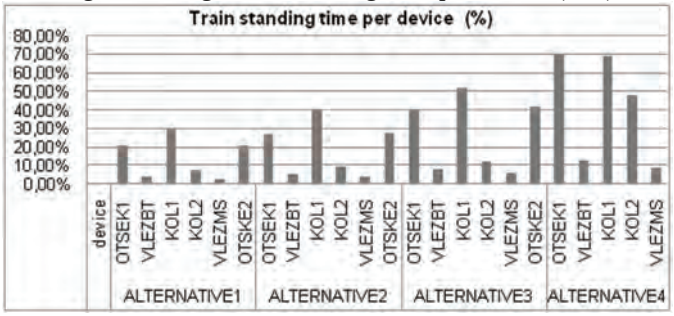


Fig. 3. Train standing time per device (%)

At the fourth alternative it can be recorded that the devices OTSEK1, OTSEK2 and KOL1 are used about 70% of the time, which means that the alternative 4 is bottleneck in these two devices for the specified quantity of traffic, while the occupation of the other devices is going from 5% to 40% with a tendency to increase in relation to decrease the interval of incoming trains.

On the basis of the obtained results for the warehouses on both entering sides, Bitola and Mesonision, it is provided that their capacity is not envisaged fully yet, but at the first alternative there is a standing of 82 minutes and arises, which is quite long time. That can tell us that the capacity of the station is not sufficient for processing the trains. This standing time grows at the alternative 4 up to 163 min and 93 min, or average in hours, 2.7 h and 1.6 h.

The testing results on alternatives are showing that there is standing of the trains in lines (on main signals) only just for alternative 3 and 4.

Next significant fact, from 22 trains, which are entering Kremencia station by the third alternative from Bitola side, 2 trains are standing on the main signal of Kremenica, while in the fourth alternative 18 of 22 trains from the station Bitola side and 18 of 18 trains from Mesonision side are standing on the main signal and waiting to enter the Kremenica station (Table 5). That means from 81% to 100% of the trains which have entered the Kremenica station are stopped on entry, with high value of standing time from 39 to 211 minutes for both sides, Bitola and Mesonision, which means all trains which have entered the Kremenica station from Mesonision side are kept at entry main signal.

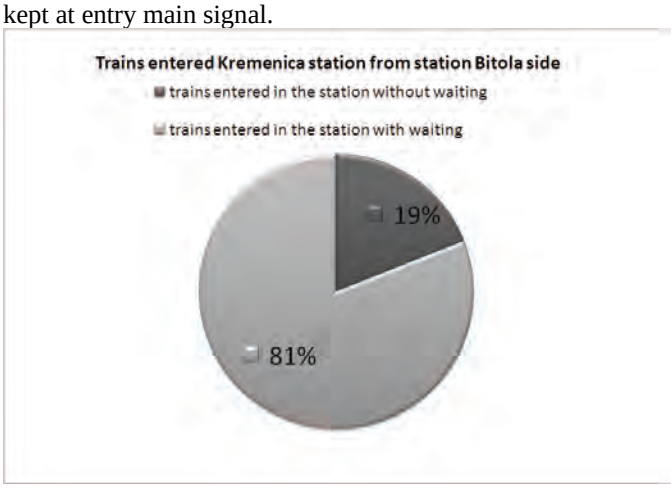


Fig. 4. Percentage of trains which entered the station Kremenica with and without standings at the entry from station Bitola side in Alternative 4

If the trains came as in alternative 1 and 2, the facilities already are sufficient, because they are not appearing waiting, according to the simulation. The analytical method has also showed that existing facilities can be satisfied, while the alternatives 3 and 4 are appearing need for increasing capacity or improving the technology of train processing, especially regarding the goods trains where are showed a big standings at entry into the system or waiting for trains processing.

When those two types of dimensioning are compared they give similar results, the analytical method showed that the expanding capacities can be considered if the trains come like in alternative 3 and 4. This means if the trains come in intervals of 120 min, and less, then a need for new equipment is shows up. In support of this obtained test results, which are showing great standing time in order to start the trains processing, it is indicating the purpose of expanding the capacity and intervention on the treatment technologies.

IV. TECHNICAL AND ORGANIZATIONAL MEASURES TO REDUCE THE RETENTION OF TRAINS AT THE BORDER STATION KREMENICA

All operations and activities carried out in border stations are necessary, because quality analysis can be divided into those which pursue public authorities and those which the railways run.

Kremenica railway station in the near future must be converted into common rail station for two administrations (MZ and CH). The prerequisite is a basic assumption under which the total retention time of passenger and goods trains could reduce on the both sides and thus could reduce the total travel time. Thus would enhance the competitiveness of the railway transport.

If Kremenica station would be a joint border station, it will accommodate all competent railway authorities for treatment and control of trains (MZ and CH staff) and state departments of customs authorities and police from Macedonia and Greece, and the presence of the competent inspection services (sanitary, environmental, veterinary, etc.).

Besides, modern trends in railway traffic are trying to use the international rail network for data transmission - HERMES and the electronic waybill form - DOCIMEL, which could significantly reduce the standings of trains at border stations.

CONCLUSION

Running the train traffic is just one of the many complex processes that take place in the complex system of railways. Railway processes are suitable for modelling and simulation, there for analytical models didn't give always the optimal solution, and the experiments in the real system could be lengthy, costly and risky. So, the model developed in this paper allows dimensioning of technology and capacity based on application of analytical models and simulation of train traffic through the border station Kremenica.

On the basis of the results obtained from the testing of the models can be concluded that the station would satisfy the needs for processing trains in interval bigger than 2 hours.

The measures offered are based on the experience of railways which belong to the EU, and members of the UIC (International Railway Union) and thus measures have to be applied by MZ and CH. So by turning the station into a common border station (common to both countries) as well as implementing new information technologies in the process of treatment and dispatching trains, would increase the processing speed of the trains. It is also essential in the joint station Kremenica all operations of the railways and the authorities to be maximally simplified and organized as a parallel work of individual operations as possible.

REFERENCES

[1] B. Radenković, M. Stanojević, A. Marković, *Računarska simulacija*, (Fakultet Organizacionih nauka, Beograd, Serbia 2001 – In serbian).

[2] M. Čičak, *Modelling Railway Traffic and Transport*, (Faculty of Traffic and Transport Engineering University of Belgrade and ŽELNID, Beograd, Serbia, 2003 - in Serbian).

[3] M. Čičak, S. Vesković, S. Mladenović, *Modeli za utvrđivanje kapaciteta železnice*, (Saobraćajni fakultet, Beograd, Serbia, 2002 - In serbian).

[4] F. Kaakai, S. Hayat, A. El Moudni, *A hybrid Petri nets-based simulation model for evaluating the design of railway transit stations*, Simulation Modelling Practice and Theory 15, (2007), 935–969.

[5] G. Stojić, S. Vesković, M. Čičak, *Modelling of technologies and capacities of technical freight stations*, Železnice, No 7-8, 2003, 207-219.

[6] H. Schlenker, *Distributed Constraint-Based Railway Simulation*, Lecture Notes in Computer Science, Volume 3392, (Springer, Berlin, 2005), 215-226.

[7] I. Awan, *Analysis of multiple-threshold queues for congestion control of heterogeneous traffic streams*, Simulation Modelling Practice and Theory 14, (2006), 712–724.

[8] L. Ke Ping, Zi You G., *An improved equation model for the train movement*, Simulation Modelling Practice and Theory 15, (2007), 1156–1162.

[9] M. Abril, F. Barber, L. Ingolotti, M. A. Salido, P. Tormos, A. Lova, *An Assessment of Railway Capacity*, Transportation Research Part E 44, (2008) 774–806

[10] J. Xun, B. Ning, P. Li Ke, *Station Model for Rail Transit System Using Cellular Automata*, Commun. Theor. Phys., Beijing, China, Vol. 51, (2009), 595–599.

[11] A. Baublys, *Principles for modelling technological processes in transport terminal*, Transport Research Institute, Vilnius Gediminas Technical University, Transport, 24 (1), (2009), 5-13.

[12] M. Čičak, G. Stojić, S. Vesković, *Simulation model for defining of technology and capacities of railway station*, XXIII Simpozijum za operaciona istraživanja, Zlatibor, Srbija, (1996), 683-686.

[13] M. Hulea, C. Avram, T. Letia, D. Muresan, S. Radu, *Distributed real-time railway simulator*, Proceedings of the 2008 High Performance, Computing & Simulation Conference, ECMS, 2008.

[14] O. Ischuka, *Model of optimization of technological process of operation of marshalling station*, Proceedings of the 6th International Scientific Conference TRANSBALTICA, 2009.

Agent-Based Simulation of Grid Resource Management Using Auction Market

Plamenka Borovska[1], Ognian Nakov[2], and Adelina Aleksieva-Petrova[3]

***Abstract* – Agent-based simulation is a new approach to modelling systems comprised of autonomous, interacting agents. The primary goal of this article is to describe the construction of an agent-based system that simulates auctions for grid resource management. This system can be used to evaluate the efficiency of auctions and as a tool to dynamically manage and distribute resources throughout grid systems.**

***Keywords* – agent-base simulation, Grid, resource management, auction.**

I. INTRODUCTION

Agent-based simulation refers to a model in which the dynamic processes of agent interaction are simulated repeatedly over time, as in systems dynamics, and time-stepped, discrete-event, and other types of conventional simulation [1].

Markets have emerged as a new paradigm for managing and allocating resources in complex systems. Several research systems have explored the use of different economic models for trading resources to manage resources in different application domains: CPU cycles, storage space, database, query processing, and distributed computing [2]. Such systems are Popcorn, Java Market, Mungi, Enhanced MOSIX, Nimrod-G and so on.

The primary goal of this article is to describe the construction of an agent-based system that simulates auctions for grid resource management. This system can be used to evaluate the efficiency of auctions and as a tool to dynamically manage and distribute resources throughout grid systems.

II. CASE STUDY

In this study, we simulate the auction mechanism of resource management.

In auctions the sellers (suppliers) submit their available resources to an auctioneer for processor time and another auctioneer for disk space at minimum prices; i.e. with the average profit per slot and per file formed on the basis of the value of the function of demand.

Buyers submit their requests by means of tasks (jobs), every task specifying the type, number and duration of resources needed by it.

Every buyer has a defined available budget ($G) with which to pay for the resources needed for its tasks, which is current for a fixed period of time.

Buyers submit requests for resources as long as they have tasks and a current budget.

The auctioneer maintains a list of available resources, running through the resources in each round in the order in which they are listed, implementing a second-price auction (in a concrete arrangement this can generally be Danish, English, etc.).

When a given resource is turned over for use by a consumer it is removed from the list of available resources and the consumer's budget decreases by a corresponding amount ($G).

After receiving resources for the current task, the buyer submits requests for its next task.

After every auction there are buyers who have already rented some resources (and for which they have already paid) who await additional resources to be provided to them from the next auctions in order to complete the current task..

In order to outline the architecture of this agent-based architecture three primary roles have been defined for agents:

- PRODUCER (seller, producer of resources)
- CONSUMER (buyer, application)
- AUCTIONEER (auctioneer, broker)

There are two types of producers: suppliers of processor time and suppliers of disk space.

Buyers (consumers) submit their requests through tasks. For each task the type, duration and number of utilized resources is described. Every buyer has a discrete budget available for paying for utilized services, which is current for a particular period of time and is subject to updates.

The auctioneer maintains a list of all the participants in the auction: CONSUMERS and PRODUCERS. Likewise, the auctioneer maintains a list of all the resources provided by the suppliers and a list of all the requests made by the buyers.

At every round the producers submit their available resources to the appropriate auctioneer at the price determined by the function of supply mentioned above.

At every round buyers submit their bids (requests), specifying the number, type and duration to use the resources, as well as the price they would pay for such resources according to their means.

It is important for us to note that at every round buyers can bid at most for one task.

Every round is conducted on the principle of the second-price auction in order to determine the winner. With this type of auction the buyer who wins is the one who has submitted

[1]Plamenka Borovska is with the Faculty of Computer Systems and Control, TU-Sofia, bl. Kliment Ohridski 8, 1000 Sofia, Bulgaria, E-mail: pborovska@tu-sofia.bg

[2]Ognian Nakov is with the Faculty of Computer Systems and Control, TU-Sofia, bl. Kliment Ohridski 8, 1000 Sofia, Bulgaria, E-mail: nakov@tu-sofia.bg

[3]Adelina Aleksieva-Petrova is with the Faculty of Computer Systems and Control, TU-Sofia, bl. Kliment Ohridski 8, 1000 Sofia, Bulgaria, E-mail: aaleksieva@tu-sofia.bg

the highest offer, thereby receiving the resource, and if there is a buyer who has submitted a second-highest offer that is above the minimum price set by the supplier, the final price for the resource is set at this second-highest bid.

III. PROGRAM IMPLEMENTATION

The Java software program JADE (Java Agent Development Framework) [3] is used in creating the forthcoming application. It facilitates the introduction of a multi-agent system in an environment that abides by FIPA specifications [4] and supports the conscription of graphical instruments to maintain and eliminate errors. The agent platform may be dispersed across computers and the configuration may be remotely controlled through a GUI (Graphical User Interface).

Three primary packages and two supplementary packages

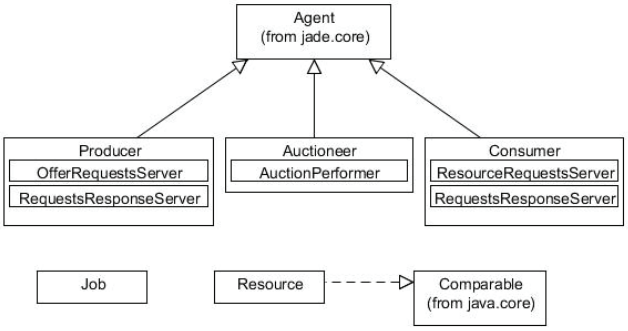


Fig. 1. General scheme of main packages

are utilized:

- Producer– with two subcategories: OfferRequestsServer and RequestsResponseServer, which records the behavior of the seller agent.
- Consumer – with two subcategories: ResourceRequestsServer and RequestsResponseServer, which records the behavior of the buyer agent.
- Auctioneer – with one subcategory: Auctionperformer, which records the behavior of the auctioneer (the resources broker).
- Job – a supplementary category that is used by the Consumer in order to generate tasks.
- Resource– a supplementary category that records each given resource (type, price and owner).

All three primary categories – Producer, Consumer and Auctioneer – are derived from the Agent category, which is a built-in JADE category.

In the setup() portion all of the characteristics of the agent are initialized and the different behaviors that the agent will express are assigned. These behaviors are described via internal categories derived from the primary category Behavior.

All of the categories describing different behaviors are part of jade.core.behaviours.*.

In the system categories the following behaviors are utilized:

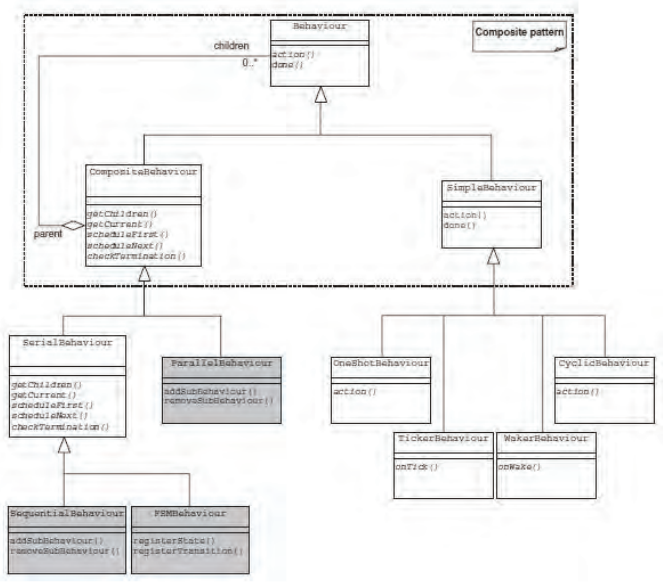


Fig. 2. The hierarchy of behavior packages in JADE

A. Producer

We have two subcategories describing two different behaviors:

OfferRequestsServer class – derived from CyclicBehaviour, which describes a periodically repeating behavior.

Each agent maintains its own message trail. The primary parts of the message are the body (contents), sender, recipient(s), the type (according to the standards of FIPA), the topic of the "conversation", etc.

In our case, the agent, in keeping with its assigned behavior, waits for a CFP (Call For Proposals) message, which serves as an "invitational" message. As long as there are no such messages on hand in the message trail the agent is blocked from using processor resources.

When such a message arrives the agent is activated and the corresponding actions are carried out – in this case it estimates the minimum resource price and sends all of its available resources at this price.

RequestsResponseServer class – derived from CyclicBehaviour. In this case this class is used for periodically receiving the auction results (whether a given agent has sold a resource and received a profit).

The agent awaits an ACCEPT_PROPOSAL message, which means "the offer has been accepted". As long as there are no such messages on hand in the message trail the agent is blocked from using processor resources.

When such a message arrives the agent is activated and the corresponding actions are carried out – in this case if a resource has been sold the profit is increased and the corresponding resources are marked as being in use for the specified duration for which has been paid.

B. Consumer

We have two subcategories describing two different behaviors:

ResourceRequestsServer category – derived from CyclicBehaviour. This class is used to periodically send requests to the auctioneer for resources needed in order to complete a given task.

The agent, in keeping with its assigned behavior, awaits a CFP (Call For Proposals) message, which serves as an "invitational" message. As long as there are no such messages on hand in the message trail the agent is blocked from using processor resources.

When such a message arrives the agent is activated and the corresponding actions are carried out – in this case the price that the agent can afford to spend is calculated based on its average speed of spending and its current budget, then it sends a request containing the number, type, duration and price it is willing to pay for a given resource to the auctioneer.

RequestsResponseServer class – derived from CyclicBehaviour. In this case this category is used to periodically receive the results from the auction (whether a given agent has won a resource and at what price).

The agent waits an INFORM message, which means "the offer has been accepted". As long as there are no such messages on hand in the message trail the agent is blocked from using processor resources.

When such a message arrives the agent is activated and the corresponding actions are carried out – in this case if it has won a resource the agent's budget is reduced and the corresponding work is marked as completed (if all the necessary resources for it have been supplied).

C. Auctioneer

Here we have one primary behavior: AuctionPerform category – derived from Behavior

In this case we have "round-specific behavior", whereby in each round the auctioneer carries out specific actions, after which it proceeds.

Round 1: The auctioneer sends messages to all agents (sellers and buyers) announcing the start of the auction, then waiting for their offers;

Round 2: After the auctioneer has announced the start of the auction it waits for all agents to send their replies – requests from buyers and offers from sellers. Only when the auctioneer has received responses from all agents does it proceed.

Round 3: After information from all participants is received then it is possible for the auction to be carried out. The action auction() is called upon, which carries out the second-price auction.

Round 4: After this auction is concluded and all resources have been distributed the corresponding messages are sent to all participants and the auctioneer moves on to the next auction.

The critical moment that needs to be mentioned is the registration of the agents into the so-called "yellow pages". Every agent is registered in them in order to easily be found in the future along with the type of services they offer as well as a brief additional description. Every agent has the ability both to register and to search for services.

In our case buyers and sellers register one service each.

This registration is subsequently used by the auctioneer in order to restrict the relevant agents according to their type.

IV. Simulation

The following parameters are given for the agents participating in the experimental formulation:

Producers:

Agent 1P → CPU Producer, 5 Slots, 6$/slot/time unit

Agent 2P → CPU Producer, 7 Slots, 8$/slot/time unit

Consumers:

Agent 1C → budget 100$, next budget refresh in 19 time units

Agent 1C Jobs:

Job1 → 4 CPU slots for 2 time units

Job2 → 3 CPU slots for 3 time units

Agent 2C → budget 150$, next budget refresh in 18 time units

Agent 2C Jobs:

Job1 → 2 CPU slots for 1 time unit

Job2 → 3 CPU slots for 1 time unit

The results from the simulation that was carried out are summarized in the following graph. Time is tracked along the X-axis, while the average price paid by all agents in each round is tracked along the Y-axis.

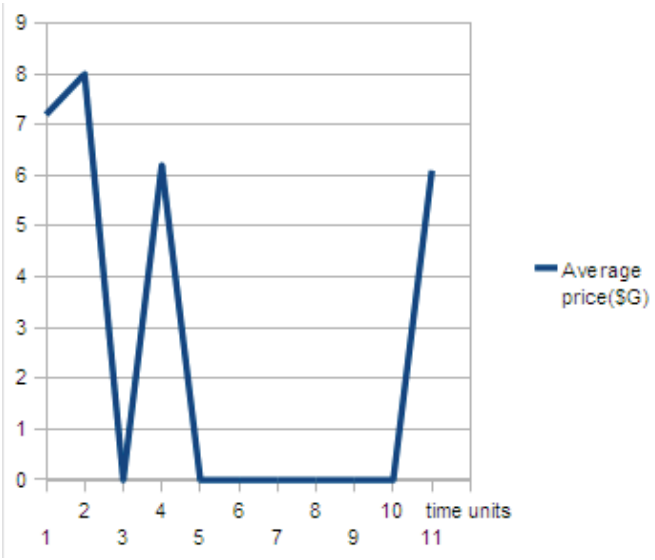


Fig. 3. Average price paid by all agent in simulation

V. Conclusion

Agent-based technologies are aimed at assisting in application development as they are based on a decentralized approach to intelligent agents. We have discussed the case study and implementation of the agent-based simulation of auction for resource management. The advantage of this

system is its agent-based architecture, which approximates to the greatest possible degree the actual organization of grid systems, along with its own requirements and problems.

REFERENCES

[1] Macal, C., Michael J. North: AGENT-BASED MODELING AND SIMULATION: ABMS EXAMPLES, Proceedings of the 2008 Winter Simulation Conference

[2] Li Chunlin, Li Layuan, The use of economic agents under price driven mechanism in grid resource management, Journal of Systems Architecture 50, 2004, pp.521–535

[3] JAva Platform DEvelopment Framework (JADE), http://jade.tilab.com/

[4] FIPA, URL: http://www.fipa.org/

[5] Qiang Tong, Zhidong Shen, Research of Mobile Agent Technology for GIS Grid, Ninth International Conference on Hybrid Intelligent Systems, 2009, 5p.

[6] Gang Wang, Tao Wen, Quan Guo, Xuebin Ma, A Knowledge Grid Architecture Based on Mobile Agent, Proceedings of the Second International, Conference on Semantics, Knowledge, and Grid, 2006, 4p.

Analysis of Identity Based Firewall Systems

Nenad Stojanovski[1] and Marjan Gušev[2]

***Abstract* – The advances in today's computer technology bring the user's experience to such level where one of the most important things is the user mobility. The rise of user mobility within enterprise networks has brought the needs to allow mobile users to use mobility with full capacity in all aspects of collaboration. This problematic issue inspired many IT vendors to develop their own security solutions that extend user mobility. The general security solution to extend user mobility was introduced with identity based firewall systems, which extend the evaluation level of TCP/IP packets by inserting user identity to the whole inspection process. In this paper we will give a detailed analysis of the current identity based firewall solutions. During the analysis we will compare in lot of details how these solutions work and what they need in order to allow user mobility. As part of our research we develop and apply methodology with indicators to evaluate how these solutions affect computer network complexity and how they affect complexity in enterprise network.**

***Keywords* – Identity based firewalls, user identity, firewalls, network security, computer networks, firewall analysis.**

I. INTRODUCTION

The advances in today's computer technology bring the user's experience to such level where one of the most important things is the user mobility. The rise of user mobility within enterprise networks has brought the needs to allow mobile users to use mobility with full capacity in all aspects of collaboration. This problematic issue inspired many IT vendors to develop their own security solutions that extend user mobility.

The main purpose of this paper is to analyze identity based firewall solutions that can be found on today's IT market. In the first chapter we give a general review on how firewall systems work and what intelligence is present behind the filtering mechanisms. The second chapter gives an analysis of the major players in the identity based firewall market. The analysis consists of information from technical character, where we present the findings of what is needed for these solutions to properly function. The last chapter gives a summary about the analysis of the firewall systems.

[1]Nenad Stojanovski is with Makedonski Telekom AD, Orce Nikolov BB, 1000 Skopje, Macedonia, E-mail: nenad.stojanovski@telekom.mk

[2]Marjan Gušev is with the Faculty of Natural Sciences and Mathematics, Ss. Cyril and Methodius University,Arhimedova b.b., PO Box 162, 1000 Skopje, Macedonia E-mail: marjan@ii.edu.mk

II. FIREWALL TECHNOLOGIES

A. Today's firewall technology

Times During the past two decades, Information Technology has evolved at such level that the complexities of the applications that are used by the users introduce the needs of extra security layers, aside from those present by the applications. Because of these needs, the firewall technology was introduces to help protect the assets that are sensitive and that shouldn't be accessed by everyone. It can be easily said that the firewall is a security gateway which is made up of the following components:

- Security policy – is a special type of policy that resides on the gateway, which contains all the rules that allow/deny access to certain resources. The rules are designed using IP addresses, TCP/UDP ports or applications
- The firewall is a software or a hardware device – it comes in the form of an appliance or in the form of an software add-on to the operating system
- Extra security features - as an addition to the firewall security policy, the firewall implements extra features that help protect the assets behind the firewall
- Demilitarized Zone or DMZ - a network segment that exists between the trusted and untrusted networks. DMZs offer some level of protection to the hosts that exist within that segment of the network.

Although, firewalls are introduced as a security mechanics it must be pointed out that there are advantages and disadvantages in using firewalls. The main advantage is the extra centralized security then bring, as well as the total cost of ownership of having a centralized firewall. The disadvantages are that they pose a bottleneck to the network, as well as false sense of total security when we speak about the insider attack.

In the beginning, the firewall at first had the task to protect the network at the IP level, layer 3 of the TCP/IP model. As time passed and applications evolved there were new challenges that had to be fulfilled by the firewall. In order to fulfill the needs, firewall systems had to evolve to a level where they could actively support the needs of modern IT technology. Today's firewall systems support all layers starting from the IP layer and ending to the Application layer.

Firewall systems are usually placed at layer 3, 4 or 5, depending on the control and protection they have to offer to the assets. The firewall systems that operate at layer 3 and 4 are called packet filtering firewalls. Their purpose is to filter IP and ICMP traffic, as well as TCP/UDP ports. A firewall

system at layer 5 is also known as application gateway and has the purpose to filter the traffic based on the application that generates the traffic.

B. Packet filtering firewalls

Packet filtering firewalls were the first firewall system that were invented. To be more precise, the first packet filtering devices were routers which had the ability to filter traffic based on layer 3 and 4 information. Packet filtering firewalls should be able to do the filtering based on the following fields: source IP address, destination IP address, TCP/UDP source port, TCP/UDP destination port.
Example firewall rules are shown on Figure 1. Some packet filtering firewalls allow additional layer 4 options to be added to the rules, like TCP flags that sometimes are very vital for enhancement of the TCP/IP communication between hosts. Packet filtering firewalls also support stateful packet inspection. The state oriented connection tracking add a state table to the firewall in which the firewall keeps information about the incoming or outgoing connections that go through the firewall system. In this way, the firewall system adds the flavor of inspecting if the incoming or outgoing packet correlates with an entry in the state table of the firewall. In this way, the firewall adds an extra mechanism to the network which protects against certain types of network attacks.

C. Application Level Firewalls

Application layer firewall is a computer networking firewall operating at the application layer of the TCP/IP protocol stack. These types of firewall systems can be implemented as a piece of software running on a single computer, or a stand-alone piece of hardware. Often, it is a host using various

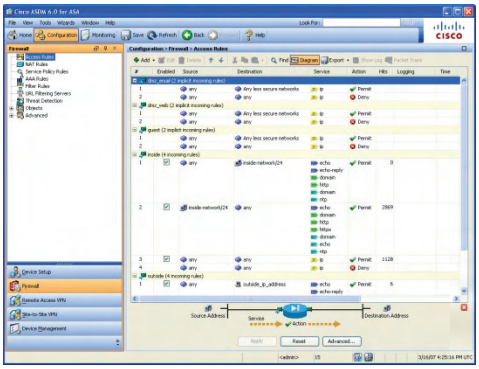


Fig. 1. Firewall policy example

forms of proxy servers to proxy traffic before passing it to a gateway router. Because it acts on the application layer, it may inspect the contents of the traffic, blocking specified content, such as certain websites, viruses, and attempts to exploit known logical flaws in client software.

Application layer firewalls may intercept all packets traveling to or from an application. They block other packets, if configured, dropping them without acknowledgment to the sender. In principle, application firewalls can prevent all unwanted outside traffic from reaching protected machines. One of the newest additions to the host based application level firewalls is the possibility to add rules that would allow or deny access to certain applications. It is also possible to allow different types of privileges to the application. Another branch of application firewalls is the web application firewall. It can be an appliance, server plug-in, or filter that applies a set of rules to an HTTP conversation. Generally, these rules cover common attacks such as Cross-site Scripting (XSS) and SQL Injection. By customizing the rules to your application, many attacks can be identified and blocked. The effort to perform this customization can be significant and needs to be maintained as the application is modified.

Issues regarding the use of application level firewalls might be administrative tasks where they can be very complex for configuration, since they bring extra levels of complexity. This means that the total cost of ownership is much bigger because most of these firewalls need individual maintenance and support. Figure 2 shows the performance comparison between firewalls that filter layer 3, 4 and layer 5 traffic.

B. Identity based firewall systems

In the last few years the windows domain infrastructure has spread rapidly among companies. This has brought the ability to uniquely identify users in this domain, since every user has been given a unique domain user name. Moreover, with the rapid development of new devices, like smart phones, laptops and the lowering of the prices for them, the need of user

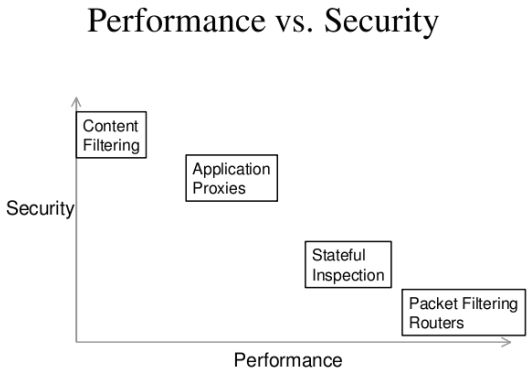


Fig. 2. Performance comparison between different types of firewalls

mobility within these companies has rises. The whole concept of mobility asks that the users are mobile within their work environment, which forces them to use different IP addresses. By using different IP addresses, they are limited in the access to their resources. In order to solve the issues with the mobility, security companies have started to design and develop new concepts that could be implemented into the current firewall technology. These concepts should allow users to extend their mobility and let them be mobile in such ways that they would be able to access resources from where ever they are.

The whole idea behind the identity based firewall concept is to use the domain user name to build firewall rules. This can be done by embedding it with the standard firewall rules, but instead of using a source or destination IP, the rule will use the user name.

III. Analysis of Identity Based Firewalls

A. Analysis of Microsoft Internet Security and Acceleration server

One of the youngest firewall systems around is the Microsoft Internet Security and Acceleration server, or Microsoft ISA. Microsoft Internet Security and Acceleration Server (ISA Server) is described by Microsoft as an "integrated edge security gateway". Originating as Microsoft Proxy Server, ISA is a Firewalling & Security product based on Microsoft Windows primarily designed to securely publish web servers and other server systems, provide stateful, Application-Layer Firewalling, act as a VPN endpoint, and provide Internet Access for client systems in a Business Networking environment.

ISA Server 2006 is a multi-featured and multi-purpose security product that can be deployed in a variety of ways to meet the unique requirements of virtually any organization. As an integrated firewall, Web proxy and VPN server and gateway, ISA Server can be configured to act in each of these roles or be set up to provide only a subset. As an integrated solution, ISA server offers the following features: network layer firewall, an application layer inspection security gateway, forward and reverse Web proxy and caching server, remote access VPN server, site to site VPN gateway.

When it comes to user mobility, the ISA server offers some interesting features in the domain of user mobility. ISA server solves the mobility issues by using a dual formula. It means that the ISA server solves the mobility issues by presenting two solutions for a certain problem.

As said before, ISA server is one of the few servers that solves the user mobility. The technology used to allow user mobility is user based access rules. By introducing user based access rules, ISA server allows the administrators to connect layer 3 or 4 information with layer 5 information, the user name. This type of filtering would allow unique filtering based on user names and group.

The ISA firewall enables this functionality by offering a piece of software to be installed on the clients, which is called a firewall client.

The firewall client software transparently sends user information to the ISA server. This information comprise of user credentials (user name and NTLM hash). In this way, the ISA server has information about the clients in the network and it is possible to build firewall rules based on user names or groups. The user has to be connected with an account from the Active Directory or with a mirrored account on the ISA server. For example, if you have an Active Directory domain, users should log on to the domain, and the ISA Server 2006 firewall must be a member of the domain. The ISA Server 2006 firewall is able to authenticate the user and allows or denies access based on the user's domain credentials.

The other way to enforce user mobility is by using the ISA web proxy. This way, clients don't need to install extra software. This approach brings some limitations, which allow only rules for the web traffic. Every other traffic is filtered the good old fashion way via the use of IP addresses.

The approach that the ISA server uses has some disadvantages. One of the disadvantages is that when the web proxy approach is used, only the web traffic gets the possibility for user mobility. The firewall client approach brings the disadvantage that the ISA firewall is dependent on the firewall client, and vice versa. If the link between the firewall client and the ISA fails, the user of that system will lose the mobility.

Another general disadvantage of the ISA firewall is that is only supports Microsoft Windows environments. User mobility on other operating systems, like Linux, Mac OS X is not support.

D. Analysis of NuFW firewall

Another interesting player from the identity based firewall market is the NuFW. NuFW is a firewall solution developed

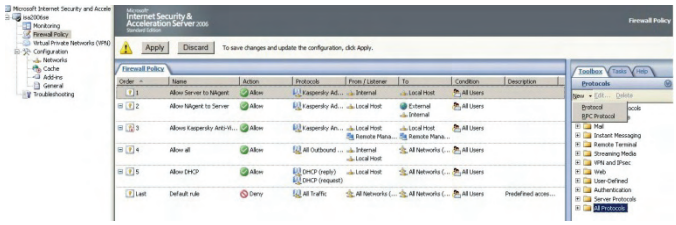


Fig. 3. Example User Based policy on Microsoft ISA 2006

by the french open source company INL. NuFW is based on the Linux operating system. The core of NuFW is the Linux firewall which is extended with the Firewall add-on Netfilter. NuFW extends the previously described Netfilter solution by adding the charm of user identity to the firewall. With NuFW, the firewall permissions follow an authenticated user instead of a PC's address. The NuFW concept allows admins to define not only user based firewall rules, but also rules that enforce policies upon IP addresses. The Identity concept allows that the security policy can be enforced for a user no matter where he or she is on the network. This also lends itself to marrying the firewall with an SSO (single sign on) authentication system. NuFW offers the following features: authenticates any connection that goes through the security gateway or only from/to a chosen subset or a specific protocol, filters packets with criteria such as application and OS used by the users, offers Single Sign On to clients.

In order for NuFW to work, it requires that he client install the NuFW client. Same as with the ISA server, this client sends user information to the NuFW firewall. NuFW can me authenticated by using and LDAP server or by using a database which stores user name information. NuFW uses the

following authentication algorithm, which is shown on figure 3:

NuFW clients can be installed on almost every platform. In general all major platforms are support, Windows, Mac OS X and Linux.

The NuFW approach along with its advantages has some disadvantages. The first disadvantage is the same as with Microsoft ISA server, and it is because NuFW uses agents to authenticate users on the hosts. Another disadvantage is that NuFW generates a big overhead because on every packet is sends authentication request. In high load networks this can be a big issue, because the network will be extra utilized because of NuFW authentication packets.

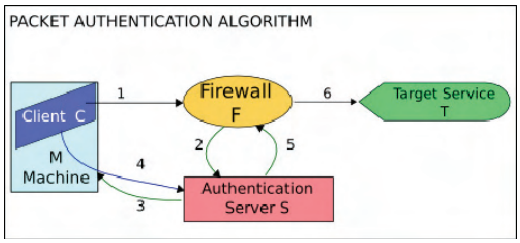


Fig. 3. NuFW authentication algorithm

E. Analysis of Checkpoint User Authority

One of the most advanced solutions regarding identity based firewalls is Checkpoint's User Authority. Check Point UserAuthority brings Web and network applications into one centrally managed security framework by leveraging Check Point's proven networking, encryption and authentication technologies. UserAuthority transparently integrates "best of breed" authentication mechanisms into network applications, enabling intelligent authorization decisions based on a connection's security context: user identity and profile information, encryption and authentication parameters, networking information and desktop security parameters. UserAuthority is the "security glue" that binds Web and network applications to users, Check Point VPN-1/FireWall-1 to create a Secure Virtual Network for the enterprise.

The key features of UserAuthority can be defined as the following: identity based firewall rules, identity based remote access rules, single sign-on to web applications.

The approach that User Authority uses is similar to the ones used with Microsoft ISA server and NuFW. One of the biggest differences that UserAuthority brings into the identity firewall schema is the ability to integrate it with remote access. By doing this, the user gets all of his resources anywhere.

In order to work, UserAuthority needs the network clients to install Checkpoint's SecureAgent. This software is agent based software that sends user credentials to the UserAuthority. By logon, SecureAgent detects the user name and sends the information to the UserAuthority server. When the user tries to pass through a security gateway to access his or her resources, the security gateway asks the UserAuthority server to identify the user that tries to access the resource. Upon successfully verifying the user, the security gateway grants access to the resources.

The disadvantage that this solution brings is the dependency on the SecureAgent, because if the agent doesn't work properly the whole concept of identity based firewall will not work.

IV. Conclusion and Future Work

Figures From the analysis we can conclude that every solution analyzed had the need of using an agent. This can lead to a conclusion that it is the weakness of these identity based firewall because the whole filtering is based upon information that is sent by the clients. If the clients fail to send information, then the whole user mobility concept fails. This brings us to the summary that the identity based firewall technology as it has some flaws.
Our future work on this topic will be to design an identity based firewall which will allow identity based traffic filtering without using software that has to be installed on every client PC. The solution will operating system features which will allow us to filter the traffic by using the user identity. In this way we will achieve that without installing agents on user workstations.
Comparison between the analyzed systems and the system that we will design is shown on table1.

	Microsoft ISA 2006	NuFW	CheckPoint UserAuthority	Our Design
L3, L4 filtering	✓	✓	✓	✓
Identity based filtering without agents	✓	×	×	✓
Identity based filtering with agents	✓	✓	✓	×
Network overhead	×	✓	×	✓
Multiple OS Support	×	✓	×	×

Table 1: comparison chart between the analyzed solutions and our design

References

[1] "CheckPoint UserAuthority", www.checkpoint.com.
[2] Microsoft ISA server 2006, www.microsoft.com
[3] NuFW, www.nufw.org
[4] Amon, C., Shinder, T., Carasik-Henmi, A.: The Best Damn Firewall Book Period, Syngress, 2003
[5] Noonan, W., Dubrawsky, I.:Firewall Fundamentals, Cisco Press, 2006
[6] Zwicky, E., Cooper, S., Chapman, D.:Building Internet Firewalls, O'Reilly Media Inc., 2000.

SESSION CS

Control Systems

Computer Automatic Control for the Engine Test Bench Schenck W-150

Saša Jovanović[1], Milan Milovanović[2], Miroslav Ravlić[3] and Dušan Nestorović[4]

***Abstract* – In this paper, one method for automatisation of the engine test bench and the application for computer controlled execution of test for deposit estimation in the inlet system of petrol engines is shown. It is presented how the test procedure is executed automaticaly by PC with built-in card for control and data acquisition, and also using an appropriate user defined software solution.**

***Keywords* – Automatic control, Data acquisition, Test bench**

I. INTRODUCTION

One big laboratory for engine testing is placed in the Institute for Automobiles – Zastava Car factory, Kragujevac, Serbia. Among the hydraulic test benches, there are also two electrical motor test bench produced by Schenck type W-150, and one motor-generator test bench produced by Siemens, purchased about 40 years ago. These test benches are in very good conditions regarding to mechanic parts and also electronic, analogue computer, because of good maintenance and appropriate handling.

This test bench, Schenck W-150, has a predisposition for automatic control, then, an analysis was done and a proposal to upgrade this for automatic laboratory testing was given.

Methodology for computer data acquisition and automatic control was developed. All necessary parameters for control and observation during the testing were defined. Then, the equipment which has to be built in the laboratory and the other one which can be used on the other places was specified.

II. TEST BENCH AUTOMATIZATION

Engine test bench "Schenck W-150" is with an appropriate electronic system which gives opportunity to control the load by external voltage. So, it was enough to make a new contact for voltage supply from D/A converter in the PC.

Throttle valve control was by hand. To do testing procedure automatically, it was necessary to introduce automatic control. One very simple solution was chosen. Step motor and rotational potentiometer as a throttle position sensor were purchased. Then, control software driver was written. Throttle valve control and throttle position measurements can be done by LPT i.e. PC parallel port [1], or by digital I/O from data acquisition and control PC card [2].

Engine test bench already had a taho generator connected with an analogue indicative instrument. Signal from taho generator is leading to the connecting panel of the acquisition card.

Also, test bench has an already assembled force transducer by "HBM", which is very reliable and high quality, and an appropriate analogue amplifier with digital display. It was necessary, only, to derive analogue voltage signal to the data acquisition card.

To measure fuel consumption, it was necessary to upgrade fuel tank with flush fuel pump and with installation for recirculation. For this purpose, volumetric gauge by "Pierburg" are used, because they do not affect on the fuel pressure in the delivery line.

Air consumption measurement is done using mass flow gauge by hot wire, "Degussa" or air mass flow sensor "Bosch" assembled in serial car production. Both sensors have analogue voltage outputs which are suitable for connecting with data acquisition card.

Ambient conditions, i.e. interior temperature and absolute atmosphere pressure, are measured by one integrated sensor which is in serial car production [3].

To measure engine oil temperature, temperatures of the inlet and outlet water for engine cooling, air temperature in the intake system, etc., type of semi-conductor sensors LM334 and LM335 are used. These sensors give absolute temperature in the range of 0÷100°C. Both of them are suitable for connection with data acquisition PC card, and they are very cheap. It is only necessary to make good, protective, metal cases for these sensors.

Exaust gas temperature is very high, and this can be measured only by thermocouple sensors type NiCrNi or PtRhPt. Signal from these sensors should be amplified before data acquisition card.

Content of exaust gas is measured by portable gas analysers "Pierburg" and "Beckman". From both of them it is possible to connect voltage outputs with data acquisition card.

Stechiometric mixture content, i.e. "Lambda factor", is measured by specific device "Lambda meter" whith an analogue voltage output, proper for connection with data acquisition card.

For data acquisition, i.e. data collecting in the PC and test bench control, PC card "ED2000" [4], is used, because this card can solve all problems, and we have a great experiance using it during a couple of years, and big library for various applications which are already done.

[1]Saša Jovanović is with the Zastava automobiles, Kosovska 4, Kragujevac, Serbia, E-mail: piter@kg.ac.rs

[2]Milan Milovanović is with the Zastava automobiles, Kosovska 4, Kragujevac, Serbia, E-mail: siljakg@sbb.rs

[3]Miroslav Ravlić is with the Prizma, Z.Đinđića bb, Kragujevac, Serbia, E-mail: m.ravlic@prizma.co.rs

[4]Dušan Nestorović is with the VTS Kragujevac, Kosovska 12, Kragujevac, Serbia, E-mail: dunestor@gmail.com

III. TESTING GASOLINE ADDITIVES

For improving the quality of gasoline produced for public distribution, i.e. for improving all characteristics of OTO engines (torque, power, speed, fuel consumption, exaust gas emission, etc.), petrol producers in cooperation with car producers, intensively are working and developing new methods and procedures for qualification of petrol quality.

Based on the existing method by Coordination European Council (CEC F–04–A–87), NIS (Serbian Petrol Industry) gave approval to Zastava Car factory to develop test method for qualification of gasoline with new additives quality [5]. This test method rates the quality of gasoline with new additives regarding to the deposits which this fuel apply onto the intake system elements (i.e. carburettor, inlet manifold, inlet tunnels, inlet valves and combustion chamber), for the period of 40 hours, see Table 1.

For execution "CEC test", i.e. for testing gasoline with different additives, test bench was modified and upgraded. Because of specific test conditions, a simple device for intake air heating to the temperature of 88 ± 2 ^{o}C is designed and realized. This device is consisted of one tank with inlet and outlet air channel. Outlet air is distributed to the engine intake manifold and air filter inlet. Tank has 5 electric heaters, each power of 1.1 *kW* which are in switching ON/OFF mode, and one heater, with same power, has regulation by thyristor. Block scheme of the complete system is shown in Fig 1.

Before the test starts, it is necessary to take engine warn-up stage, until the working conditions, and if entirely new engine, for the first time, is on the test bench, then it is necessary to take running-in stage for new engine. "CEC test" itself, is executing on the test bench with stimulated testing conditions, i.e. engine should suck hot air, which is heated to the temperature of the 88 ^{o}C.

TABLE I
STAGES FOR CEC TEST

Stage	Duration (min)	Engine Speed (rev./min)	Moment (Nm)	Intake Air Temperat. (^{o}C)
1	0.5	1200 ± 50	2.0 max	88 ± 2
2	1.0	3000 ± 50	35.3 ± 2	88 ± 2
3	1.0	1300 ± 50	29.6 ± 2	88 ± 2
4	2.0	1850 ± 50	32.5 ± 2	88 ± 2

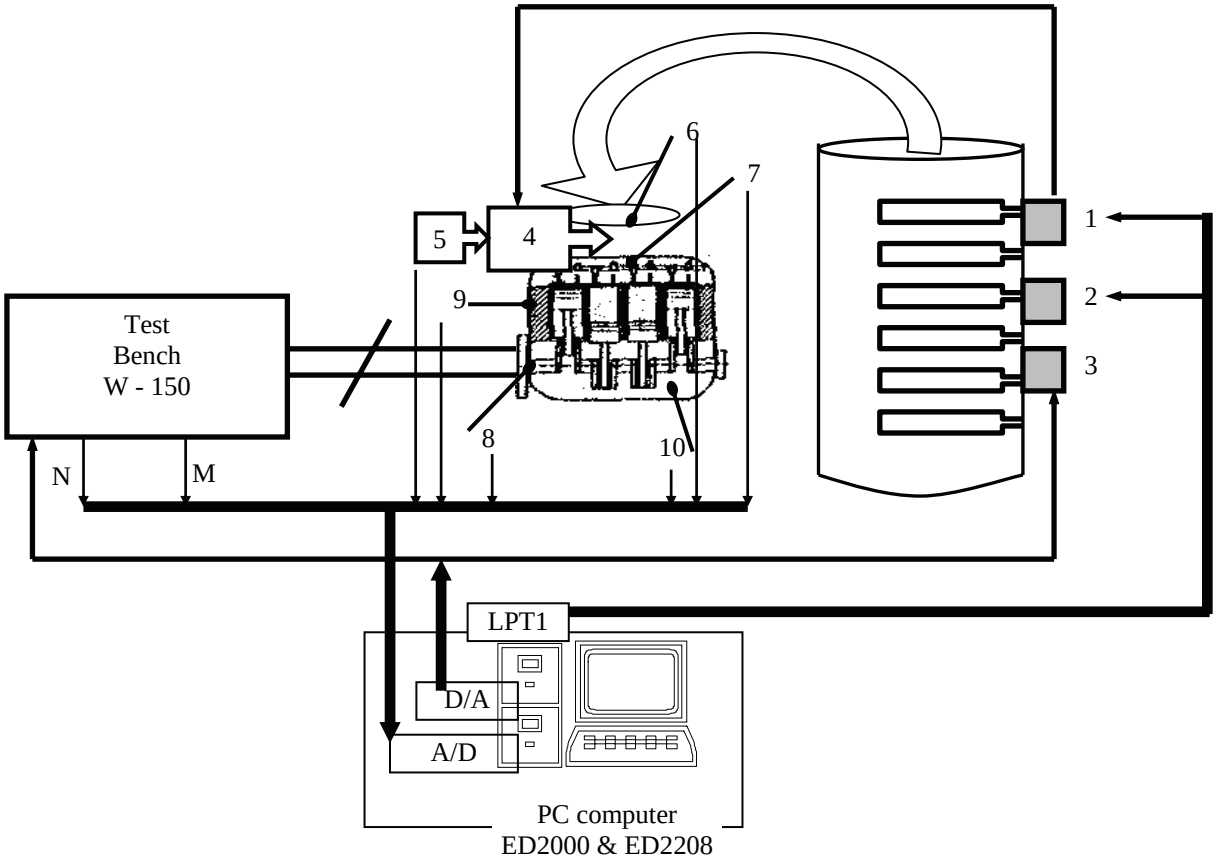


Fig. 1. Block scheme of automatized test bench W-150 ready for execution "CEC test":
1. Step motor driver; 2. Electric relay for heaters; 3. Regulation by Thyristor; 4. Step motor with reductuon gear; 5. Rotational potenciometar; 6. Intake air temperature sensor; 7. Intake air pressure sensor; 8. Engine cooling water inlet temperature sensor; 9. Engine cooling water outlet temperature sensor; 10. Engine oil temperature sensor.

IV. APPLICATION SOFTWARE

PC cards for data acquisition and control ED2000 are using by engineers at Zastava automobiles for a long period. For comfortable working and implementation these cards, a drivers library were writen in Turbo Pascal, firstly, and then also in Delphi programming language.

Test execution is completely automated. Appropriate application software, writen in Delphi programming language, using PC parallel port and A/D, D/A cards for data acquisition and control, ED2000 and ED2208, respectively, and actuators, is controlling test bench working and throttle opening. By controlling the throttle at engine carburettor, the change of engine mode in desired moment can be done automaticaly.

All measured data are aquired and saved in data file with one second sampling time.

Simultaneously with engine mode changing and holding, application software takes care about intake air temperature retaining. Heaters relay are switched on/off by parallel port LPT1, and one output of the D/A card ED2208 is used to control thyristor regulator, for fine regulation of the power of one heater. The response of the intake air temperature changing is very slow regarding to the moment of heater switching on. Because of that, it is very difficult to make very good and successful regulation if disturbances are very frequent and impulsive.

The fact is that stages for CEC test are known in advance, and that unexpected disturbances are very rare. This simplifies control algorithm for air temperature. Regulation is done in the closed loop, but there is a state prediction which makes minimal variations of the temperatures in the transient modes.

The application for users and CEC test operators is shown in Fig 2. This application is used for control and data acquisition. When application software is started, after iPC cards initiations, step motor firstly is closing the throttle. This is done with predefined and sufficient number of steps so the throttle is reliable closed. After that, the throttle is opened to maximal position. By this way, the exact throttle position can be determinated later.

Hereafter, the application is positioned in the stage "ručni rad", i.e. "manual operation". In this stage two scroll bars are opened (below the heaters status in Fig 2.). One scroll bar is for setting the reference engine speed (i.e. the load for test bench) and the other one is for setting opening of throttle. In this way, the basic function of the test bench is retained, and it is possible to start the engine for the first time. Also, in this mode it is possible to switch on/off five heaters by pressing the buttons "Povecaj" and "Smanji", i.e. "Increase" and "Decrease", respectively.

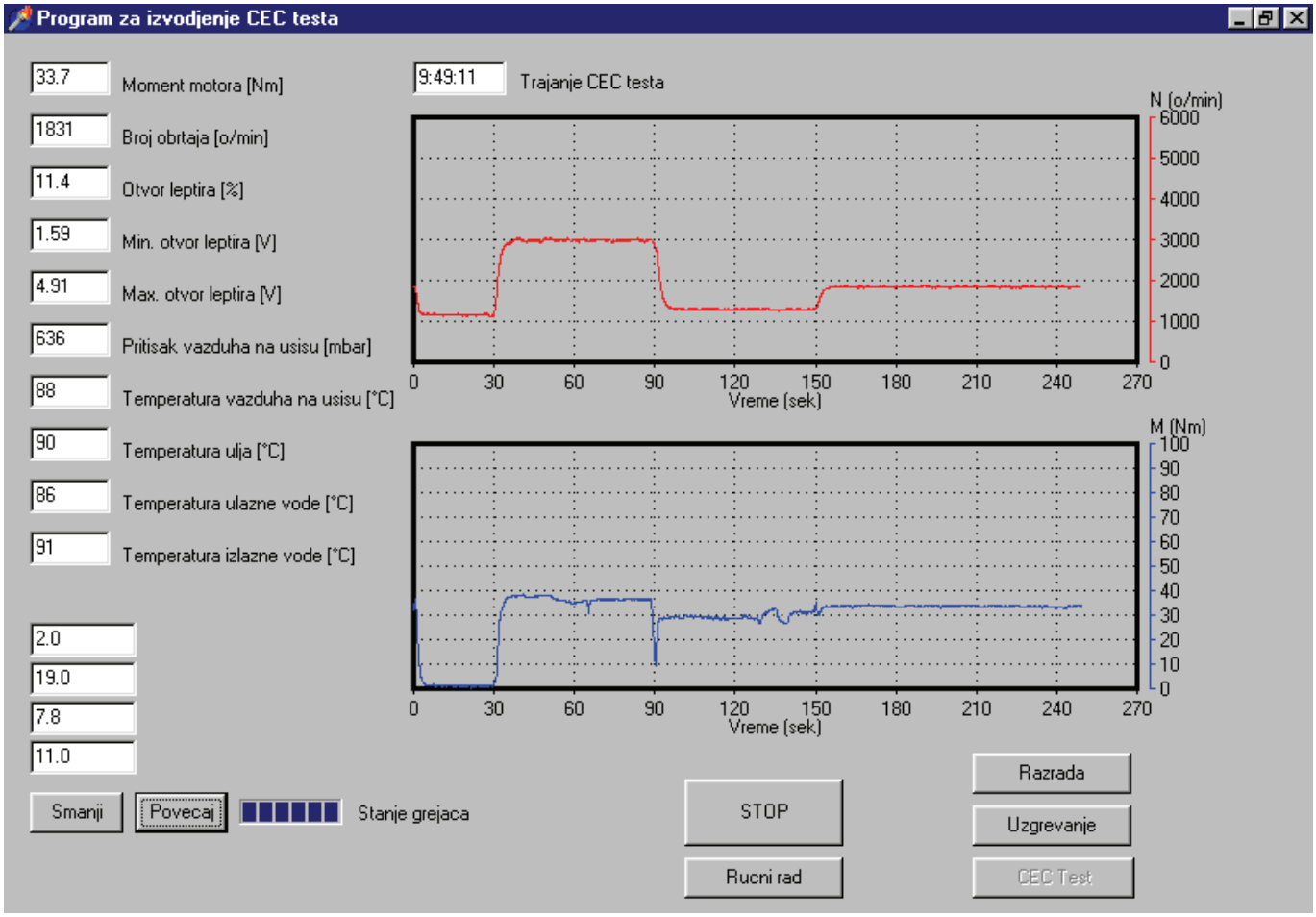


Fig. 2. Application for CEC test execution at the engine test bench

After the automatic mode starting, (stages "Uzgrevanje" i.e. warm-up, "Razrada" i.e. run-in or "CECTest", the possibility for setting up engine speed or throttle opening are not allowed. Also, buttons to heater switch on/off have no function, but they are showing the status for active heaters.

When new engine is mounted for the first time at the test bench, automatic work is runing in the mode "Razrada", i.e. run-in. Operating conditions are specially defined.

Mode "Uzgrevanje" i.e. warm-up, is also specially defined, and it is used always before the CEC test. When operating engine conditions are fulfilled, the application automatically is turning off test engine because of necessity to measure the weight of engine oil. After that, by established methodology, the test engine is starting again and the mode "CEC Test" for automatic operation is activated.

PC card ED2000 can contain eight differential analogue inputs, then this number of most important inputs are taken for data acquisition in the data file and are shown on the display. Also, these data are serving for test condition checking.

As an illustration of working in laboratory for engine testing, the photo of test engine, produced by DMB Serbia, 1301ccm, assembled carburettor with two throats, and some parts of the equipement is shown in Fig. 3.

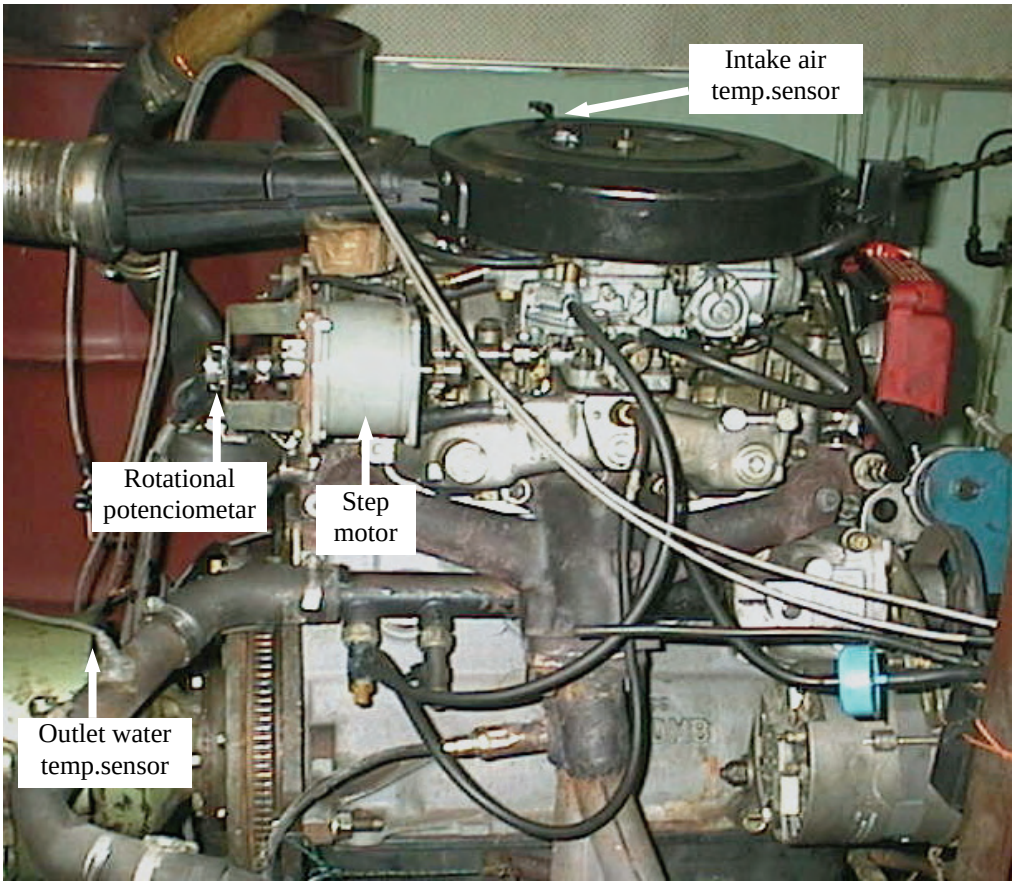


Fig. 3. Test engine photo

V. Conclusion

This example shows how it is possible to completely automate worth and specific lab equipement, with very low investment. Hereafter, it is possible to use this equipement for various experiments and tests, only making an adequate software. This software is user friendly, and training for operators is very short and simple.

Because of its capabilities, test method shown in this paper, can be used not only for qualification of gasoline with different additives, but also for testing and qualification all engine fluids. For example, different engine oils can be tested using the same fuel. Qualification can be done in the same way as for fuels, by measuring and comparing accumulated deposits into the intake system.

References

[1] S. Jovanović, "Control and Data Acquisition using Computer Standard Parallel Port", YUINFO 2003, Proceedings on CD, Kopaonik, Serbia, 2003.

[2] R. Isermann, "*Digital Control Systems*", Springer-Verlag, Berlin, Germany, 1989.

[3] Bosch Gmbh, "*Automotive Handbook*", 3rd Edition, VDI-Verlag, Düsseldorf, Germany, 1993.

[4] Electronic Design, "*ED2000 User's Manual*", Beograd, Serbia, 1993.

[5] D. Nestorović, S. Jovanović and P. Šundović, "Methodology for Testing the Quality of Gasoline on the Engine Test Bench", YUNG 4P 2000, Proceedings, Vrnjačka Banja, Serbia, 2000.

Design of Information and Communications System Based on 3D MEMS Inertial Sensors

Emil Iontchev[1], Ivaylo Simeonov[2] and Rosen Miletiev[3]

Abstract – **The current paper represents the design of the information and communication system which is based on MEMS inertial sensors, which may be useful for a control of the moving objects such as unmanned airplanes, robots, etc. It consists from 3D inertial accelerometers, 3D angular rate sensors (gyroscopes), 3D magnetometer and 3D thermometer. Also it may be extended with different type of sensors such as three-axis accelerometers, one axis gyroscope, inclinometer sensor, pressure sensors, etc. to monitor the acceleration or relative shift of the selected parts of the moving object. The system also has capability to realize a M2M functions due to the integrated GPS receiver and GSM modem. The block diagram and the communication protocol review are accomplished in the paper.**

Keywords – **IMU, MEMS, inertial sensors**

I. Introduction

Interest in the development of microelectro-mechanical systems (MEMS) has mushroomed during the past decade. In the most general sense, MEMS attempts to exploit and extend the fabrication techniques developed for the integrated circuit (IC) industry to add mechanical elements to the electrical circuits to make integrated microsystems for perception and control of the physical world. MEMS sensors allow the implementation of a lot of different functions, as free-fall detection, car navigation, map browsing, gaming, menu scrolling, motion control, vibration monitoring, antitheft and many others. The MEMS sensors are used to measure the frequency, amplitude (strength) and spectrum (signature) of vibrations, enabling the ability to perform active monitoring of moving objects such as unmanned autonomous vehicles [1-3]. MEMS sensors are also widely used in the integrated navigation systems [4-5] by using Extended Kalman filter.

The current paper represents a design of the information and communications system, based on 3D MEMS inertial sensors, which is capable of measurement and control of many individual moving parts and the entire body of the selected object (robot, vehicle, etc.). The paper discusses the used communication protocols, block diagrams of the information and communications system and sensor PCB. Also the installed MEMS inertial sensors are mentioned.

[1]Emil Iontchev is with the Higher School of Transport "T. Kableshkov" 158 Geo Milev Street, Sofia 1574, Bulgaria, E-mail: e_iontchev@yahoo.com

[2]Ivaylo Simeonov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: ivosim@abv.bg

[3]Rosen Miletiev is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria. E-mail: miletiev@tu-sofia.bg

II. System Block Diagram

The system block diagram is based on the following main blocks (Fig.1):

- Control and communication unit (CCU)
- Inertial measurement unit (IMU)
- Single Board Computer (SBC)
- Database server (DBS)
- Sensors

The functional connections between modules are established by digital interfaces, shown on the diagram. Their choice is based on the following criteria:

- Number of connected nodes
- Transmission speed
- Communication channel noise immunity
- End terminal device distance

The main system module is recognized as inertial measurement unit (IMU) where is installed the main 3D inertial sensor. It consists of 3D 14bit accelerometer, 3D 14bit angular rate sensor, 3D 14bit magnetometer and 3D 12bit thermometer. These sensors are integrated in the single chip solution – ADIS16405 produced by Analog Devices. The sensor data are transmitted to the PIC microcontroller via SPI interface. This bus also is connected to the main FLASH memory (MMC/SD card up to 2GB) and 14bit inclinometer with 2 axes accelerometer type ADIS16209. The inclinometer data are compared with the Extended Kalman Filter (EKF) output to calculate the filter response and accuracy.

The IMU block also contains a high sensitive GPS receiver type LEA 5S, produced by UBLOX. This receiver has 50-channel u-blox 5 engines with over 1 million effective correlators, SuperSense® Indoor GPS: -160 dBm tracking sensitivity and also supports AssistNow Online and AssistNow Offline A-GPS services and SBAS (WAAS, EGNOS, MSAS, GAGAN). The navigation data are sent via UART interface to the PIC microcontroller while the same data are transmitted via high speed USB interface to the single board computer (SBC) according to NMEA 0183 protocol. These data may also be sent to the SBC via RS232 interface with a hardware handshaking (RTS/CTS). This scheme allows sending navigation data via two independent communication channels to the SBC.

The IMU block also communicates with external sensors via high speed CAN interface, which allows connecting up to 112 nodes to the bus. These additional sensors send data about accelerations of the other object parts. The environment parameter data such as pressure and temperature are also sent to the microcontroller. The system stability and security is highly increased if the identical sensors are installed to the system due to the sensor reservation.

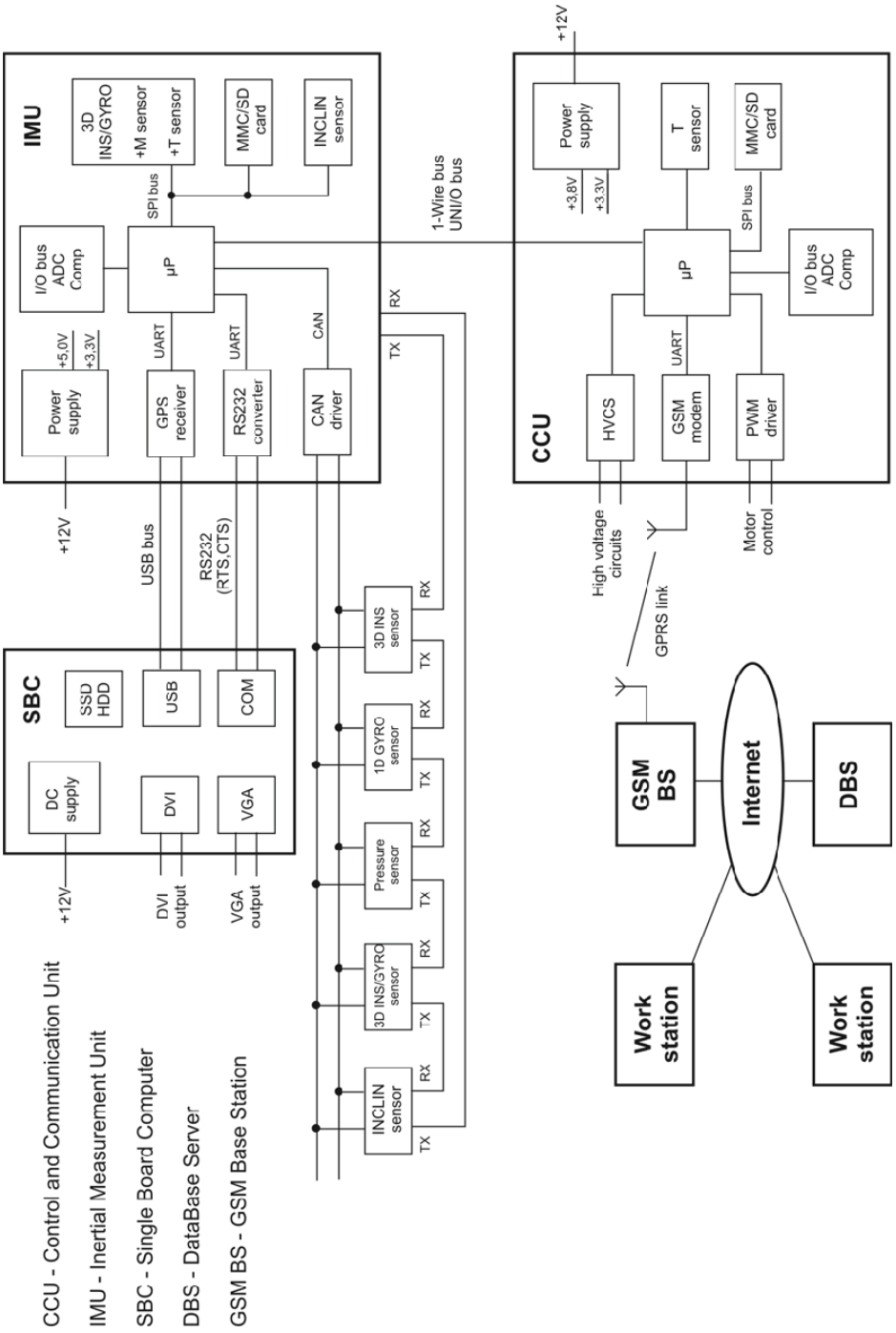


Figure 1. System block diagram

This system performance may be also achieved by using the secondary UART interface. The RX pin of this interface of one device is connected to the TX pin to another and the system commands are sent consecutively to the bus.

The IMU block may support different types of sensors, such as:

- 3-axis 12bit accelerometer (3D INS) type LIS3LV02DQ
- 1-axis 14bit angular rate sensor (1D GYRO) type ADIS16100
- Pressure sensor type SPD015A
- 3-axis 14bit accelerometer (3D INS) and 3-axes 14bit angular rate sensor (3D GYRO) with extended temperature compensation type ADIS16355 or ADIS16405
- 14bit inclinometer with 2 axis accelerometer type ADIS16209.

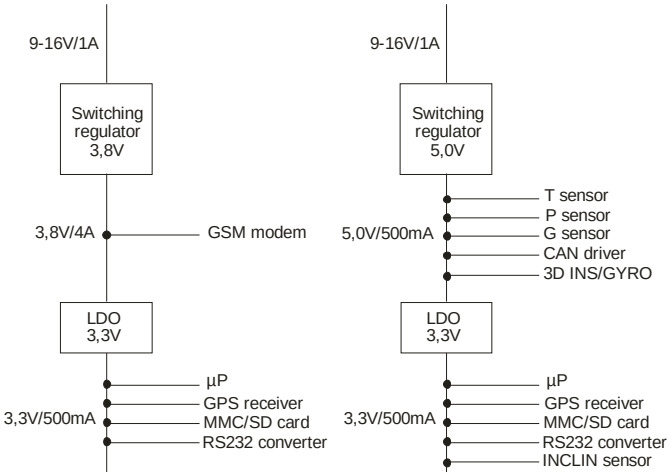


Figure 2. System supply diagram

The IMU block also supports different types of analog extensions such as 10bit ADC, 2-channel comparator with an internally generated Vref and general purpose I/O pins (GPIO).

The system supply diagram is shown at Figure 2. the external supply (9÷16V) is fed to the switching regulators, which forms 3.8V/4A to supply the GSM modem and 5.0V/500mA to supply the 3-axis inertial sensor and CAN driver. The additional LDO regulators produce a stable 3.3V power supply for onboard microcontrollers, GPS receiver, MMC/SD card and other inertial sensors.

The IMU data or states (pitch, yaw and roll information) are sent to the communication unit CCU via 1-Wire interface. According to the received information the CCU block made decisions to control the external devices such as motors via PWM driver, valves or another high voltage devices or circuits (HVCs), etc. The gathered data may also be transmitted to the database server via GPRS link over the mobile network. This function implements a M2M connected to established real-time telemetry of the remote device. The database data may be read via secure SSL channel by another remote user.

The communication and control functions of the system are performed by the CCU module.

The block diagram of the sensor unit is shown at Figure 3. It consists of a CAN transceiver MC2551 produced by Microchip Inc., microcontroller (µP), power supply and MEMS inertial sensor. The communication protocol between the sensor and microcontroller is SPI based. The sensor unit also the integrated USART module of the microcontroller to established serial connection with other sensors or to personal computer (PC) via software emulated serial protocol. It supplies by the external 5V power supply which is fed directly to the CAN transceiver and microcontroller through input LC low pass filter, which reduces the high frequency noise. The MEMS sensor is supplied by single ended 3,3V power supply which is generated by the Low Quiescent Current LDO MCP1700 in 3-Lead Plastic Small Outline Transistor SOT23 package.

III. System application

The proposed system may be used such as control, monitoring and communication system for a humanoid robot system (Figure 4). The shown figure represents the main moving parts of the robot while the system unit (ICS) is installed in the robot body.

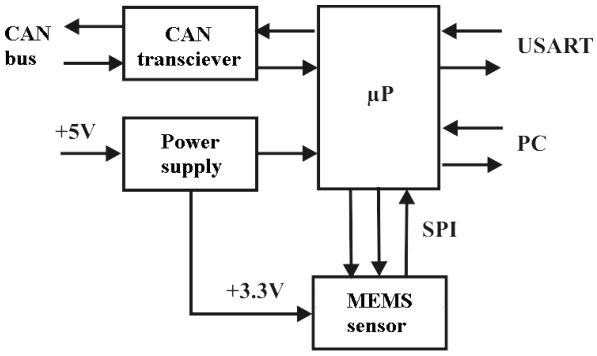


Figure 3. Sensor block diagram

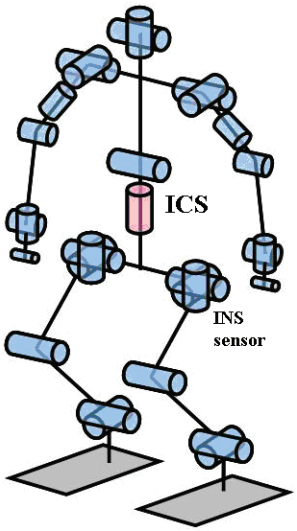


Figure 4. Robot's kinematic

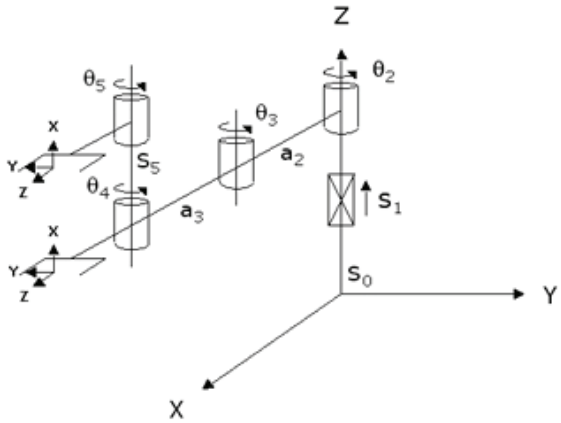


Figure 5. XYZ coordinate system orientation

If the xyz coordinate system of the sensor unit is orientated randomly according to the XYZ coordinate system of the body (Figure 5), the sensor accelerations according to the XYZ system may be calculated according to the Euler transformations [6]:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = R_z(\chi) R_N(\theta) R_Z(\phi) \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \quad (1)$$

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -\sin\chi & \cos\chi \\ 0 & \cos\chi & \sin\chi \end{pmatrix} \begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \quad (2),$$

where χ, ϕ and θ represents the rotation angles (Figure 6).

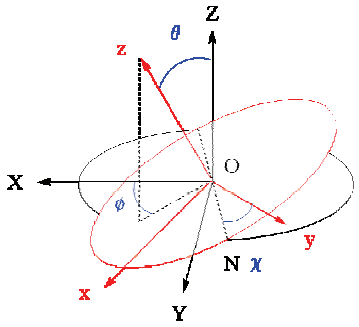


Figure 6. Rotation angles

The accelerations may be also used for calculation of the absolute speed and shifting of the selected part or relative speed and shifting between two selected nodes. These parameters are calculated by a numerical integration of the measured accelerations according to the following equations:

$$\begin{aligned} v_{i+1,x} &= v_{i,x} + (a_{i+1,x} - a_{i,x})\Delta t \\ v_{i+1,y} &= v_{i,y} + (a_{i+1,y} - a_{i,y})\Delta t \\ v_{i+1,z} &= v_{i,z} + (a_{i+1,z} - a_{i,z})\Delta t \\ S_{i+1,x} &= S_{i,x} + v_{i,x}\Delta t + \frac{1}{2}(a_{i+1,x} - a_{i,x})(\Delta t)^2 \\ S_{i+1,y} &= S_{i,y} + v_{i,y}\Delta t + \frac{1}{2}(a_{i+1,y} - a_{i,y})(\Delta t)^2 \\ S_{i+1,z} &= S_{i,z} + v_{i,z}\Delta t + \frac{1}{2}(a_{i+1,z} - a_{i,z})(\Delta t)^2 \end{aligned} \quad (3).$$

It is clearly visible that the integration error may be reduced while the sampling frequency is increased. The chosen 3D inertial sensor is capable to measure accelerations up to 2560 times per second.

Other application of this system might be to assess of track geometry and quality by prediction of the vehicle dynamics. Recently, interest has grown for improving the conventional track geometry inspection methods by use of a real-time simulation system, which provides performance-based assessments to identify track sections that are likely to produce dynamically high and/or unsafe vehicle response.

When the system is applied to railway track the most important signals of the vehicle response are the wheel-rail contact forces and the accelerations of the vehicle body. With two acceleration sensors witch are attached to the bearings of the axle we can measure of corrugation and other isolated faults on the rail surface. Two acceleration sensors attached to the floor of the vehicle together with sensors recording the relative distance of the axle to the vehicle floor form the basis for level and alignment signals. Lateral acceleration sensors and gyros produce the raw data for a spatial curve. The combination of INS sensors with a GPS allows for an absolute and kinematics three-dimensional description of the track inventory.

The system can measure the rail acceleration while the train passed through the place where inertial sensors are placed. The signals from them can be used in railway diagnostics to the detection of wheel flat faults.

IV. Conclusion

The current paper represents a design of the information and communications system, based on 3D MEMS inertial sensors, which is capable of inertial measurement of many individual moving parts and compare them to the movement of the entire body of the selected object (robot, vehicle, etc.).

This allows calculation of the velocity and shifting of the moving parts in the XYZ coordinate system and establishment of the object position and vector movement. The system is based on the latest digital MEMS accelerometers and gyroscopes and supports many eternal devices such as other digital systems or analog sensors and is capable to send all data to the remote device due to the integrated GSM modem.

The developed information and communication system is a foundation of the next generation digital three-dimensional IMU systems and Event Data Recorders (EDRs), according to IEEE1616A standard.

Acknowledgement

This paper was prepared and supported by the National Fund under contract number No.DTK02/2-2009.

References

[1]. Peter Gibbens, Ben Grocholsky, Keith Willis, Hugh F. Durrant-Whyte - *A Low-Cost, Redundant Inertial Measurement Unit for Unmanned Air Vehicles*, International Journal of Robotics Research, Vol. 19, No. 11, 1089-1103 (2000)

[2]. Jan Wendel, Oliver Meister, Christian Schlaile and Gert F. Trommer - *An integrated GPS/MEMS-IMU navigation system for an autonomous helicopter*, Aerospace Science and Technology, Volume 10, Issue 6, September 2006, Pages 527-533

[3]. J.F. Vasconcelos, J. Calvario, P. Oliveira,C. Silvestre - GPS Aided IMU for unmanned air vehicles, 5th IFAC/EURON Symposium on Intelligent Autonomous Vehicles, Instintro Superior Tecnico, Lisboa, Portugal, July 5-7, 2004

[4]. Alison K. Brown, Yan Lu - *Performance Test Results of an Integrated GPS/MEMS Inertial Navigation Package*, Proceedings of ION GNSS 2004, Long Beach, California, September 2004, pp.1-8

[5]. Sungsu Park Chin-Woo Tan - *GPS-Aided Gyroscope-Free Inertial Navigation Systems*, University of California, Berkeley, 2002, Paper UCB-ITS-PRR-2002-22

[6]. G. Arfken, *Mathematical Methods for Physicists*, 3rd. ed., Academic Press: New York 1985

Full Perturbation Analysis of the Discrete-time LMI Based H_∞ Quadratic Stability Problem

Andrey S. Yonchev[1], Mihail M. Konstantinov[2] and Petko H. Petkov[3]

***Abstract* – The quadratic H_∞ control ensuring closed-loop performance γ can be implicitly realized by the solutions *Q, Y* of a system of linear matrix inequalities (LMIs). The paper is concerned with performing linear full perturbation analysis for the discrete-time LMI based H_∞ quadratic stability problem. The sensitivity analysis of the perturbed matrix inequalities is considered in a similar manner as for perturbed matrix equations, after introducing a suitable right hand part, which is slightly perturbed. The proposed approach leads to tight linear perturbation bounds for the LMIs' solutions to the H_∞ quadratic stability problem. Numerical example is also presented**

***Keywords* – Full perturbation analysis, H_∞ quadratic stability, LMI based synthesis, Linear systems.**

I. INTRODUCTION

In many control problems, the design constraints have a simple reformulation in terms of linear matrix inequalities (LMIs). This is hardly surprising, given that LMIs are direct byproducts of Lyapunov based criteria, and that Lyapunov techniques play a central role in the analysis and control of linear systems, see [1,2] and the literature therein.

The H_∞ control problem is a good illustration of this point. Indeed, the H_∞ constraints can be expressed as a single matrix inequality via the bounded real lemma [3]. Even though the H_∞ control problem has a solution in terms of Riccati equations [4], the LMI approach remains valuable for several reasons. First it is applicable to all plants without restrictions on infinite or pure imaginary invariant zeros. Secondly, it offers a simple and insightful derivation of the Riccati based solvability conditions [5]. In addition, the LMI based H_∞ control is practical thanks to the availability of efficient convex optimization algorithms, based on the interior point method [6], and software [7].

In this paper we propose an approach to perform full linear perturbation analysis of the LMI based H_∞ quadratic stability problem via introducing a suitable right hand part in the considered matrix inequalities.

We use the following notations: $R^{m\times n}$ - the space of real $m\times n$ matrices; $R^n = R^{n\times 1}$; I_n - the identity $n\times n$ matrix; e_n - the unit $n\times 1$ vector; M^T - the transpose of M; $M^\dagger$ - the pseudo inverse of M; $\|M\|_2 = \sigma_{\max}(M)$ - the spectral norm of M, where $\sigma_{\max}(M)$ is the maximum singular value of M; $vec(M) \in R^{mn}$ - the column-wise vector representation of $M \in R^{m\times n}$; $\Pi_{m,n} \in R^{mn\times mn}$ - the vec-permutation matrix, such that $vec(M^T) = \Pi_{m,n} vec(M)$; $M \otimes P$ - the Kroneker product of the matrices M and P. The notation ":=" stands for "equal by definition".

The remainder of the paper is organized as follows. In Section 2 we shortly present the problem set up and objective. Section 3 describes the performed linear full perturbation analysis of the LMI based H_∞ control problem. Section 4 presents a numerical example before we conclude in Section 5 with some final remarks.

II. PROBLEM SET UP AND OBJECTIVE

Consider the linear discrete-time system

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k), \\ y(k) &= Cx(k) + Du(k) \end{aligned} \tag{1}$$

where $x(k) \in R^n$, $u(k) \in R^m$, and $y(k) \in R^r$ are the system state, input and output vectors respectively, and A, B, C, D are constant matrices of compatible size.

We consider an LMI approach to solve the H_∞ quadratic stability problem, as stated in [8]

$$\begin{bmatrix} -P^{-1} & A & B & 0 \\ A^T & -P & 0 & C^T \\ B^T & 0 & -\gamma I & D^T \\ 0 & C & D & -\gamma I \end{bmatrix} < 0, P > 0, \tag{2}$$

which is actually an Eigenvalue Problem (EVP) with respect to the variables P and γ. Here we assume that the optimal closed-loop performance γ_{opt} of the system (1) is already obtained.

In order to obtain quadratic H_∞ stability and to ensure closed-loop performance γ it is necessary to design a state-feedback control $u=Kx$. To transform LMI (2) we apply Schur complement argument [9] to obtain the following inequality:

$$\begin{bmatrix} -P^{-1} & (A+BK) & 0 & 0 \\ (A+BK)^T & -P & 0 & (C+DK)^T \\ 0 & 0 & -\gamma I & 0 \\ 0 & C+DK & 0 & -\gamma I \end{bmatrix} < 0, P > 0, \tag{3}$$

with respect to the variables K, P and γ. We pre-and post-multiply inequality (3) by $diag\{I, P^{-1}, I, I\}$ and introduce change of variables such that $Q=P^{-1}$ and $Y=KP^{-1}$ to obtain the following system of LMIs:

[1]Andrey S. Yonchev is with the Faculty of Automatics, Technical University, 1000 Sofia, Bulgaria, E-mail: ajonchev@mail.bg

[2]Mihail M. Konstantinov is with the Faculty of Mathematics, UACG, 1046 Sofia, Bulgaria, E-mail: mmk_fte@uacg.bg

[3]Petko H. Petkov is with the Faculty of Automatics, Technical University, 1000 Sofia, Bulgaria, E-mail: php@tu-sofia.bg

$$\begin{bmatrix} -Q & (AQ+BK) & 0 & 0 \\ (AQ+BY)^T & -Q & 0 & (CQ+DY)^T \\ 0 & 0 & -\gamma I & 0 \\ 0 & CQ+DY & 0 & -\gamma I \end{bmatrix} < 0, Q > 0, \quad (4)$$

The main objective of the paper is to perform a linear sensitivity analysis of the LMI system (4) near the optimal value of γ, needed to solve the H_∞ quadratic stability problem.

Suppose that the matrices *A, B, C, D* are subject to perturbations $\Delta A, \Delta B, \Delta C, \Delta D$ and assume that they do not change the sign of the LMI system (4). The full perturbation analysis of the discrete-time LMI based H_∞ quadratic stability problem is aimed at determining perturbation bounds of the LMIs (4) near the optimal value of γ, as functions of the perturbations in the data *A, B, C, D* and in $\gamma_{opt.}$

III. Linear Full Perturbation Analysis

We perform perturbation analysis of the LMI (4) for the discrete-time system (1)

$$\begin{bmatrix} -(Q+\Delta Q) & ABQY^T & 0 & 0 \\ ABQY & -(Q+\Delta Q) & 0 & CQDY^T \\ 0 & 0 & -(\gamma+\Delta\gamma) & 0 \\ 0 & CQDY & 0 & -(\gamma+\Delta\gamma) \end{bmatrix} < 0, \quad (5)$$

where $ABQY^T = (Q+\Delta Q)(A+\Delta A)^T + (Y+\Delta Y)^T (B+\Delta B)^T$

$ABQY = (A+\Delta A)(Q+\Delta Q) + (B+\Delta B)(Y+\Delta Y)$,

$CQDY^T = (Q+\Delta Q)(C+\Delta C)^T + (Y+\Delta Y)^T (D+\Delta D)^T$,

$CQDY = (C+\Delta C)(Q+\Delta Q) + (D+\Delta D)(Y+\Delta Y)$. We have to study the effect of the perturbations $\Delta A, \Delta B, \Delta C, \Delta D$ and $\Delta\lambda$ on the perturbed LMI solutions $Q^* + \Delta Q$ and $Y^* + \Delta Y$, where Q^*, Y^* and $\Delta Q, \Delta Y$ are the nominal solution of the inequality (4) and the perturbations, respectively. The essence of our approach is to perform perturbation analysis of the inequality (4) in a similar manner as for a proper matrix equation after introducing a suitable right hand part, which is slightly perturbed. Thus for LMI (5) we have:

$$\begin{bmatrix} -(Q^*+\Delta Q) & ABQY^{*T} & 0 & 0 \\ ABQY^* & -(Q^*+\Delta Q) & 0 & CQDY^{*T} \\ 0 & 0 & -(\gamma_{opt}+\Delta\gamma) & 0 \\ 0 & CQDY^* & 0 & -(\gamma_{opt}+\Delta\gamma) \end{bmatrix} = L^* + \Delta L_1 < 0, \quad (6)$$

where $ABQY^{*T} = (Q^*+\Delta Q)(A+\Delta A)^T + (Y^*+\Delta Y)^T (B+\Delta B)^T$,

$ABQY^* = (A+\Delta A)(Q^*+\Delta Q) + (B+\Delta B)(Y^*+\Delta Y)$

$CQDY^{*T} = (Q^*+\Delta Q)(C+\Delta C)^T + (Y^*+\Delta Y)^T (D+\Delta D)^T$,

$CQDY = (C+\Delta C)(Q^*+\Delta Q) + (D+\Delta D)(Y^*+\Delta Y)$ and L^* is obtained using the nominal LMI

$$\begin{bmatrix} -Q^* & Q^*A^T + Y^{*T}B^T & 0 & 0 \\ AQ^* + BY^* & -Q^* & 0 & Q^*C^T + Y^T * D^T \\ 0 & 0 & -\gamma_{opt} & 0 \\ 0 & CQ^* + DY^* & 0 & -\gamma_{opt} \end{bmatrix} = L^* < 0, \quad (7)$$

The matrix ΔL_1 is due to the data and closed-loop performance perturbations, the rounding errors and the sensitivity of the interior point method that is used to solve the LMIs.

Using the relation (7) the perturbed equation (6) may be written as

$$\Delta_Q + \Omega_Q = \Delta L_1, \quad (8)$$

where

$$\Delta_Q = \begin{bmatrix} -\Delta Q & \Delta Q A^T & 0 & 0 \\ A\Delta Q & -\Delta Q & 0 & \Delta Q C^T \\ 0 & 0 & 0 & 0 \\ 0 & C\Delta Q & 0 & 0 \end{bmatrix},$$

$$\Omega_Q = \begin{bmatrix} 0 & Q^*\Delta A^T + \Delta Y^T B^T + Y^{*T}\Delta B^T & 0 & 0 \\ \Delta A Q^* + B\Delta Y + \Delta B Y^* & 0 & 0 & Q^*\Delta C^T + \Delta Y^T D^T + Y^{*T}\Delta D^T \\ 0 & 0 & -\Delta\gamma I & 0 \\ 0 & \Delta C Q^* + D\Delta Y + \Delta D Y^* & 0 & -\Delta\gamma I \end{bmatrix},$$

Here the terms of second and higher order are neglected. The relation (8) may be written in a vector form as

$$vec(\Delta_Q) + vec(\Omega_Q) = vec(\Delta L_1), \quad (9)$$

where

$vec(\Delta_Q) = [-I, A\otimes I, 0, 0, I\otimes A, -I, 0, C\otimes I, 0, 0, 0, 0, 0, I\otimes C, 0, 0]^T$

$*vec(\Delta Q) = T\Delta q,$

$vec(\Omega_Q) =$

$$= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ (I\otimes Q^*)\Pi_{n^2} & (B\otimes I)\Pi_{n\times m} & (I\otimes Y^{*T})\Pi_{m^2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ (Q^*\otimes I) & (I\otimes B) & (Y^*\otimes I) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & (D\otimes I)\Pi_{n\times m} & 0 & (I\otimes Q^*)\Pi_{n^2} & 0 & (I\otimes Y^*)\Pi_{mf^2} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -e & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & (I\otimes D) & 0 & (Q^*\otimes I) & 0 & (Y^*\otimes I) \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -e & 0 \end{bmatrix} \times$$

$$\times \begin{bmatrix} vec(\Delta A) \\ vec(\Delta Y) \\ vec(\Delta B) \\ vec(\Delta C) \\ \Delta\gamma \\ vec(\Delta D) \end{bmatrix} = \left[T_{t1}, T_{t2}, T_{t3}, T_{t4}, T_{t5}, T_{t6}\right]\Delta_{aybcd\gamma} = T_t \Delta_{aybcd\gamma}.$$

Further we obtain the expression

$$T\Delta q + T_{t1} vec(\Delta A) + T_{t2} vec(\Delta Y) + T_{t3} vec(\Delta B) + T_{t4} vec(\Delta C) + T_{t5}\Delta\gamma + \\ + T_{t6} vec(\Delta D) = vec(\Delta L_1). \tag{10}$$

Finally the relative perturbation bound for the solution $Q*$ of the LMI (4) has the form

$$\frac{\|\Delta q\|_2}{\|vec(Q^*)\|_2} \le \frac{1}{\|vec(Q^*)\|_2}\left(T_1\frac{\|vec(\Delta A)\|_2}{\|vec(A)\|_2} + T_2\frac{\|vec(\Delta Y)\|_2}{\|vec(Y^*)\|_2} + T_3\frac{\|vec(\Delta B)\|_2}{\|vec(B)\|_2}\right) \\ + \frac{1}{\|vec(Q^*)\|_2}\left(T_4\frac{\|vec(\Delta C)\|_2}{\|vec(C)\|_2} + T_5\frac{|\Delta\gamma|_2}{|\gamma_{opt}|_2} + T_6\frac{\|vec(\Delta D)\|_2}{\|vec(D)\|_2}\right) \\ + \frac{1}{\|vec(Q^*)\|_2}\left(L_1\frac{\|vec(\Delta L_1)\|_2}{\|vec(L^*)\|_2}\right) \tag{11}$$

where

$$\frac{T_1}{\|vec(Q^*)\|_2} = \frac{\|T^+\|_2\|T_{t1}\|_2\|vec(A)\|_2}{\|vec(Q^*)\|_2}, \frac{T_2}{\|vec(Q^*)\|_2} = \frac{\|T^+\|_2\|T_{t2}\|_2\|vec(Y^*)\|_2}{\|vec(Q^*)\|_2},$$

$$\frac{T_3}{\|vec(Q^*)\|_2} = \frac{\|T^+\|_2\|T_{t3}\|_2\|vec(B)\|_2}{\|vec(Q^*)\|_2}, \frac{T_4}{\|vec(Q^*)\|_2} = \frac{\|T^+\|_2\|T_{t4}\|_2\|vec(C)\|_2}{\|vec(Q^*)\|_2},$$

$$\frac{T_5}{\|vec(Q^*)\|_2} = \frac{\|T^+\|_2\|T_{t5}\|_2|\Delta\gamma|_2}{\|vec(Q^*)\|_2}, \frac{T_6}{\|vec(Q^*)\|_2} = \frac{\|T^+\|_2\|T_{t6}\|_2\|vec(D)\|_2}{\|vec(Q^*)\|_2},$$

$$\frac{L_1}{\|vec(Q^*)\|_2} = \frac{\|T^+\|_2\|vec(L^*)\|_2}{\|vec(Q^*)\|_2}.$$

may be considered as individual relative condition numbers of the LMI (4) with respect to the perturbations $\Delta A, \Delta B, \Delta C, \Delta D, \Delta Y$ and $\Delta\gamma$.

In a similar way the relative perturbation bounds for the solution $Y*$ of the LMI (4) may be obtained using the following expression

$$\Delta_Y + \Omega_Y = \Delta L_2, \tag{12}$$

where

$$\Delta_Y = \begin{bmatrix} 0 & \Delta Y^T B^T & 0 & 0 \\ B\Delta Y & 0 & 0 & \Delta Y^T D^T \\ 0 & 0 & 0 & 0 \\ 0 & D\Delta Y & 0 & 0 \end{bmatrix},$$

$$\Omega_Y = \begin{bmatrix} -\Delta Q & \Delta Q A^T + Q^*\Delta A^T + Y^{*T}\Delta B^T & 0 & 0 \\ A\Delta Q + \Delta A Q^* + \Delta B Y^* & 0 & 0 & \Delta Q C^T + Q^*\Delta C^T + Y^{*T}\Delta D^T \\ 0 & 0 & -\Delta\gamma I & 0 \\ 0 & C\Delta Q + \Delta C Q^* + \Delta D Y^* & 0 & -\Delta\gamma I \end{bmatrix},$$

Here the terms of second and higher order are neglected. The relation (12) may be written in a vector form as

$$vec(\Delta_Y) + vec(\Omega_Y) = vec(\Delta L_2), \tag{13}$$

where

$$vec(\Delta_Y) = [0, (B\otimes I)\Pi_{n\times m}, 0, 0, (I\otimes B), 0, 0, (D\otimes I)\Pi_{n\times m}, 0, 0, 0, 0, 0, (I\otimes D), 0, 0]^T \\ * vec(\Delta Y) = W\Delta y,$$

$$vec(\Omega_Y) = \begin{bmatrix} 0 & -I & 0 & 0 & 0 & 0 \\ (I\otimes Q^*)\Pi_{n^2} & (A\otimes I) & (I\otimes Y^{*T})\Pi_{nl^2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ (Q^*\otimes I) & (I\otimes A) & (Y^*\otimes I) & 0 & 0 & 0 \\ 0 & -I & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & (C\otimes I) & 0 & (I\otimes Q^*)\Pi_{n^2} & 0 & (I\otimes Y^*)\Pi_{nl^2} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -e & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & (I\otimes C) & 0 & (Q^*\otimes I) & 0 & (Y^*\otimes I) \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -e & 0 \end{bmatrix} \times$$

$$\times \begin{bmatrix} vec(\Delta A) \\ vec(\Delta Q) \\ vec(\Delta B) \\ vec(\Delta C) \\ \Delta\gamma \\ vec(\Delta D) \end{bmatrix} = \left[W_{t1}, W_{t2}, W_{t3}, W_{t4}, W_{t5}, W_{t6}\right]\Delta_{aqbcd\gamma} = W_t\Delta_{aqbcd\gamma}.$$

Further we obtain the expression

$$W\Delta y + W_{t1} vec(\Delta A) + W_{t2} vec(\Delta Q) + W_{t3} vec(\Delta B) + W_{t4} vec(\Delta C) + W_{t5}\Delta\gamma + \\ + W_{t6} vec(\Delta D) = vec(\Delta L_2). \tag{14}$$

Finally the relative perturbation bound for the solution $Y*$ of the LMI (4) has the form

$$\frac{\|\Delta y\|_2}{\|vec(Y^*)\|_2} \le \frac{1}{\|vec(Y^*)\|_2}\left(W_1\frac{\|vec(\Delta A)\|_2}{\|vec(A)\|_2}+W_2\frac{\|vec(\Delta Q)\|_2}{\|vec(Q^*)\|_2}+T_3\frac{\|vec(\Delta B)\|_2}{\|vec(B)\|_2}\right)$$
$$+\frac{1}{\|vec(Y^*)\|_2}\left(W_4\frac{\|vec(\Delta C)\|_2}{\|vec(C)\|_2}+W_5\frac{|\Delta\gamma|_2}{|\gamma_{opt}|_2}+W_6\frac{\|vec(\Delta D)\|_2}{\|vec(D)\|_2}\right)$$
$$+\frac{1}{\|vec(Y^*)\|_2}\left(L_2\frac{\|vec(\Delta L_2)\|_2}{\|vec(L^*)\|_2}\right) \quad (15)$$

where

$$\frac{W_1}{\|vec(Y^*)\|_2}=\frac{\|W^\dagger\|_2\|W_{t1}\|_2\|vec(A)\|_2}{\|vec(Y^*)\|_2},\ \frac{W_2}{\|vec(Y^*)\|_2}=\frac{\|W^\dagger\|_2\|W_{t2}\|_2\|vec(Q^*)\|_2}{\|vec(Y^*)\|_2},$$
$$\frac{W_3}{\|vec(Y^*)\|_2}=\frac{\|W^\dagger\|_2\|W_{t3}\|_2\|vec(B)\|_2}{\|vec(Y^*)\|_2},\ \frac{W_4}{\|vec(Y^*)\|_2}=\frac{\|W^\dagger\|_2\|W_{t4}\|_2\|vec(C)\|_2}{\|vec(Y^*)\|_2},$$
$$\frac{W_5}{\|vec(Y^*)\|_2}=\frac{\|W^\dagger\|_2\|W_{t5}\|_2|\Delta\gamma|_2}{\|vec(Y^*)\|_2},\ \frac{W_6}{\|vec(Y^*)\|_2}=\frac{\|W^\dagger\|_2\|W_{t6}\|_2\|vec(D)\|_2}{\|vec(Y^*)\|_2},$$
$$\frac{L_2}{\|vec(Y^*)\|_2}=\frac{\|W^\dagger\|_2\|vec(L^*)\|_2}{\|vec(Y^*)\|_2}.$$

may be considered as individual relative condition numbers of the LMI (4) with respect to the perturbations $\Delta A, \Delta B, \Delta C, \Delta D, \Delta Q$ and $\Delta\gamma$.

IV. Numerical Examples

Consider the discrete-time system (1), where

$$A=\begin{bmatrix}1 & 0.01\\ -0.0067 & 0.9966\end{bmatrix}, B=\begin{bmatrix}0 & 0 & 0 & 0\\ -0.004 & -0.0007 & -0.001 & 0.0033\end{bmatrix},$$
$$C=\begin{bmatrix}-0.6667 & -0.333\\ 0 & 1\\ 2 & 0\\ 1 & 0\end{bmatrix}, D=\begin{bmatrix}-0.4 & -0.0667 & -0.1 & 0.333\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{bmatrix}.$$

The perturbations in the system matrices of the discrete-time system are chosen as

$$\Delta A = A\times 10^{-i}, \Delta B = B\times 10^{-i}, \Delta C = C\times 10^{-i},$$
$$\Delta D = D\times 10^{-i}, \Delta L_1 = L^*\times 10^{-i}, \Delta L_2 = L^*\times 10^{-i}, \Delta\gamma = \gamma_{opt}\times 10^{-i}$$
$$\Delta Q^* = Q^*\times 10^{-i}, \Delta Y = Y^*\times 10^{-i} \text{ for } i = 8,7,..,4.$$

The perturbed solutions $Q^*+\Delta Q$ and $Y^*+\Delta Y$ are computed based on the method derived in [5] and using the software [7]. The relative perturbation bounds for the solutions Q^* and Y^* of the LMIs (4) are obtained by the linear bounds (11) and (15), respectively.

The results obtained for different values of i are shown in the following table

TABLE I

i	$\frac{\|\Delta q\|_2}{\|vec(Q^*)\|_2}$	Bound (11)	$\frac{\|\Delta y\|_2}{\|vec(Y^*)\|_2}$	Bound (15)
8	$3.59*10^{-8}$	$4.92*10^{-7}$	$1.54*10^{-8}$	$5.32*10^{-7}$
7	$3.59*10^{-7}$	$4.92*10^{-6}$	$1.54*10^{-7}$	$5.32*10^{-6}$
6	$3.59*10^{-6}$	$4.92*10^{-5}$	$1.54*10^{-6}$	$5.32*10^{-5}$
5	$3.59*10^{-5}$	$4.92*10^{-4}$	$1.54*10^{-5}$	$5.32*10^{-4}$
4	$3.59*10^{-4}$	$4.92*10^{-3}$	$1.54*10^{-4}$	$5.32*10^{-3}$

The obtained perturbation bounds (11) and (15), based on the presented solution approach, are close to the real relative perturbation bounds $\frac{\|\Delta q\|_2}{\|vec(Q^*)\|_2}$ and $\frac{\|\Delta y\|_2}{\|vec(Y^*)\|_2}$, thus they are good in sense that they are tight.

V. Conclusion

The linear full perturbation analysis of the discrete-time LMI based H_∞ control problem has been studied. Tight perturbation bounds, which are linear functions of the data perturbations, have been obtained for the matrix inequalities determining the problem solution. Based on these results we have presented numerical examples to explicitly reveal the performance and applicability of the proposed approach to analyze the sensitivity of the discrete-time LMI based H_∞ control problem.

References

[1] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*. SIAM, Philadelphia, 1994.

[2] A. Yonchev, M. Konstantinov, and P. Petkov. *Linear Matrix Inequalities in Control Theory*, Demetra, Sofia, 2005. ISBN 954:9526-32-1 (in Bulgarian).

[3] I. R. Peterson, B. D. O. Anderson, and E. A. Jonkheere, *A first principles solution to the non-singular H_∞ control problem*. Int. J. Robust Non. Cont. 1, pp. 171-185, 1994.

[4] J.C.Doyle, K.Glover, P.P.Khargonekar, and B.A.Francis, *State-Space Solutions to Standard H_2 and H_∞ Control Problems*. IEEE Transactions on Automatic Control 34, pp. 831-847, 1989.

[5] P. Gahinet and P. Apkarian. *A linear matrix inequality approach to H∞ control*. Int. J. Robust and Nonlinear Control. 4:421-448, 1994.

[6] Y. Nestorov and P. Gahinet, *Interior-Point Polynomial Algorithms in Convex Programming*, SIAM, Philadelphia, PA,1994.

[7] P. Gahinet, A. Nemirovski, A. Laub, and M. Chilali. *LMI Control Toolbox for Use with MATLAB*. The MathWorks, Inc., 2000

[8] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*. SIAM, Philadelphia, 1994.

[9] G. E. Dullerud, and F. Paganini. *A Course in Robust Control Theory*. Springer-Verlag, New York, 2000.

Sensitivity Analysis of the Continuous-time LMI Based Bounded Output Energy Problem

Andrey S. Yonchev[1], Mihail M. Konstantinov[2] and Petko H. Petkov[3]

***Abstract* – The bounded output energy problem is connected with obtaining a control law that ensures closed-loop stability, specified performance and minimizes the output energy. The specified control problem can be implicitly realized by the solutions Q, Y of a system of linear matrix inequalities (LMIs). The paper is concerned with obtaining linear perturbation bounds for the continuous-time LMI based bounded output energy problem, which are linear functions of the data perturbations. The sensitivity analysis of the perturbed matrix inequalities is considered in a similar manner as for perturbed matrix equations, after introducing a suitable right hand part, which is slightly perturbed. The proposed approach leads to tight linear perturbation bounds for the LMIs' solutions to the bounded output energy problem. Numerical example is also presented.**

***Keywords* – Sensitivity analysis, bounded output energy problem, LMI based synthesis, Linear systems.**

I. INTRODUCTION

In many control problems, the design constraints have a simple reformulation in terms of linear matrix inequalities (LMIs). This is hardly surprising, given that LMIs are direct byproducts of Lyapunov based criteria, and that Lyapunov techniques play a central role in the analysis and control of linear systems, see [1,2] and the literature therein. The bounded output energy problem is a good illustration of this point.

The effectiveness of LMI approach remains valuable for several reasons. To begin with it is applicable to all plants without restrictions on infinite or pure imaginary invariant zeros. In addition, the LMI based design is practical thanks to the availability of efficient convex optimization algorithms, based on the interior point method [3], and software [4].

In this paper we propose an approach to perform linear sensitivity analysis of the LMI based bounded output energy problem via introducing a suitable right hand part in the considered matrix inequalities.

We use the following notations: $R^{m\times n}$ - the space of real $m \times n$ matrices; $R^n = R^{n\times 1}$; I_n - the identity $n \times n$ matrix; e_n - the unit $n \times 1$ vector; M^T - the transpose of M; $M^\dagger$ - the pseudo inverse of M; $\| M \|_2 = \sigma_{\max}(M)$ - the spectral norm of M, where $\sigma_{\max}(M)$ is the maximum singular value of M; $vec(M) \in R^{mn}$ - the column-wise vector representation of $M \in R^{m\times n}$; $\Pi_{m,n} \in R^{mn\times mn}$ - the vec-permutation matrix, such that $vec(M^T) = \Pi_{m,n} vec(M)$; $M \otimes P$ - the Kroneker product of the matrices M and P. The notation ":=" stands for "equal by definition".

The remainder of the paper is organized as follows. In Section 2 we shortly present the problem set up and objective. Section 3 describes the performed linear full perturbation analysis of the LMI based bounded output energy problem. Section 4 presents a numerical example before we conclude in Section 5 with some final remarks.

II. PROBLEM SET UP AND OBJECTIVE

Consider the linear continuous-time system

$$\begin{aligned} \dot{x} &= Ax + Bu, \\ y &= Cx \end{aligned} \qquad (1)$$

where $x \in R^n$, $u \in R^m$, and $y \in R^r$ are the system state, input and output vectors respectively, and A, B, C are constant matrices of compatible size.

Bounded output energy problem means for a given initial state $x(0)$ to find a control law, which minimizes the output energy

$$\int_0^\infty y^T y dt.$$

It is also necessary to find a quadratic Lyapunov function $V(x) = x^T P x,\ P > 0$ such that

$$\frac{d}{dt} V(x) < -y^T y.$$

In order to solve the bounded energy control problem and to ensure closed-loop stability and specified performance it is necessary to design a state-feedback controller $u = Kx$.

We consider an LMI approach to solve the bounded output energy problem, as stated in [5]

$$x^T[(A+BK)^T P + P(A+BK)]x < -y^T y, P > 0. \qquad (2)$$

Unfortunately inequality (2) is not an LMI with respect to the decision variables K and P. That is why we perform the substitution $Q = P^{-1}, Q > 0$ and $Y = KP^{-1}$ to obtain the following system of LMIs:

[1]Andrey S. Yonchev is with the Faculty of Automatics, Technical University, 1000 Sofia, Bulgaria, E-mail: ajonchev@mail.bg

[2]Mihail M. Konstantinov is with the Faculty of Mathematics, UACG, 1046 Sofia, Bulgaria, E-mail: mmk_fte@uacg.bg

[3]Petko H. Petkov is with the Faculty of Automatics, Technical University, 1000 Sofia, Bulgaria, E-mail: php@tu-sofia.bg

$$(AQ+BY)^T+(AQ+BY)-(CQ)^T CQ>0 \qquad (3)$$

To transform LMI (3) we apply Schur complement argument [6] to obtain the following inequality:

$$\begin{bmatrix}(AQ+BY)^T+(AQ+BY) & QC^T\\ CQ & -I\end{bmatrix}, Q>0 \qquad (4)$$

with respect to the variables *Q* and Y.

The main objective of the paper is to perform a linear sensitivity analysis of the LMI system (4) needed to solve the bounded output energy problem.

Suppose that the matrices *A, B, C* are subject to perturbations $\Delta A, \Delta B, \Delta C$ and assume that they do not change the sign of the LMI system (4). The sensitivity analysis of the continuous-time LMI based bounded output energy problem is aimed at determining perturbation bounds of the LMIs (4) as functions of the perturbations in the data *A, B, C*.

III. Linear Full Perturbation Analysis

We perform sensitivity analysis of the LMI (4) for the continuous-time system (1).

$$\begin{bmatrix}ABQY^T+ABQY & qQC^T\\ qQC & -I\end{bmatrix}<0 \qquad (5)$$

where $ABQY^T=(Q+\Delta Q)(A+\Delta A)^T+(Y+\Delta Y)^T(B+\Delta B)^T$

$ABQY=(A+\Delta A)(Q+\Delta Q)+(B+\Delta B)(Y+\Delta Y)$,

$qQC^T=(Q+\Delta Q)(C+\Delta C)^T$,

$qCQ=(C+\Delta C)(Q+\Delta Q)$. We have to study the effect of the perturbations $\Delta A, \Delta B, \Delta C$ on the perturbed LMI solutions $Q^*+\Delta Q$ and $Y^*+\Delta Y$, where Q^*, Y^* and $\Delta Q, \Delta Y$ are the nominal solution of the inequality (4) and the perturbations, respectively. The essence of our approach is to perform sensitivity analysis of the inequality (4) in a similar manner as for a proper matrix equation after introducing a suitable right hand part, which is slightly perturbed. Thus for LMI (5) we have:

$$\begin{bmatrix}ABQY^{*T}+ABQY^* & qQC^{*T}\\ qQC^* & -I\end{bmatrix}=N^*+\Delta N_1<0 \qquad (6)$$

where $ABQY^{*T}=(Q^*+\Delta Q)(A+\Delta A)^T+(Y^*+\Delta Y)^T(B+\Delta B)^T$,

$ABQY^*=(A+\Delta A)(Q^*+\Delta Q)+(B+\Delta B)(Y^*+\Delta Y)$

$qQC^{*T}=(Q^*+\Delta Q)(C+\Delta C)^T$,

$qCQ^*=(C+\Delta C)(Q^*+\Delta Q)$ and N^* is obtained using the nominal LMI

$$\begin{bmatrix}AQ^*+Q^*A^T+BY^*+Y^{*T}B^T & Q^*C^T\\ CQ^* & -I\end{bmatrix}=N^*<0\ . \qquad (7)$$

The matrix ΔN_1 is due to the data and closed-loop performance perturbations, the rounding errors and the sensitivity of the interior point method that is used to solve the LMIs.

Using the relation (7) the perturbed equation (6) may be written as

$$\Delta_Q+\Omega_Q=\Delta N_1, \qquad (8)$$

where

$$\Delta_Q=\begin{bmatrix}A\Delta Q+\Delta QA^T & \Delta QC^T\\ C\Delta Q & 0\end{bmatrix},$$

$$\Omega_Q=\begin{bmatrix}\Delta AQ^*+Q^*\Delta A^T+B\Delta Y+\Delta Y^TB^T+\Delta BY^*+Y^{*T}\Delta B^T & Q^*\Delta C^T\\ \Delta CQ^* & 0\end{bmatrix}.$$

Here the terms of second and higher order are neglected. The relation (8) may be written in a vector form as

$$vec(\Delta_Q)+vec(\Omega_Q)=vec(\Delta N_1), \qquad (9)$$

where

$$vec(\Delta_Q)=\left[A\otimes I+I\otimes A, C\otimes I, I\otimes C, 0\right]^T vec(\Delta Q)$$
$$=T\Delta q,$$

$$vec(\Omega_Q)=$$
$$=\begin{bmatrix}(Q^*\otimes I)+(I\otimes Q^*)\Pi_{n^2} & (I\otimes B)+(B\otimes I)\Pi_{nm} & (Y^*\otimes I)+(I\otimes Y^{*T})\Pi_{m^2} & 0\\ 0 & 0 & 0 & (I\otimes Q^*)\Pi_{n^2}\\ 0 & 0 & 0 & (Q^*\otimes I)\\ 0 & 0 & 0 & 0\end{bmatrix}$$

$$\times\begin{bmatrix}vec(\Delta A)\\ vec(\Delta Y)\\ vec(\Delta B)\\ vec(\Delta C)\end{bmatrix}=\left[T_{t1}, T_{t2}, T_{t3}, T_{t4}\right]\Delta_{aybc}=T_t\Delta_{aybc}.$$

Further we obtain the expression

$$T\Delta q+T_{t1}vec(\Delta A)+T_{t2}vec(\Delta Y)+T_{t3}vec(\Delta B)+T_{t4}vec(\Delta C)=$$
$$=vec(\Delta N_1). \qquad (10)$$

Finally the relative perturbation bound for the solution Q^* of the LMI (4) has the form

$$\frac{\|\Delta q\|_2}{\|vec(Q^*)\|_2} \le \frac{1}{\|vec(Q^*)\|_2}\left(T_1\frac{\|vec(\Delta A)\|_2}{\|vec(A)\|_2}+T_2\frac{\|vec(\Delta Y)\|_2}{\|vec(Y^*)\|_2}+T_3\frac{\|vec(\Delta B)\|_2}{\|vec(B)\|_2}\right)$$
$$+\frac{1}{\|vec(Q^*)\|_2}\left(T_4\frac{\|vec(\Delta C)\|_2}{\|vec(C)\|_2}+N_1\frac{\|vec(\Delta N_1)\|_2}{\|vec(N^*)\|_2}\right) \quad (11)$$

where

$$\frac{T_1}{\|vec(Q^*)\|_2}=\frac{\|T^\dagger\|_2\|T_{t1}\|_2\|vec(A)\|_2}{\|vec(Q^*)\|_2}, \frac{T_2}{\|vec(Q^*)\|_2}=\frac{\|T^\dagger\|_2\|T_{t2}\|_2\|vec(Y^*)\|_2}{\|vec(Q^*)\|_2},$$
$$\frac{T_3}{\|vec(Q^*)\|_2}=\frac{\|T^\dagger\|_2\|T_{t3}\|_2\|vec(B)\|_2}{\|vec(Q^*)\|_2}, \frac{T_4}{\|vec(Q^*)\|_2}=\frac{\|T^\dagger\|_2\|T_{t4}\|_2\|vec(C)\|_2}{\|vec(Q^*)\|_2},$$
$$\frac{N_1}{\|vec(Q^*)\|_2}=\frac{\|T^\dagger\|_2\|vec(N^*)\|_2}{\|vec(Q^*)\|_2}.$$

may be considered as individual relative condition numbers of the LMI (4) with respect to the perturbations $\Delta A, \Delta B, \Delta C$ and ΔY.

In a similar way the relative perturbation bounds for the solution Y^* of the LMI (4) may be obtained using the following expression

$$\Delta_Y + \Omega_Y = \Delta N_2, \quad (12)$$

where

$$\Delta_Y = \begin{bmatrix} B\Delta Y + \Delta Y^T B^T & 0 \\ 0 & 0 \end{bmatrix},$$
$$\Omega_Y=\begin{bmatrix} A\Delta Q+\Delta Q A^T+\Delta A Q^*+Q^*\Delta A^T+\Delta B Y^*+Y^{*T}\Delta B^T & \Delta Q C^T+Q^*\Delta C^T \\ C\Delta Q+\Delta C Q^* & 0 \end{bmatrix}.$$

Here the terms of second and higher order are neglected. The relation (12) may be written in a vector form as

$$vec(\Delta_Y)+vec(\Omega_Y)=vec(\Delta N_2), \quad (13)$$

where

$$vec(\Delta_Y)=[(I\otimes B)+(B\otimes I)\Pi_{n\times m},0,0,0]^T vec(\Delta Y)$$
$$=W\Delta y,$$

$$vec(\Omega_Y)=$$
$$=\begin{bmatrix} (Q^*\otimes I)+(I\otimes Q^*)\Pi_{n^2} & (I\otimes A)+(A\otimes I) & (Y^*\otimes I)+(I\otimes Y^{*T})\Pi_{m^2} & 0 \\ 0 & (C\otimes I) & 0 & (Q^*\otimes I) \\ 0 & (I\otimes C) & 0 & (I\otimes Q^*)\Pi_{n^2} \\ 0 & 0 & 0 & 0 \end{bmatrix}$$
$$\times\begin{bmatrix} vec(\Delta A) \\ vec(\Delta Q) \\ vec(\Delta B) \\ vec(\Delta C) \end{bmatrix}=[W_{t1},W_{t2},W_{t3},W_{t4}]\Delta_{ytc}=W_t\Delta_{ytc}.$$

Further we obtain the expression

$$W\Delta y+W_{t1}vec(\Delta A)+W_{t2}vec(\Delta Q)+W_{t3}vec(\Delta B)+W_{t4}vec(\Delta C)=$$
$$=vec(\Delta N_2). \quad (14)$$

Finally the relative perturbation bound for the solution Y^* of the LMI (4) has the form

$$\frac{\|\Delta y\|_2}{\|vec(Y^*)\|_2} \le \frac{1}{\|vec(Y^*)\|_2}\left(W_1\frac{\|vec(\Delta A)\|_2}{\|vec(A)\|_2}+W_2\frac{\|vec(\Delta Q)\|_2}{\|vec(Q^*)\|_2}+W_3\frac{\|vec(\Delta B)\|_2}{\|vec(B)\|_2}\right)$$
$$+\frac{1}{\|vec(Y^*)\|_2}\left(W_4\frac{\|vec(\Delta C)\|_2}{\|vec(C)\|_2}+N_2\frac{\|vec(\Delta N_2)\|_2}{\|vec(N^*)\|_2}\right) \quad (15)$$

where

$$\frac{W_1}{\|vec(Y^*)\|_2}=\frac{\|W^\dagger\|_2\|W_{t1}\|_2\|vec(A)\|_2}{\|vec(Y^*)\|_2}, \frac{W_2}{\|vec(Y^*)\|_2}=\frac{\|W^\dagger\|_2\|W_{t2}\|_2\|vec(Q^*)\|_2}{\|vec(Y^*)\|_2},$$
$$\frac{W_3}{\|vec(Y^*)\|_2}=\frac{\|W^\dagger\|_2\|W_{t3}\|_2\|vec(B)\|_2}{\|vec(Y^*)\|_2}, \frac{W_4}{\|vec(Y^*)\|_2}=\frac{\|W^\dagger\|_2\|W_{t4}\|_2\|vec(C)\|_2}{\|vec(Y^*)\|_2},$$
$$\frac{N_2}{\|vec(Y^*)\|_2}=\frac{\|W^\dagger\|_2\|vec(N^*)\|_2}{\|vec(Y^*)\|_2}.$$

may be considered as individual relative condition numbers of the LMI (4) with respect to the perturbations $\Delta A, \Delta B, \Delta C$ and ΔQ.

IV. Numerical Examples

Consider the continuous-time system (1), where

$$A=\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & \frac{mm^2*el^2*g}{((M+mm)*J+M*mm*el^2)} & \frac{-(J+mm*el^2)*fc}{((M+mm)*J+M*mm*el\wedge 2)} & \frac{-mm*el*fp}{((M+mm)*J+M*mm*el^2)} \\ 0 & \frac{((M+mm)*mm*el*g)}{((M+mm)*J+M*mm*el^2)} & \frac{-mm*el*fc}{((M+mm)*J+M*mm*el^2)} & \frac{-(M+mm)*fp}{((M+mm)*J+M*mm*el^2)} \end{bmatrix},$$
$$B=\begin{bmatrix} 0 \\ 0 \\ \frac{(J+mm*el\wedge 2)}{((M+mm)*J+M*mm*el^2)} \\ \frac{-mm*el}{((M+mm)*J+M*mm*el^2)} \end{bmatrix}, C=\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, D=\begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

here

M=0.768; mm=0.104; el=0.1735; J=2.829e-3;
g=9.81; fc=0.5; fp=0.1035e-3.

The perturbations in the system matrices of the continuous-time system are chosen as

$$\Delta A = A\times 10^{-i}, \Delta B = B\times 10^{-i}, \Delta C = C\times 10^{-i},$$
$$\Delta N_1 = N^*\times 10^{-i}, \Delta N_2 = N^*\times 10^{-i},$$
$$\Delta Q^* = Q^*\times 10^{-i}, \Delta Y = Y^*\times 10^{-i} \text{ for } i = 8,7,..,4.$$

The perturbed solutions $Q^* + \Delta Q$ and $Y^* + \Delta Y$ are computed based on the method derived in [7] and using the software [4]. The relative perturbation bounds for the solutions Q^* and Y^* of the LMIs (4) are obtained by the linear bounds (11) and (15), respectively.

The results obtained for different values of i are shown in the following table

TABLE I

i	$\frac{\|\Delta q\|_2}{\|vec(Q^*)\|_2}$	Bound (11)	$\frac{\|\Delta y\|_2}{\|vec(Y^*)\|_2}$	Bound (15)
8	$6.34*10^{-8}$	$9.98*10^{-7}$	$3.23*10^{-8}$	$1.53*10^{-7}$
7	$6.34*10^{-7}$	$9.98*10^{-6}$	$3.23*10^{-7}$	$1.53*10^{-6}$
6	$6.34*10^{-6}$	$9.98*10^{-5}$	$3.23*10^{-6}$	$1.53*10^{-5}$
5	$6.34*10^{-5}$	$9.98*10^{-4}$	$3.23*10^{-5}$	$1.53*10^{-4}$
4	$6.34*10^{-4}$	$9.98*10^{-3}$	$3.23*10^{-4}$	$1.53*10^{-3}$

The obtained perturbation bounds (11) and (15), based on the presented solution approach, are close to the real relative perturbation bounds $\frac{\|\Delta q\|_2}{\|vec(Q^*)\|_2}$ and $\frac{\|\Delta y\|_2}{\|vec(Y^*)\|_2}$, thus they are good in sense that they are tight.

V. CONCLUSION

The linear sensitivity analysis of the continuous-time LMI based bounded output energy problem has been studied. Tight perturbation bounds, which are linear functions of the data perturbations, have been obtained for the matrix inequalities determining the problem solution. Based on these results we have presented numerical examples to explicitly reveal the performance and applicability of the proposed approach to analyze the sensitivity of the continuous-time LMI based bounded output energy problem.

REFERENCES

[1] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*. SIAM, Philadelphia, 1994.

[2] A. Yonchev, M. Konstantinov, and P. Petkov. *Linear Matrix Inequalities in Control Theory*, Demetra, Sofia, 2005. ISBN 954:9526-32-1 (in Bulgarian).

[3] Y. Nestorov and P. Gahinet, *Interior-Point Polynomial Algorithms in Convex Programming*, SIAM, Philadelphia, PA,1994.

[4] P. Gahinet, A. Nemirovski, A. Laub, and M. Chilali. *LMI Control Toolbox for Use with MATLAB*. The MathWorks, Inc., 2000

[5] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*. SIAM, Philadelphia, 1994.

[6] G. E. Dullerud, and F. Paganini. *A Course in Robust Control Theory*. Springer-Verlag, New York, 2000.

[7] P. Gahinet and P. Apkarian. *A linear matrix inequality approach to H∞ control*. Int. J. Robust and Nonlinear Control. 4:421-448, 1994.

Managing the Process of Restructuring the National Railway Systems Using Fuzzy Logic

Gordan Stojić[1], Ilija Tanackov[1], Milan Marković[2], Jovan Tepić[1], Siniša Sremac[1]

***Abstract* – Carrying out the restructuring of the national railway systems inside and outside of EU is conducted by different level of dynamics which causes different levels of national railway reforms. The different level of restructuring has influence on the railway market establishment and its liberalisation. The EU institutions' tendency is to synchronise the process of restructuring of the national railway systems. The initial point of synchronisation is the estimation of the level of restructuring of the national railway systems. This paper presents a model of estimation of the level of restructuring of the national railway system using fuzzy logic.**

***Keywords* – estimate, restructuring, railway system, fuzzy logic**

I. INTRODUCTION

The reform level of a particular subject is often the matter of discussion and mutual misunderstandings. The evaluation is given according to the experience, intuition and subjective attitudes of particular institutions and experts. The institutions are: the Government, Ministries, Agencies, Companies, the EU organs, Financial Institutions (international: World Bank-WB, European Bank for Reconstruction and Development-EBRD, International Monetary Fund-IMF , European Investment Bank-EIB, or domestic banks) and experts that work on their own behalf and represent a particular organisation (market surveying, regulative compatibilities and the EU reforms and etc.). However, uncertainty, regarding the input data necessary for the certain decision making, is also present. This implies that all the parameters of evaluation are characterised by uncertainty, subjectivity, inaccuracy and ambiguity. There is a similar problem with the railway reform evaluation. The theory of fuzzy sets (scattered – inarticulate sets), is a very suitable tool for the treatment of uncertainty, subjectivity, ambiguity and inaccuracy [14].

The establishment of the railway reform levels was conducted several times in the past during the process of the national railway system reconstruction. The railway reforms represent a continuous process. The certain railway systems are at the beginning of the reforms, some of them have advanced and some have applied a large number of measures in the process of reform implementation.

The evaluation was mostly necessary for the approval of financial funds for the reforms and railway recovery but also to determine the level of the railway harmonisation with the European ones. The manner, in which the railway reform evaluation was conducted, is not always scientifically founded, except in the few cases which are to be analysed in this paper. The reformation level evaluation was mostly conducted by a heuristic evaluation method and it was based on the individual opinions, reasoning and intuitions and without any mathematical method application. The similar report was the one on major economic aspects of the EU railway reforms [10], mutual influence within the railway systems in the research of the South East Europe Transport Observatory organisation on the reforms of railways [12]. The significant evaluation of railway reforms in Europe with the reference to the infrastructural approach was given in the European Conference of Ministers of Transport report (ECMT) [5].

The WB described the railway system reforms as low, medium and high [1]. The IBM Business Consulting Services conducted the research for the evaluation of market liberalisation level in 27 European countries [7]. The Benchmarking method for the evaluation of legal and practical obstacles of the approach to the railway market was included in the research. Rank order of the countries regarding the railway transport market liberalisation was conducted by LIB Index (Fig. 1). The LIB index presents information on the relative degree of market opening in the European rail transport markets.

Fig. 1. Railway reform evaluation using LIB index

One of the latest approaches to the railway reform evaluation, applying the SWOT analysis, is presented in this paper [13].

In the literature can be found applicated hybrid models such as A'WOT analysis (combining the SWOT analysis and the Analytic Hierarchy Process - AHP). A'WOT analysis is applied for strategic planning of rural tourism [8].

According to this analysis, it can be determined which countries have the highest level of market liberalisation and the most extensive application of the EU Directives (Fig. 2).

[1]University of Novi Sad, Faculty of Technical Sciences, Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia e-mail: gordan@uns.ac.rs, ilijat@uns.ac.rs,

[2]University of Belgrade, Faculty of Transport and Traffic Engineering, Vojvode Stepe 305, 11000 Belgrade, Serbia

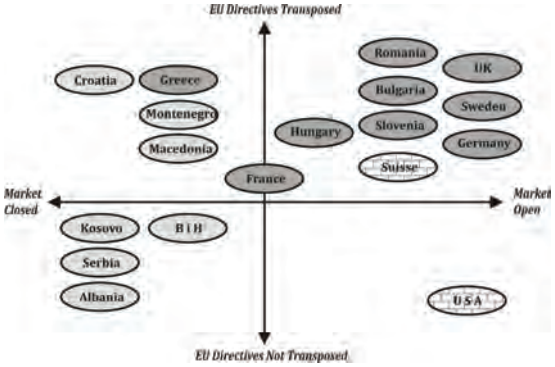


Fig. 2. Railway reform evaluation using SWOT analysis

The SWOT analysis is a widely applied method of analysis. The application of this analysis for the evaluation of railway system reformation also implies the use of intuition and subjective evaluation. However, the SWOT and A'WOT analysis are not only used as a direct research method. The results on SWOT analysis are too often only a superficial and imprecise listing or an incomplete qualitative examination of internal and external factors [3].

Apart from the previously mentioned methods, the reference also comprises the application of the Computable General Equilibrium (CGE) and Dynamic General Equilibrium (DGE) method for the evaluation of regulatory measure influences on liberalisation and market balance determination [2] [6] [8] [11]. The Benchmarking analysis for the railway system establishment was used in the paper [2].

The statistic analysis does not take uncertainty into consideration. Particular elements in the analysis are usually insufficiently precise, and their values' estimation is subjective. CGE and DGE methods employ a statistical data base.

The comparative Benchmarking method is based on the concept of an innovative way of positive practice and the use of experience. The initial problem with the application of the method is in the procedure of reducing it to the simple method of comparison or innovation via copying. On the other hand, railway reforms have not reached their final phase in any country since they represent a sustained process. This led to the following questions: Can the Benchmarking method, SWOT and A'WOT analysis be used for the evaluation of reform levels if none of the railway system is completely reformed? Which reform will be used to conduct a comparison in order to evaluate the reform level? We again come to the subjective evaluations, especially in the field of weighed factor determination.

II. FUZZY MODEL

A is defined as a fuzzy output variable which represents the evaluation of a railway reform level, where uncertainty, in terms of the performance of dynamics, the number of taken reform steps, the defined transport policy that is carried out, social and political reasons of the authority's readiness to conduct the reforms, political changes etc, is present. Assuming that the reforms can be: "Low", "Medium" and "High", and that the quantity evaluation is from 0 to 10, the membership functions and are defined (Fig. 3).

The following input variables have been defined in this paper for modelling and evaluation of the railway reforms: (1) Preparation for the railway reform; (2) Railway reforms criteria fulfilment; (3) The number of Railway Undertakings (RUs).

The European Union issued the Directive 91/440/EEC in order to improve the quality of railway services and to ensure more efficient railway's participation in the transport market in the same manner [4]. The basic aim of the Directive is to simplify the European national railway adjustment in order to meet the needs of the unique market and to improve their efficiency by ensuring self-management of railway companies and improving the financial structure of the companies. 18 years have passed since the Directive was issued in 1991 and this has been a slow process so far. This indicates that the problem is serious and that there is no capacity to solve it. Furthermore, it indicates the influence of public railway companies and political reasons to keep them under control.

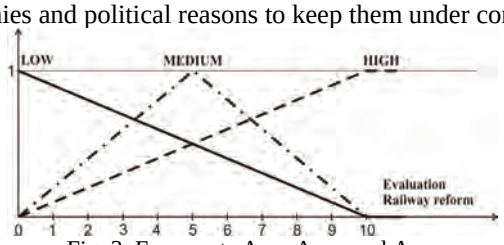


Fig. 3. Fuzzy sets A_{low}, A_{med}, and A_{high}

The restructuring of public railway companies is a process developed in three phases: (1) enforcing the law on railway (the EU Basic Regulative-Directive 91/440/EEC); (2) changing the public railway company into holding; (3) separating the infrastructural and transportation businesses into completely separate companies. Apart from the basic regulative, there is also an additional EU railway regulative in terms of the first, second and third directive sets.

Practice and review on the experience of the European countries reveal that there are two approaches in the restructuring process and they are: gradual and radical. A gradual approach implies the longer transitional period (railway law and holding). The radical approach to the restructure process means "rapid" and "sudden" transformation of the railway company regarding its organisation and its relation with the country without any previous transitional period. This means that at least two independent companies are formed: infrastructure and transport.

Different approaches to the phases of public railway company restructures in terms of duration, moment and manner of transition into the next phase imply the favourability of the theory on fuzzy sets, when it is necessary to take the following into consideration: uncertainty, subjectivity and inaccuracy of the restructure process.

The fuzzy variable B, showing the phase of the countries' readiness for the railway reforms, is defined. Assume the existence of "Unprepared" and "Prepared" countries for the railway reforms. The membership functions are defined (Fig. 4a).

The criteria number and types represent an important element according to which the level of the railway system reforms is evaluated. The following criteria for the evaluation of the railway system reforms are defined:

K_1 – "New regulative". If the first set is enforced than K_1 criterion is fulfilled. The basic regulative enforcement in the field of railway is included in the fuzzy variable "Prepared for the railway reforms".

K_2 – "Improved management structure". The criterion K_2 is considered fulfilled when the operators are converted into stock associations or they are private companies.

K_3 – "Open access to the infrastructure". The criterion K_3 is considered fulfilled if there is a greater number of transporters in the railway system.

K_4 – "Market liberalisation". The criterion K_4 is considered fulfilled if there is at least one international operator on the railway market.

K_5 – "Commercial business of the companies on the market". The criterion K_5 is considered fulfilled when, except for the established regulation, the state provides financial means for infrastructure controllers only for the development of railway infrastructure.

K_6 – "Subventions in the public transport (PSO)". The criterion K_6 is considered fulfilled when the authorised organs provide financial means for agreed duties that are related to the transport of general interest.

K_7 – "Adjustment of the employees' number and structure". The criterion K_7 is considered fulfilled when the rationalisation of employees was made.

In literature, defining a criterion, as well as the evaluation of levels of reforms, is based on experience, intuition and subjective evaluation [1] [7] [13]. Variable C fuzzy is a variable that evaluates the fulfillment of the criterion number that is needed. Let us predict that there is: "Unsatisfactory", "Small", "Satisfactory" and "Great" fulfillment of a criterion number (implementation of criteria).

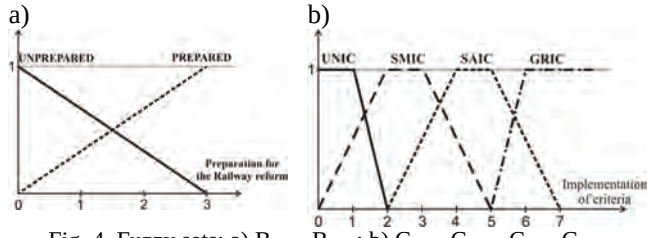


Fig. 4. Fuzzy sets: a) B_{unpr}, B_{prep}; b) C_{unic}, C_{smic}, C_{saic}, C_{gric}

The membership functions to the set C for unsatisfactory – UNIC, small – SMIC, satisfactory – SAIC and great impletion of criteria are defined (Fig. 4b).

Let us predict that there can be "Small", "Medium" and "Large" number of RUs. The fuzzy variable D that evaluates the number of RUs in a railway system is defined. The membership functions of the set D for small, medium and large number of RUs are defined (Fig. 5).

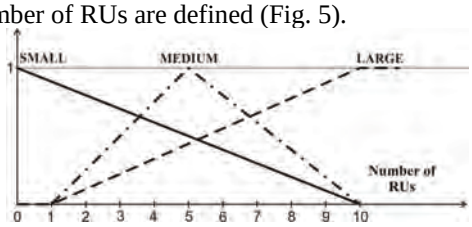


Fig. 5. Fuzzy sets D_{small}, D_{med} and D_{large}

Fuzzy logic is the base of the fuzzy system. It enables making decisions based on uncompleted information. The models based on fuzzy logic consist of the so called "IF-THEN" rules. Fuzzy logic for validation of levels of railway reforms consists of the following, equally competitive rules:

1. If (Preparation is UNPREPARED) and (Criteria is UNIC or SMIC) and (RUs is SMALL) then (Railway Reform is LOW) – weight: (1)
2. If (Preparation is UNPREPARED) and (Criteria is UNIC or SMIC) and (RUs is MEDIUM) then (Railway Reform is LOW) – weight: (0.5)
3. If (Preparation is PREPARED) and (Criteria is UNIC or SMIC) and (RUs is SMALL) then (Railway Reform is LOW) – weight: (1)
4. If (Preparation is PREPARED) and (Criteria is UNIC or SMIC) and (RUs is MEDIUM) then (Railway Reform is LOW) – weight: (0.5)
5. If (Preparation is PREPARED) and (Criteria is SAIC) and (RUs is ANY) then (Railway Reform is MEDIUM) – weight: (1)
6. If (Preparation is PREPARED) and (Criteria is GRIC) and (RUs is ANY) then (Railway Reform is HIGH) – weight: (1)

III. The Tests Results

The incoming variables in fuzzy systems represent the so called linguistic variables. The outcome is given in a continual phase. An adequate level of belonging is determined for all possible outcome sums of variables. After being observed, the levels of belonging of particular outcome sums of variables are to be made by defuzzification.

Defuzzification is the process of producing a quantifiable result in fuzzy logic, given fuzzy sets and corresponding membership degrees. It is typically needed in fuzzy control systems. These will have a number of rules that transform a number of variables into a fuzzy result, that is, the result is described in terms of membership in fuzzy sets. A useful defuzzification technique must first combine the results from the rules. The most typical fuzzy set membership function has the graph of a triangle. If this triangle were to be cut in a straight horizontal line somewhere between the top and the bottom, and the top portion were to be removed, the remaining portion forms a trapezoid. Typically, the first step of defuzzification is chopping off parts of the triangle to form trapezoids (or other shapes if the initial shapes were not triangles). In the most common technique, the trapezoids from all input functions are then superimposed one upon the other, forming a single geometric shape. Then, the centroid of this shape, called the fuzzy centroid, is calculated. The x coordinate of the centroid is the defuzzified value.

The authors decided to use the center of area (COA) method defuzzification and Mamdani fuzzy inference systems.

Probably the best known defuzzification operator is the center of gravity defuzzification (COG) method. It is basic general deffuzification method that computes the center of gravity of the area under the membership function. The value

x^* of the output, which is resulting from the COA method, is given in the following equation:

$$x^* = \sum_{i=x_{\min}}^{x_{\max}} x_i \cdot \mu(x_i) \Big/ \sum_{i=x_{\min}}^{x_{\max}} \mu(x_i) \qquad (1)$$

Where the $\mu(x_i)$ is membership function. The formula shows that COG calculates the expected value when A is considered to be probability distribution.

Testing the size assessment of the countries was carried out on randomly selected samples of the EU member countries and aspirants for EU membership

The results of the assessment of the size of the countries in randomly selected sample are shown in the following figure.

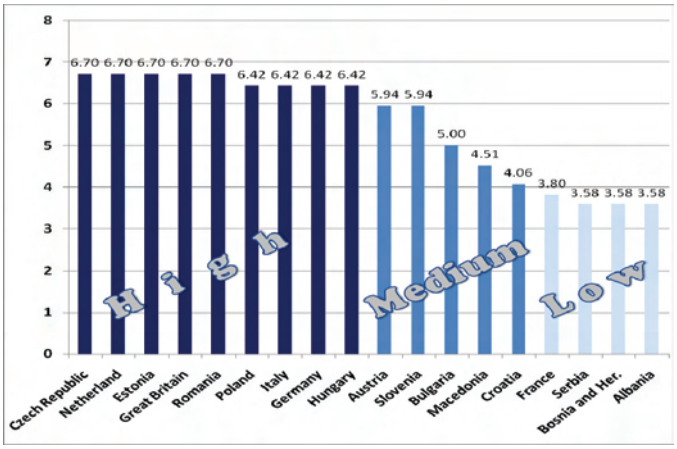


Fig. 6. Railway reforms grade of chosen countries

The results may be an indicator for the competent institutions, especially in countries with "low" reforms, to make appropriate criteria in order to acceleration of reforms. This would lead to improvement of the functioning of European railway system. Development of railway transportation may be difficult if railway system reforms are not on the same level, especially in neighbouring countries (France - German and Italy, Serbia-Macedonia etc.). If you do not establish a balance of railway reform Europe striving for liberalization of the railway market, introduction of competition, modernizing and accelerating the transport, the establishment of a single European railway system and similar, can be a problem.

IV. Conclusion

The railway reform evaluation is a very important process by which a reached level of a reform can be measured. Railway reforms becoming as similar as possible, is of a great importance for the stabilisation of the transport market. The European railway system must not be divided into non-synchronised railway national levels of reforms, because that would not contribute to making a unique European transport market.

Railway reform levels were often evaluated based on experts' beliefs or by using statistical or inadequate methods. Relevant parameters that were used in methods consisted of uncertainty, subjectivity, ambiguity and more than one meaning.

This paper shows a new way of modelling and evaluating the process of restructuring of the national railway systems by using the fuzzy logic. Railway reforms of randomly chosen countries have been evaluated by testing a model. By appropriate modification of rules and variable sums, the railway reform evaluation model can also be used in other fields of industry such as, the evaluation of the market liberalisation in general and its particular elements, then, in law system reforms evaluation, approaching EU countries, harmonisation of EU member countries, in various reform fields.

References

[1] Amos, P.: Reform, Commercialization and Private Sector Participation in Railways in Eastern Europe and Central Asia, TP-4, The World Bank Group, Washington, 2005.

[2] Balistreri, E. J., et al.: Modeling services liberalization: The case of Kenya, Economic Modelling 26, 2009.

[3] Chang, H. H., Huang, W.C.: Application of a quantification SWOT analytical method, Mathematical and Computer Modelling 43, 2006.

[4] Council of the European Union, Council Directive 91/440/EEC on the development Community's railways". Brussels: 29 July 1991.

[5] ECMT: Railway Reform & Charges the Use of Infrastructure, Paris, 2005.

[6] Feraboli, O.: A Dynamic General Equilibrium Analysis of Jordan's Trade Liberalisation, PhD dissertation, Chemnitz University of Technology, Faculty of Economics and Business Administration, Germany, 2006.

[7] IBM Business Consulting Services: Summary of the Study Rail Liberalisation Index 2007, Market Opening: Rail Markets of the Member States of the European Union, Switzerland and Norway in comparison, Brussels, October 2007.

[8] Kajanus, M, et al.: The use of value focused thinking and the A'WOT hybrid method in tourism management, Tourism Management 25, 499–506, 2004.

[9] Konan, D. E., Maskus K. E.: Quantifying the impact of services liberalization in a developing country, Journal of Development Economics 81, 2006.

[10] Pietrantonio, L. D., et al.: The Economics of EU Railway Reform, Bruges European Economic Policy Briefings – BEEP briefing n°8, College of Europe, Brugge, 2004.

[11] Ruiz Estrada, M.A.: Trade Liberalization Evaluation Methodology Theoretical Framework, SciTopics pages on Economics, Econometrics and Finance, 2009

[12] SEETO: Rail and Inter-modal Working Group 5th Session, Belgrade, 2007.

[13] SEETO: Exchange Of Information On Regional Legal Framework For Access To Railway Network And Draft Regulatory Manual, First Railway Reform Workshop For Task: Access To Railway Network, Belgrade, Serbia, February 2009.

[14] Teodorović, D., Kikuchi, S.: Fuzzy Sets in Traffic and Transport Systems, Preface, Fuzzy Sets and Systems 116, 2000.

Multiple-Model Model Predictive Control for Nonlinear Systems

Goran S. Stojanovski[1] and Mile J. Stankovski[2]

***Abstract* – In this paper we present a multi-parametric algorithm for control of nonlinear systems. Recently proposed algorithms for hybrid control of nonlinear systems introduce big computational burden. In order to reduce the optimization problem we use the multiple-model approach in. The algorithm is based on linearization of the nonlinear plant in multiple operational points. These operational points are chosen after detailed analysis of the plant's behaviour and the set of referent inputs. Then we construct model based predictive controller for each defined linearized model of the system, and we connect these states into one switched MPC. This controller automatically switches the prediction model according to the user instructions.**

***Keywords* – Model predictive control, nonlinear control, switching control.**

I. INTRODUCTION

Process industries need an easy to setup predictive controller that costs low, and maintains an adaptive behavior which accounts for time-varying dynamics as well as potential plant miss-modeling. MPC has the ability to fulfill the expectation of the engineers and to successfully control complex processes.

As presented in [1] the essence of model-based predictive control (MBPC) or model predictive control (MPC) lies in optimization of the future process behavior with respect to the future values of the executive (or manipulated) process variables. Throughout this report the abbreviation MPC shall be used. The use of linear, non-linear, hybrid and time-delay models in model-based predictive control is motivated by the drive to improve the quality of the prediction of inputs and outputs, as well as to reduce the computer burden during the optimization [1-5].

This paper is organized as follows: we briefly present the latest developments in MPC theory, and especially nonlinear model predictive control in section 2. Then in section 3 we introduce the techniques for multiple-model MPC. In section 4 we present the simulation result performed on a model of a tank reservoir. At the end in section 5 we point out the conclusions from the paper and give some further research headings in the area of nonlinear and multiple-model model predictive control.

[1]Goran S. Stojanovski is with the Faculty of Electrical Engineering and Information Technology, Ruger Boskovic bb, 1000 Skopje, Macedonia, E-mail: goranst@feit.ukim.edu.mk

[2]Mile J. Stankovski is with the Faculty of Electrical Engineering and Information Technology, Ruger Boskovic bb, 1000 Skopje, Macedonia, E-mail: milestk@feit.ukim.edu.mk

II. THE BASICS OF MODEL PREDICTIVE CONTROL

A. The General Idea of MPC

The general idea behind MPC is simple indeed. If we have a reliable model of the system, represented as in (1) or similar, we can use it for predicting the future system behavior. At each consecutive time of sampling k the controls inputs (2) are calculated,

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k);\ x(0) = x_0 \\ y(k+1) &= Cx(k) + Du(k); \end{aligned} \tag{1}$$

$$\begin{aligned} u(k) = [&u(k \mid k),\ u(k+1 \mid k),... \\ &u(k+N_u-2 \mid k),\ u(k+N_u-1 \mid k)] \end{aligned} \tag{2}$$

where A,B,C,D are the system matrices; N_u represents the length of the control horizon and the notation $u(k+p \mid k)$ means the prediction of the control input value for the future time $k+p$ calculated at time k. These control inputs are calculated in such a way as to minimize the difference between the predicted controlled outputs $y(k+p \mid k)$ and foreseen set points $r(k+p \mid k)$ for these outputs, over the prediction horizon N_y, $(p = 1,2,...N_y)$. Then only the first element from the calculated control inputs is applied to the process, i.e. $u(k) = u(k \mid k)$. At the next sample time $(k+1)$, we have a new measurement of the process outputs and the whole procedure is repeated.

The most common used cost function is the Quadratic, and it can be formulated as:

$$\begin{aligned} J(k) = &\sum_{p=N_1}^{N_y} \left\| w(k+p \mid k) - y(k+p \mid k) \right\|^2 + \\ &\lambda \sum_{p=0}^{N_u-1} \left\| \Delta u(k+p \mid k) \right\|^2 \end{aligned} \tag{3}$$

B. Nonlinear MPC

When we control nonlinear systems, usually we consider the following definition

$$\dot{x}(t) = f(x(t), u(t)), \quad x(0) = x_0 \in \mathrm{X}_0 \tag{4}$$

subject to the input and state constraints

$$u(t) \in U, \quad x(t) \in X, \quad \forall t \geq 0, \tag{5}$$

where $x(t) \in X \subseteq \Re^n$ is the system state and $u(t) \in U \subset \Re^m$ is the input applied to the system. In this case we need to use more complex nonlinear programming in order to compute the optimal solution of the nonlinear control problem.

III. Multiple-Model Model Predictive Control for Nonlinear Systems

For control the process is approximated with *p* linear affine models that built a hybrid PWA state space model as presented in [6]

$$\begin{aligned} x(k+1) &= A^i x(k) + B^i u(k) + f^i \\ y(k) &= C^i x(k) + D^i u(k) + g^i \\ \text{if } &\begin{bmatrix} x(k) \\ u(k) \end{bmatrix} \in P^i, i \in \{1,...., p\} \end{aligned} \tag{6}$$

where k is the discrete time index, A^i, B^i, C^i, D^i state space matrices, f^i, g^i the affine vectors, $u \in \Re^m$ input, $x \in \Re^n$ state, and P^i valid region of the state+input space in $\Re^{m+n}$. The system is subject to input and state constraints. For each region P^i a model exists and for it the corresponding mp-MPC controller is designed. The currently active model is determined by Model selection algorithm from estimated state values. Each time step the active controller computes the control signal. The control scheme is presented in Fig. 1.

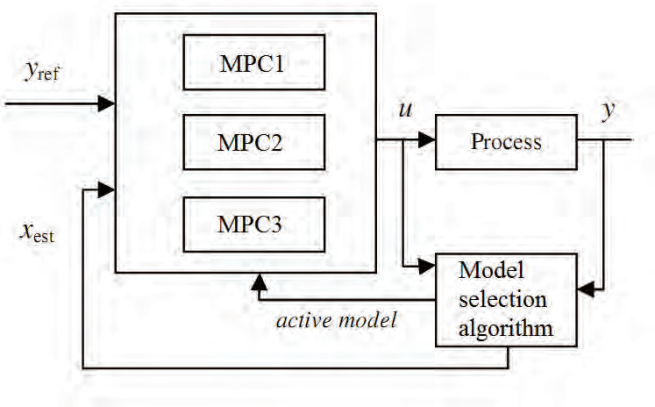


Fig. 1. Multiple-model predictive control scheme

This model predictive controllers use linear models of the nonlinear system to predict the future behavior. The models are linearized around different working points of the plant.

Model selection algorithm is the most important part of the multiple-model MPC. Usually it is a function depending on the inputs and outputs of the system which results with appropriate model of the system. In more complicated systems Kalman filter is used to estimate the system states, and afterwards the algorithm selects the appropriate model.

IV. Simulation of MMMPC on a Water Tank Model

In this paper we will use multiple-model MPC algorithm control a nonlinear reservoir tank model. This model is originally proposed in MathWorks Inc. MATLAB software.

Water enters a tank from the top and leaves through an orifice in its base. The rate that water enters is proportional to the voltage, *V*, applied to the pump. The rate that water leaves is proportional to the square root of the height of water in the tank. The water tank is illustrated on Fig. 2.

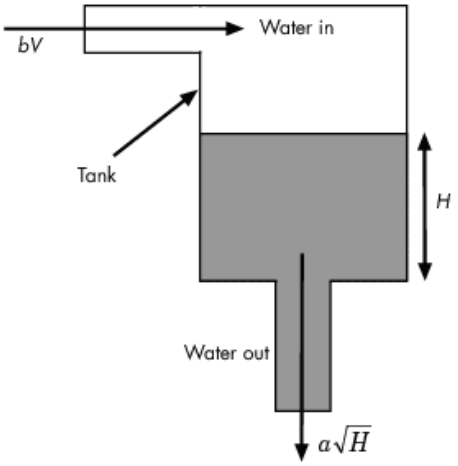


Fig. 2. Schematic Diagram for the Water-Tank System

The dynamics of the process can be described with the following equation

$$\frac{d}{dt}Vol = A\frac{dH}{dt} = bV - \sqrt{H} \tag{7}$$

where *Vol* is the volume of water in the tank, *A* is the cross-sectional area of the tank, *b* is a constant related to the flow rate into the tank, and *a* is a constant related to the flow rate out of the tank. The equation describes the height of water, *H*, as a function of time, due to the difference between flow rates into and out of the tank.

The equation contains one state, *H*, one input, *V*, and one output, *H*. It is nonlinear due to its dependence on the square-root of *H*.

The tank does have physical limitations. Maximum height ca ne 10 meter (that is the height of the tank), and the input voltage (which is proportional to the fluid flow in the tank) can vary between 0 and 24 VDC.

The linearization of the water tank model is realized using Matlab functions from the Control Systems Toolbox in five separate points depending of the only state of the system – height of the water tank *H*. Since the height is limited with both minimal and maximal value, we can easily decide which linearization points we should use. In this example we use linearization point at *H*=2m; *H*=4m; *H*=6m; *H*=8m; *H*=10m.

The realization of the switched model predictive controller is done it Matlab Simulink platform. The Simulink file containing the Switched MPC and the controlled process if presented on

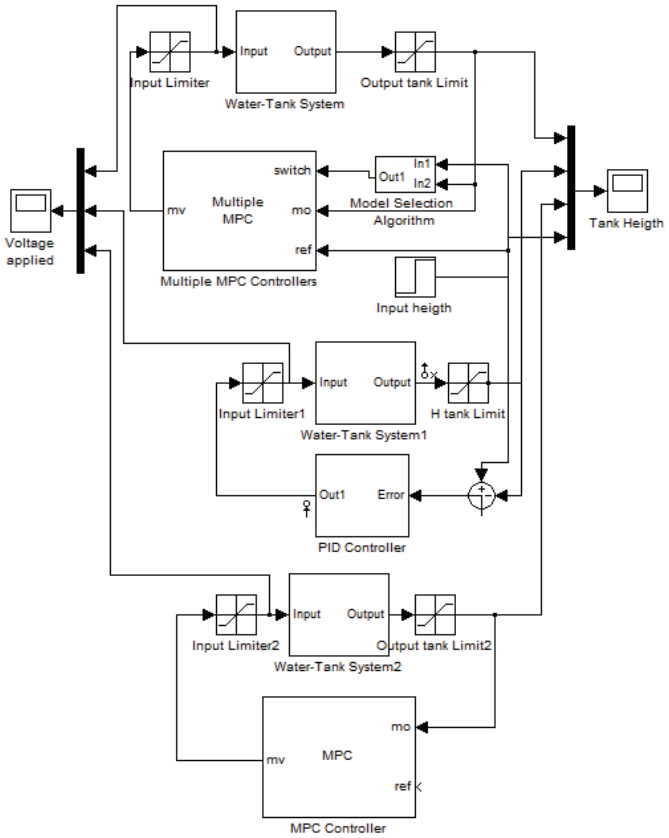


Fig. 3. Matlab Simulink model for simulation of the water tank system (M-M MPC, PID and MPC control systems, from top down)

We are presenting the results from an experiment for settling the water height at levels 2, 10 and 6 meters. The results are presents in figures 4, 5 and 6 retrospectively.

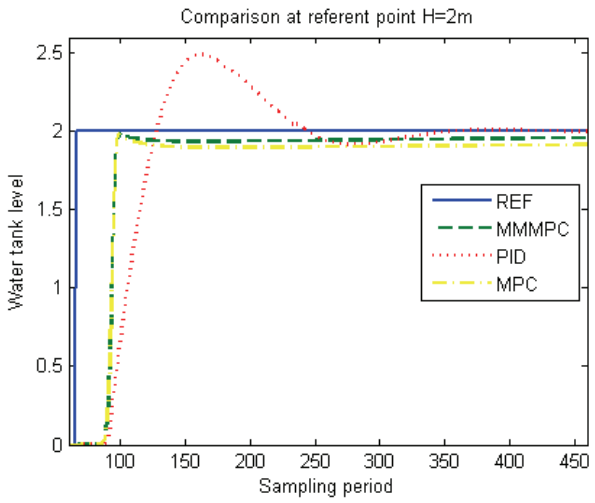


Fig. 4. Results from the simulation of the water tank system when we have referent height of 2 meters.

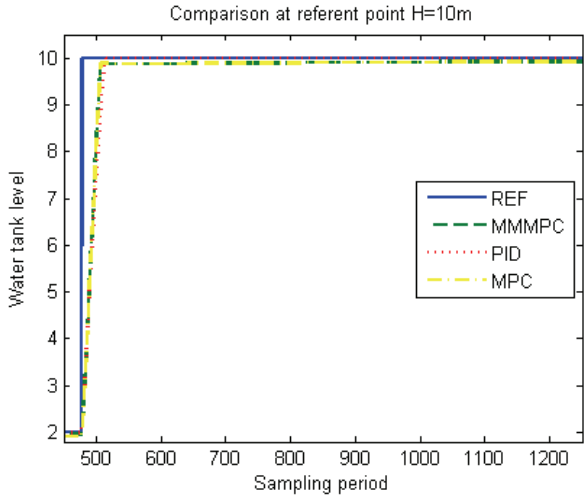


Fig. 5. Results from the simulation of the water tank system when we have referent height of 10 meters.

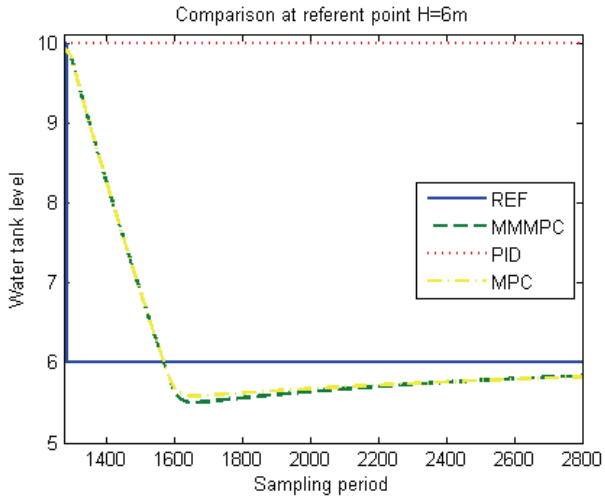


Fig. 6. Results from the simulation of the water tank system when we have referent height of 6 meters.

It is obvious that the multiple-models MPC has the best performance comparing with the PID and the MPC with only one point of linearization. On Fig. 4. we can notice that the PID controller (which is properly tuned) has significant overshot compared to the MPC controllers. Also the MMMPC has better performance compared with MPC linearized only in operating point with height 10 m.

On Fig. 5. we can notice that all tree controller have almost the same behavior. Nevertheless the PID controller is not “aware” of the physical limitation of the water tank, which can be seen from the control signal during this period. The input voltage is high compared to the control signals of the MMMPC and MPC controllers. During this period of time, the water tank is full, and yet the control signal is high, which results with overflowing of the water tank. Here we must point out that the model based predictive controllers have big advantage compared to the conventional PID, because of the possibility to enter the input and output constraints.

The results presented on Fig. 6. are directly affected by the previous state of the system. The MMMPC and the MPC,

have similar behavior, but as time passes, the MMMPC leans towards smaller steady state error than the MPC. In this case the PID controller continuously has the value of 10 m, which is result on an integrated value of the controller during the foregoing period.

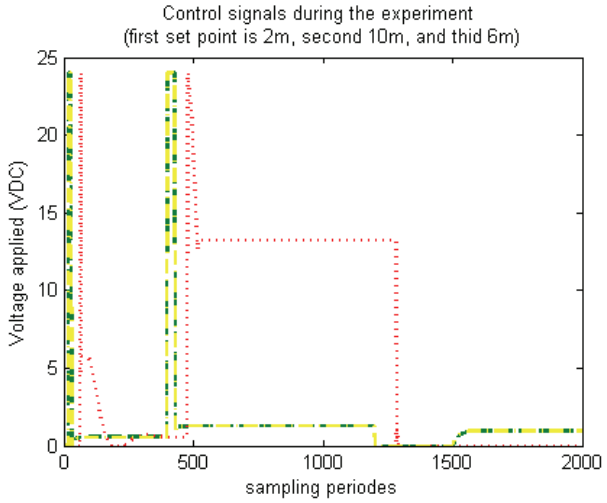


Fig. 7. Control signals during the whole experiment.

Fig 7. presents us the control signals during the experiment. On this picture we can notice that the control signals for the MPC and MMMPC are almost equivalent, but there is big difference compared to the PID control signal. The reasons why are explained before in the paper.

V. Conclusions

The model predictive control techniques have great advantage compared to the conventional control techniques, but still they are model-dependent. This means that the performance of the MPC is directly affected by the precision of the model. When we are controlling nonlinear processes and plants, the model mismatch will rather be bigger than smaller. In some particular cases we can easily partition the system and linearize in multiple points. We have shown that these linearized models can be used for constructing a multiple-model MPC, which has better characteristic compared to the MPC linearized around only one operating point.

There are possibilities for further research especially in defining proper model selection algorithm and determining the proper operation points of the system.

References

[1] F. Allgöwer, T. Badgwell, J. Qin, J. Rawlings, and S. Wright, “Nonlinear predictive control and moving horizon estimation - an introductory overview”, In P.M. Frank (ed.) *Advances in Control: Highlights of ECC'99*, pp. 391-449. Springer-Verlag, Berlin Heidelberg, DE 1999.

[2] G.M. Dimirovski, A. Dourado, E. Ikonen, U. Kortela, J. Pico, B. Ribeiro, M.J. Stankovski, and E. Tulunay, “Learning control of thermal systems”, in K. J. Astrorm, P. Albertos, M. Blanke, A. Isidori, W. Schaufelberger, R. Sanz (ed.), *Control of Complex Systems*, pp. 317-337. Springer-Verlag, London, UK 2001.

[3] G.M. Dimirovski, A.T. Dinibutun, M. Vukobratovic and J. Zhao, “Optimizing supervision and control for industrial furnaces: Predictive control based design”, in V. Sgurev, G. Dimirovski and M. Hadjiski (ed.), *Automatic Systems for Building the Infrastructure in Developing Countries – Global and Regional Aspects DECOM-04*, 17-28. Union for Automation & Informatics of Bulgaria and the IFAC, Sofia, BG 2004.

[4] Y. Jing, and G.M. Dimirovski, “Decentralized stabilization control for composite systems with time delays and uncertainties”, in B. Solaiman (ed.) *Proceedings of the 3rd IEEE Conference Information and Communication Technologies*, Damascus, Syria. Volume 1, 1404–1409. The IEEE, Piscataway, NJ, USA 2006.

[5] J. Zhao, “Hybrid control for global stabilization of a class of systems”. In T. Leung and H. Qin (eds.), *Advanced Topics in Nonlinear Control systems*, volume 40, Chapter 4, 129–160. World Scientific, Singapore, SG, 2001.

[6] B. Pregelj, and S. Gerkšič, “Multiple Model Approach to Multi-Parametric Model Predictive Control of a Nonlinear Process – simulation case study” http://as.utia.cas.cz/files/113.pdf

Switched Fuzzy Control Systems: Concepts for Simulation

Vesna M. Ojleska, Tatjana Kolemishevska-Gugulovska[1] and Georgi M. Dimirovski[2]

***Abstract* –In this paper we propose a concept for building simulation algorithms for switched fuzzy systems. Using these concepts we have shown the whole process of modeling, stability analysis and design of stabilizing controllers for the hovercraft-vehicle, as a typical nonholonomic system. MATLAB Simulator (Simulink model) and special MATLAB functions for the switched fuzzy controller of the hovercraft vehicle, along with the numerous simulation results, verify the correctness of the proposed concepts.**

***Keywords* – Control, fuzzy systems, hybrid systems, switched systems, switching, hovercraft.**

I. INTRODUCTION

Switched system is special class of hybrid dynamical systems, which consist of a family of continuous-time or discrete-time subsystems and a rule that orchestrates the switching among them. The last couple of decades have witnessed an enormous growth of interest of the class of switched systems in combination with the even larger class of hybrid systems [4][5][6][7], as these systems have a wide range of potential applications. From the middle of the 1980's, there have appeared a number of analysis/synthesis problems for Takagi-Sugeno (T-S) fuzzy systems [1][2][3].

Following the remarkable developments in theory, applications, and the industrial implementations of fuzzy control systems, recently [8] switched systems have been extended further to encompass *switched fuzzy systems*. In general, a switched fuzzy system involves fuzzy systems among its sub-systems or an alternative fuzzy-switching law, or the both (least explored case). Resent developments in this area promote a new direction in the control of dynamic systems [9] [10] [11] [12] [13] [14] [15], and it is clear that the field of switched fuzzy systems is becoming very popular.

To our best knowledge it may well be found that, up to know the most of the research in this area is focused on representation modeling, stability analysis and controller design, that guarantees stabilization and certain system performance. We are expecting that these results can be used as a good platform for solving real world control problems. Thus, in this paper we are exploring the performance of switched fuzzy control systems, and also we are proposing the simulation scheme algorithms, that can be used in modelling and design of switched fuzzy controllers for real world systems. Hovercraft vehicle, as a typical nonholonomic system, is used for verifying the simulations.

To begin with, in Section 2 we present the basic concepts of switched fuzzy systems with levels of structure, their representation modelling and stability analysis, taken from [9][10][11][12]. In Section 3 we are proposing the simulation scheme algorithms. In Section 4 we present the whole process of modelling and design of the switched fuzzy controllers, using the previously suggested simulation schemes, for the hovercraft vehicle. In Section 5 we give the essential conclusions of this study.

[1]Vesna M. Ojleska and Tatjana Kolemishevska-Gugulovska are with SS Cyril and Methodius University, Faculty of Electrical Engineering and Information Technologies (FEEIT), Skopje, R. of Macedonia, E-mail: {vojleska,tanjakg}@feit.ukim.edu.mk.

[2]Georgi M. Dimirovski is with Dogus University, School of Engineering, Istanbul, TR-34722, R. of Turkey, and with SS Cyril and Methodius University, FEEIT, Skopje, R. of Macedonia, E-mail: gdimirovski@dogus.edu.tr.

II. CONCEPTS OF SWITCHED FUZZY SYSTEMS, REPRESENTATION MODELLING AND STABILITY ANALYSIS

Since in [15] we have given a detailed overview of the achievements in the field of switched fuzzy systems, followed by the comparative study for this kind of systems, here, we will only provide a discussion of some of the key principles of the joint concept, using the representation modelling given in [9],[10],[11],[12], the so-called "switched fuzzy system with levels of structure".

The switching fuzzy model from [9][10][11][12] is given with:

Region Rule i :

$$\textbf{IF } z_1(t) \textbf{ is } N_{i1} \textbf{ and } \Lambda \textbf{ and } z_p(t) \textbf{ is } N_{ip}, \qquad (1)$$

THEN

Local Plant Rule l :

$$\textbf{IF } z_1(t) \textbf{ is } M_{il1} \textbf{ and } \Lambda \textbf{ and } z_p(t) \textbf{ is } M_{ilp},$$

$$\textbf{THEN} \begin{cases} \dot{x}(t) = A_{il}x(t) + B_{il}u(t), \\ y(t) = C_{il}x(t), \end{cases} \quad l = 1,2,\text{K}\ ,r \quad i = 1,2,\text{K}\ ,m$$

Here, m is the number of regions partitioned on the premise parts space. $N_{ij}(z(t))$ is a crisp set, where $N_{ij}(z(t)) = \begin{cases} 1 & , z(t) \in N_{ij} \\ 0 & , o.w \end{cases}$; r is the number of rules of the local models; M_{ilj} is fuzzy set; $x(t) \in R^n$ is the state vector, $u(t) \in R^m$ is the input vector, $y(t) \in R^q$ is the output vector, $A_{il} \in R^{n \times n}$, $B_{il} \in R^{n \times m}$, and $C_{il} \in R^{q \times n}$; $z(t) = \lfloor z_1(t), \text{K}\ , z_p(t) \rfloor$ are known premise variables that can be functions of the state variables, external disturbances, and/or time.

Region Rule i :

$$\textbf{IF } z_1(t) \textbf{ is } N_{i1} \textbf{ and } \Lambda \textbf{ and } z_p(t) \textbf{ is } N_{ip}, \qquad (2)$$

THEN

Local Control Rule l :

$$\textbf{IF } z_1(t) \textbf{ is } M_{il1} \textbf{ and } \Lambda \textbf{ and } z_p(t) \textbf{ is } M_{ilp},$$

$$\textbf{THEN} \quad u(t) = -F_{il}x(t), \qquad l = 1,2,\text{K}\ ,r \quad i = 1,2,\text{K}\ ,m.$$

From (1), it is clear that the switching fuzzy model has two levels of structure: region rule level and local fuzzy rule level. The switching fuzzy model (1) is inferred by fuzzily blending the linear system models $\dot{x}(t) = A_{il}x(t) + B_{il}u(t)$ and switching the global T-S fuzzy models, defined on every region.

In [9], authors propose a new PDC to design a stable switching fuzzy controller for the switching fuzzy system (1). The structure of

the PDC fuzzy controller is given with (2), where the design purpose is to determine the local feedback gains F_{il} in the consequent parts.

Along with the representation modelling of switched fuzzy systems, authors in [9][10] present LMI stability (relaxed stability) conditions for the system that consists of the switching fuzzy model (1) and the appropriate PDC controller, given with (2). Except the stability conditions, authors in [9]-[12] present the conditions for the constraints on control inputs.

III. Proposed Simulation Algorithms

The whole process for design of switched fuzzy controller using switched fuzzy model with levels of structure, for the certain nonlinear system is shown on Fig. 1.

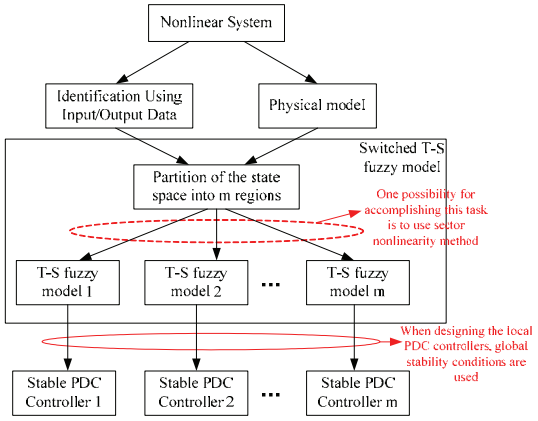


Fig. 1. Switched fuzzy controllers design.

In this paper, the simulations for performance analysis of the hovercraft-vehicle control will be made in few steps, which is shown on Fig. 2. This scheme for constructing the simulation environment can be also used for other nonlinear systems, if the parameters for the switched nonlinear model with levels of structure of a system are previously calculated. The final aim is to generate stable controllers which will satisfy certain conditions given in a form of LMIs.

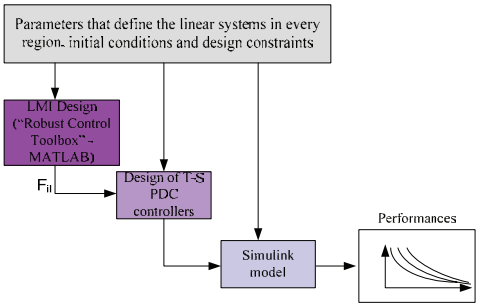


Fig. 2. Steps for controller design and performance evaluation for the given switched fuzzy model.

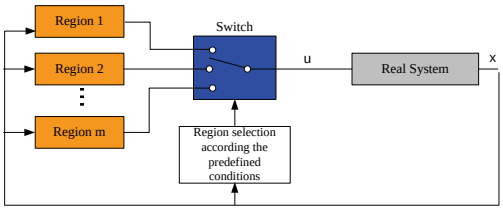


Fig. 3. Control of a real system with previously designed controllers, according the procedure given on Fig. 2.

After the local switched controllers are generated, according to the appropriate switched T-S fuzzy model with levels of structure, they can be used for controlling the real nonlinear system. Choosing the appropriate controller in certain moment (switching among these controllers) is made according the conditions which define the certain region selection. This is shown on Fig. 3. In every region there is an appropriate stable controller (in this case the controller is of PDC type).

IV. T-S Switched Fuzzy Model and Controller for the Hovercraft Vehicle

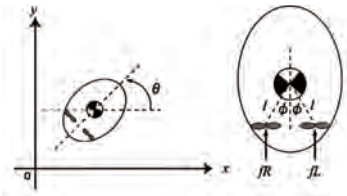


Fig. 4. Hovercraft model and its coordinate system [9].

Acording to hovercraft model of Fig. 4, the following state-space model of the hovercraft vehicle is given in [9]:

$$\dot{x}_1(t)=\dot{y}(t)=\frac{1}{M}\sin\theta(t)f_1(t),\ \dot{x}_2(t)=\dot{y}(t)=x_1(t),$$
$$\dot{x}_3(t)=\dot{\theta}(t)=\frac{l\sin\phi}{I}f_2(t),\ \dot{x}_4(t)=\dot{\theta}(t)=x_3(t) \qquad (3)$$

where $f_1(t)=f_R(t)+f_L(t)$, $f_2(t)=f_R(t)-f_L(t)$, θ is the angle of the vehicle; l is the distance between the gravity and fans; ϕ is the angle between the gravity and fans; f_R and f_L are the forces generated by the right and left side fans, respectively; M and I are the mass and the inertia, respectively. The control purpose is $\lim_{t\to\infty} y(t)=0$ and $\lim_{t\to\infty}\theta(t)=0$, by manipulating $f_R(t)$ and $f_L(t)$.

The hovercraft vehicle, is a typical nonholonomic system (the system is subject to constraints involving both, the position and velocity, whereupon it can not be stabilized by any continuous feedback law [4]). Further, we will show that this system can be stabilized with switching controller, based on the switched fuzzy-logic model.

A. Switched Fuzzy Model

Assuming that $\theta(t)\in[-179\ \ 179]$, we divide the premise variable space into three regions with nonnegative constant d. We will use sector nonlinearity concept for determing the local linear models in every region [3].

Region 1: $\theta(t)\geq d$

$\sin\theta(t)=\sum_{i=1}^{2}h_{1i}(\theta(t))a_{1i}$, where $h_{11}(\theta(t))=\frac{\sin(\theta(t))-a_{12}}{a_{11}-a_{12}}$ and $h_{12}(\theta(t))=\frac{a_{11}-\sin(\theta(t))}{a_{11}-a_{12}}$, where $a_{11}=1$ and $a_{12}=\sin(179^0)\approx 0.01$.

Region 2: $-d<\theta(t)<d$

In this region, we fix the first input, i.e. $f_1(t)=C$, where C is a positive constant. This implies $\dot{y}(t)=\frac{\sin\theta(t)}{M}C$.

$\sin\theta(t)=\left(\sum_{i=1}^{2}h_{2i}(\theta(t))a_{2i}\right)\theta(t)$, where $h_{21}(\theta(t))=\frac{\frac{\sin(\theta(t))}{\theta(t)}-a_{22}}{a_{21}-a_{22}}$ and $h_{22}(\theta(t))=\frac{a_{21}-\frac{\sin(\theta(t))}{\theta(t)}}{a_{21}-a_{22}}$, where $a_{21}=1$ and $a_{22}=\sin(d)/d$.

Region 3: $\theta(t) \le -d$

$\sin\theta(t) = \sum_{i=1}^{2} h_{3i}(\theta(t))a_{3i}$, where $h_{31}(\theta(t)) = \dfrac{\sin(\theta(t)) - a_{32}}{a_{31} - a_{32}}$ and $h_{32}(\theta(t)) = \dfrac{a_{31} - \sin(\theta(t))}{a_{31} - a_{32}}$, where $a_{31} = -1$ and $a_{32} = \sin(-179^0) \approx -0.01$.

By aggregating the above results, according the relation (1), we construct the following switched fuzzy model:

Region Rule 1 : **IF** $\theta(t) \ge d$, **THEN**
Local Plant Rule 1 : **IF** $\theta(t)$ **e** $h_{11}(\theta(t))$, **THEN** $\dot{x}(t) = A_{11}x(t) + B_{11}u(t)$
Local Plant Rule 2 : **IF** $\theta(t)$ **e** $h_{12}(\theta(t))$, **THEN** $\dot{x}(t) = A_{12}x(t) + B_{12}u(t)$

Region Rule 2 : **IF** $-d < \theta(t) < d$, **THEN**
Local Plant Rule 1 : **IF** $\theta(t)$ **e** $h_{21}(\theta(t))$, **THEN** $\dot{x}(t) = A_{21}x(t) + B_{21}u(t)$
Local Plant Rule 2 : **IF** $\theta(t)$ **e** $h_{22}(\theta(t))$, **THEN** $\dot{x}(t) = A_{22}x(t) + B_{22}u(t)$

Region Rule 3 : **IF** $\theta(t) \le -d$, **THEN**
Local Plant Rule 1 : **IF** $\theta(t)$ **e** $h_{31}(\theta(t))$, **THEN** $\dot{x}(t) = A_{31}x(t) + B_{31}u(t)$
Local Plant Rule 2 : **IF** $\theta(t)$ **e** $h_{32}(\theta(t))$, **THEN** $\dot{x}(t) = A_{32}x(t) + B_{32}u(t)$ (4)

where $u(t) = [f_1(t), f_2(t)]^T$ and $x(t) = [x_1(t) \;\; x_2(t) \;\; x_3(t) \;\; x_4(t)]$,

$$A_{11} = A_{12} = A_{31} = A_{32} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad A_{21} = \begin{bmatrix} 0 & 0 & 0 & \frac{C}{M}a_{21} \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

$$A_{22} = \begin{bmatrix} 0 & 0 & 0 & \frac{C}{M}a_{22} \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad B_{11} = \begin{bmatrix} \frac{1}{M}a_{11} & 0 \\ 0 & 0 \\ 0 & \frac{l\sin\phi}{I} \\ 0 & 0 \end{bmatrix}, \quad B_{12} = \begin{bmatrix} \frac{1}{M}a_{12} & 0 \\ 0 & 0 \\ 0 & \frac{l\sin\phi}{I} \\ 0 & 0 \end{bmatrix},$$

$$B_{21} = B_{22} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & \frac{l\sin\phi}{I} \\ 0 & 0 \end{bmatrix}, \quad B_{31} = \begin{bmatrix} \frac{1}{M}a_{31} & 0 \\ 0 & 0 \\ 0 & \frac{l\sin\phi}{I} \\ 0 & 0 \end{bmatrix}, \quad B_{32} = \begin{bmatrix} \frac{1}{M}a_{32} & 0 \\ 0 & 0 \\ 0 & \frac{l\sin\phi}{I} \\ 0 & 0 \end{bmatrix}.$$

B. Controller Design via Switched PDS

Region Rule 1 : **IF** $\theta(t) \ge d$, **THEN**
Local Control Rule 1 : **IF** $\theta(t)$ **is** $h_{11}(\theta(t))$, **THEN** $u(t) = -F_{11}x(t)$
Local Control Rule 2 : **IF** $\theta(t)$ **is** $h_{12}(\theta(t))$, **THEN** $u(t) = -F_{12}x(t)$

Region Rule 2 : **IF** $-d < \theta(t) < d$, **THEN**
Local Control Rule 1 : **IF** $\theta(t)$ **is** $h_{21}(\theta(t))$, **THEN** $u(t) = -F_{21}x(t)$
Local Control Rule 2 : **IF** $\theta(t)$ **is** $h_{22}(\theta(t))$, **THEN** $u(t) = -F_{22}x(t)$

Region Rule 3 : **IF** $\theta(t) \le -d$, **THEN**
Local Control Rule 1 : **IF** $\theta(t)$ **is** $h_{31}(\theta(t))$, **THEN** $u(t) = -F_{31}x(t)$
Local Control Rule 2 : **IF** $\theta(t)$ **is** $h_{32}(\theta(t))$, **THEN** $u(t) = -F_{32}x(t)$ (5)

The switching fuzzy controller of PDC type for the switched fuzzy model (4) can be designed accoding the relation (2), having the form (5). From (5), it is obvious that the design process depends on the local feedback gains F_{il}, which can be obtained if there is a fisible solution to the LMI stability conditions given in [9], [10].

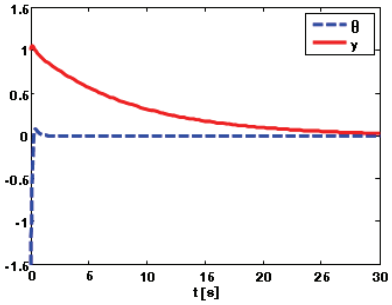
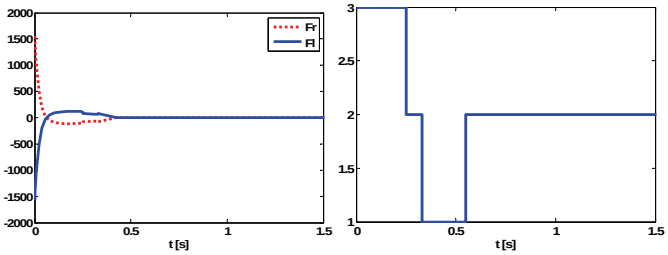


Fig. 5. Values for θ and y, using the control signals, shown on Fig. 6, when the LMI stability conditions from [9], [10] are used. Initial conditions are $x(0) = [0 \;\; 1 \;\; 0 \;\; -1.5]^T$ and (M=0.1, $\phi = \pi/4$, I=0.5, l=0.1, C=0.5, $d = \pi/50$).

From Fig.5 it is clear that the control purpose ($\lim_{t\to\infty} y(t) = 0$ and $\lim_{t\to\infty} \theta(t) = 0$) is achievable. Showing this it is obvious that the use of switched fuzzy controllers is worthwhile in comparison of ordinary T-S fuzzy controllers.



a) b)

Fig. 6. a) Control inputs $f_R(t)$ and $f_L(t)$; b).Region selection.

However, if we view the control effort, i.e. the values of $f_R(t)$ and $f_L(t)$, (see fig. 6-a), it is obvious that they have very high amplitudes. The respective switching signal that selects the appropriate region during the control period is shown on Fig. 6-b.

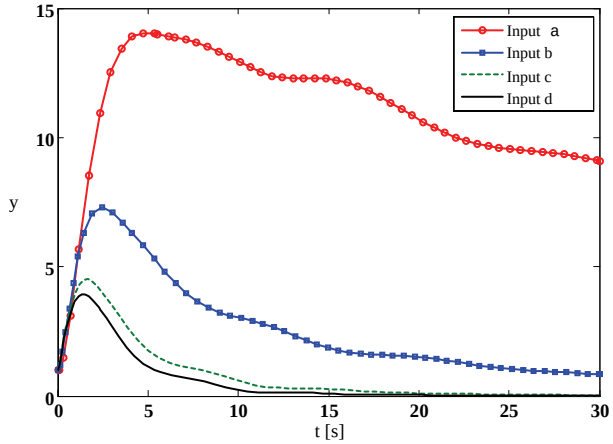


Fig. 7. Values for y, using the control signals, shown on Fig. 9.

For constraining the control inputs $f_R(t)$ and $f_L(t)$, we can use the LMI conditions given in [11], [12]. Please note that those LMI

conditions depend on the initial conditions, so whenever we would like to change the initial conditions it would be necessary to recalculate the feedback control gains.

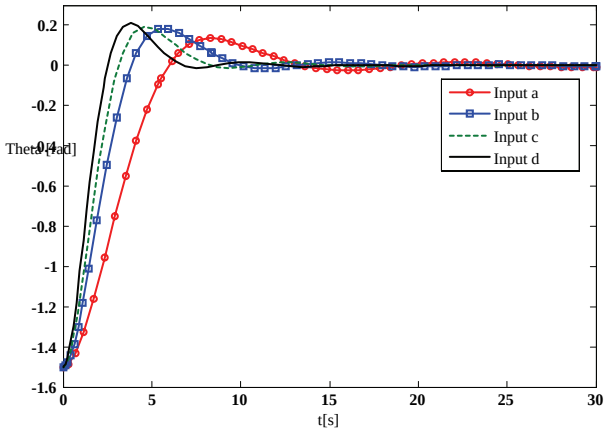


Fig. 8. Values for θ, using the control signals, shown on Fig. 9.

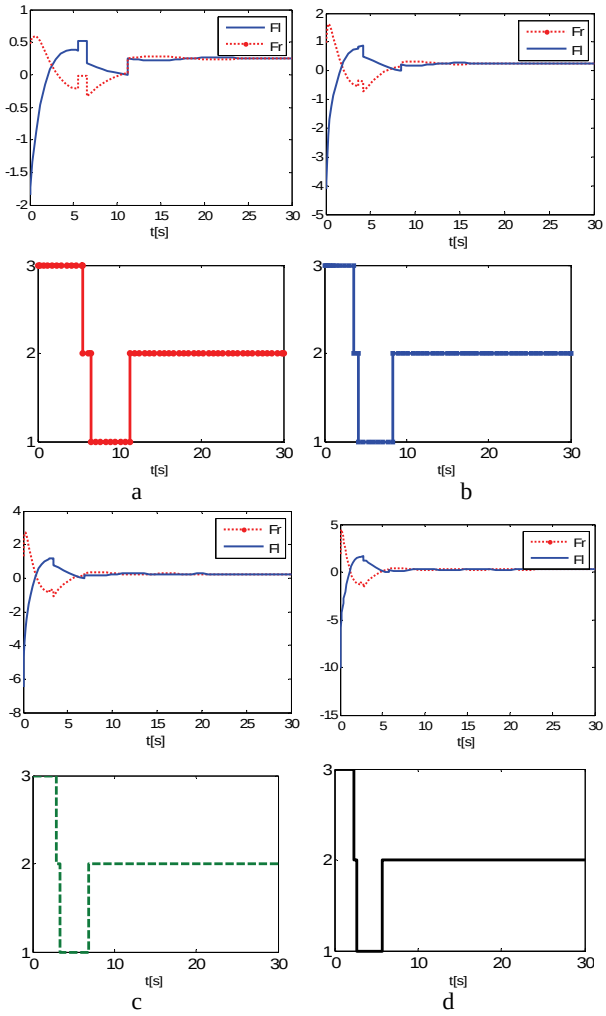


Fig. 9. Control inputs $f_R(t)$ and $f_L(t)$ and the appropriate switching signal, when the joined LMI stability conditions, along with the LMI conditions for the constraints on the control input are used. ($x(0)=[0 \quad 1 \quad 0 \quad -1.5]^T$, M=0.1, $\phi=\pi/4$, I=0.5, l=0.1, C=0.5, $d=\pi/50$).

Fig. 7 and Fig. 8 show the values of the control variables for the four different constraint inputs, given on Fig. 9-a,b,c,d. It is obvious that as much as the control inputs are constraint, as hard as the control purpose is achieved.

V. Conclusion

We have showed that the proposed mechanism for modeling and simulation of the performance of switched fuzzy systems is applicable for building and simulating the switched fuzzy control of the hovercraft vehicle, whereas we expect that these concepts can be used for building and simulating the switched fuzzy control of other nonlinear systems. Also, with the derived simulation results that show the feasibleness of the use of switched fuzzy controllers for the systems that are not stabilizible with any continuous feedback control laws, we showed that these types of controllers can be used in solving many control problems that are not solvable with other types of controllers.

References

[1] G. Feng, "A Survey on Analysis and Design of Model-Based Fuzzy Control Systems," *IEEE Transactions on Fuzzy Systems*, vol. 14, no. 5, October 2006.

[2] O. Kaynak, K. Erbatur, and M. Ertugnrl, "The fusion of computationally intelligent methodologies and sliding-mode control—A survey," *IEEE Trans. Ind. Electron.*, vol. 48, no. 1, pp. 4–17, Feb. 2001.

[3] K. Tanaka, H. O. Wang, *Fuzzy Control Systems Design and Analysis: A Linear Matrix Inequality Approach*. New York: John Wiley & Sons, 2001.

[4] D. Liberzon, *Switching in Systems and Control*. Boston, MA: Birkhauser, 2003.

[5] Z. Sun, S. S. Ge, *Switched Linear Systems - Control and Design*. London: Springer, 2005.

[6] H. Lin, P. J. Antsaklis, "Stability and Stabilizability of Switched Linear Systems: A Survey of Recent Results," *IEEE Transactions on Automatic Control*, vol. 54, NO. 2, February 2009.

[7] J. Zhao, G. M. Dimirovski, "Quadratic stability of a Class of Switched Nonlinear Systems," *IEEE Transactions on Automatic Control*, vol. 49, no. 4, April 2004.

[8] R. Palm, D. Driankov, "Fuzzy Switched Hybrid Systems - Modeling and Identification", Proc. of the *IEEE ISIC/CIRA/ISAS Joint Conference*, Gaithersburg, MD, September 14-17, 1998, pp. 130-135.

[9] K. Tanaka, M. Iwasaki, H. O. Wang, "Stabilization of switching fuzzy systems," *in Proc. of Asian Fuzzy System Symposium, 2000.*

[10] K. Tanaka, M. Iwasaki, H. O. Wang, "Stable switching fuzzy control and its application to a hovercraft type vehicle", in Proc. of the *Ninth IEEE International Conference on Fuzzy Systems, 2000*, Vol. 2, pp. 804–809.

[11] K. Tanaka, M. Iwasaki, H. O. Wang, "Stability and smoothness conditions for switching fuzzy systems," in *Proc. of the 2000 American Control Conference*, Chicago, Ilinois, June, 2000, pp. 2474-2478.

[12] K. Tanaka, M. Iwasaki, H. O. Wang, "Switching Control of an R/C Hovercraft: Stabilization and Smooth Switching," *IEEE Transactions on Systems, Man, and Cybernetics* Part B: vol. 31, no. 6, December 2001.

[13] H. Yang, G. M. Dimirovski, and J. Zhao, "Switched Fuzzy Systems: Representation Modelling, Stability Analysis and Control Design," in *Studies in Computational Intelligence 109*, pp. 155-168, Springer-Verlag, Berlin Heidelberg, DE, 2008.

[14] **V. Ojleska**, G. Stojanovski, "Switched fuzzy systems: Overview and Perspectives," *The 9 International PhDWorkshop on Systems and Control*, Izola, Slovenia, October, 2008.

[15] **V. Ojleska**, T. K.-Gugulovska, G. Dimirovski, "A Survey on Modeling, Analysis and Design of Switched Fuzzy Control Systems", *Proceedings of DECOM-IFAC'09*, Ohrid, September 26-29, Macedonia.

Sliding Hyperplane Design for Linear Systems with Unmatched External Disturbance Vector

Boban Veselić[1], Branislava Peruničić-Draženović[2] and Čedomir Milosavljević[3]

***Abstract* – A class of linear systems subjected to an unmatched external disturbance is considered within the framework of sliding mode control (SMC) theory. The paper offers a sliding hyperplane design method to minimize the effects of the unmatched disturbance upon the SM dynamics. The optimization criterion is minimization of the steady state vector norm. The suggested approach has been demonstrated on a numerical example.**

***Keywords* –Sliding mode control, Matching conditions, Unmatched disturbances, Hyperplane design.**

I. INTRODUCTION

The most prestigious property of the variable structure control systems (VSCSs) in sliding mode (SM) [1] is their invariance to parameter and exogenous disturbances, under so called matching conditions [2]. Physical meaning of these conditions is that disturbance acts through the control channel, i.e. the control is able to change coincidentally with the disturbance. Although there are many examples of systems where this requirement is fulfilled, there are still systems that structurally do not meet these conditions. While VSCSs in SM are completely insensitive to the matched uncertainties, on the other hand the SM dynamics, which is prescribed by the choice of the sliding manifold, is vulnerable to the unmatched uncertainties. In some cases sliding motion along certain manifolds may result in severe dynamics deterioration under action of unmatched uncertainties.

Several approaches are present in dealing with unmatched uncertainties, within the context of SM control (SMC): a continuous nonlinear control strategy [3], dynamical approach by means of pseudo-control inputs introduction into the reduced order system [4]-[6], a new invariance condition [7] in terms of linear matrix inequalities (LMI). All mentioned methods consider unmatched uncertainties having parametric nature, which diminish as system states approach origin. Therefore asymptotic stability can be ensured.

When it comes to the unmatched uncertainties that besides parametric contain external disturbances as well, asymptotic stability can not be attained. Although it is possible to establish a SM in such systems, the unmatched part of an external disturbance has impact on SM dynamics, forcing the system trajectory not to converge to the origin but to wander in its neighborhood along the sliding manifold. One way of addressing this problem may be to construct a sliding manifold that reduces system sensitivity in SM upon unmatched disturbances. A sliding manifold design that minimizes equivalent perturbation is suggested in [8], whereas another sliding manifold selection is proposed in [9], using the invariant ellipsoid method.

As in [8] and [9], this paper also searches for an adequate sliding hyperplane selection in a class of linear systems with scalar control and bounded unmatched external disturbance vector, which minimizes in some sense the impact of the disturbance onto the SM motion. The chosen optimization criterion here is minimization of the steady state vector norm. A systematic procedure for the sliding hyperplane design is developed, guaranteeing minimal static error. The proposed design method has been investigated on a numerical example.

II. SM IN CASE OF UNMATCHED DISTURBANCES

Consider a class of linear systems with scalar control that can be represented by the following state space model

$$\dot{\mathbf{x}}(t) = \mathbf{A}x(t) + \mathbf{b}u(t) + \mathbf{d}(t), \tag{1}$$

where $\mathbf{x} \in \Re^{n\times 1}$ is the state vector, $u \in \Re$ is the control signal, $\mathbf{A} \in \Re^{n\times n}$ is the system matrix, $\mathbf{b} \in \Re^{n\times 1}$ is the control input vector and $\mathbf{d} \in \Re^{n\times 1}$ is a bounded external disturbance vector, $\| \mathbf{d}(t) \| \le M < \infty$, $\forall t > 0$. It is assumed that the pair $(\mathbf{A}, \mathbf{b})$ is controllable and the matching condition [2]

$$\mathbf{d}(t) = \mathbf{b}\tilde{d}(t), \tag{2}$$

is not fulfilled. This means that there does not exist a function $\tilde{d}(t)$ satisfying (2). Let a sliding hyperplane $\sigma(\mathbf{x}) = 0$ be defined by the switching function

$$\sigma(\mathbf{x}) = \mathbf{c}\mathbf{x}, \quad \mathbf{c} = [c_1, c_2, \Lambda\ , c_n], \tag{3}$$

which includes the state space origin as the equilibrium point.

SM reaching and existence condition $\sigma\dot{\sigma} \le -\eta\,|\sigma|$, $\eta > 0$ can be attained by an appropriate discontinuous control

$$u = -(\mathbf{cb})^{-1}(\mathbf{cAx} + \alpha \operatorname{sgn}\sigma) \tag{4}$$

if the switching gain overcomes disturbance $\alpha > \| \mathbf{c} \| M$.

[1]B. Veselić is with the Faculty of Electronic Engineering, University of Niš, 18000 Niš, Serbia, E-mail: boban.veselic@elfak.ni.ac.rs.

[2]B. Peruničić-Draženović is with the Faculty of Electrical Engineering, University of Sarajevo, 71000 Sarajevo, Bosnia and Herzegovina, E-mail: brana_p@hotmail.com.

[3]Č. Milosavljević is with the Faculty of Electrical Engineering, University of Istočno Sarajevo, 71000 Sarajevo, Bosnia and Herzegovina, E-mail: cedomir.milosavljevic@elfak.ni.ac.rs

Hence, SM along $\sigma(\mathbf{x})=0$ will be attained in a finite time. System dynamics in ideal SM is obtained by transformation of (1) into the regular form by coordinate change $\mathbf{x}=\mathbf{P}_1\bar{\mathbf{x}}$, where the transformation matrix is given by [1]

$$\mathbf{P}_1=\mathbf{M}_c\begin{bmatrix} a_2 & a_3 & \Lambda & a_n & 1\\ a_3 & a_4 & \Lambda & 1 & \\ \mathrm{M} & \mathrm{M} & \mathrm{O} & & \\ a_n & 1 & & \mathbf{0} & \\ 1 & & & & \end{bmatrix}. \tag{5}$$

$\mathbf{M}_c$ is the controllability matrix and a_i, $i=1,\Lambda,n$ are the coefficients of the characteristic polynomial $\det(s\mathbf{I}-\mathbf{A})=s^n+a_ns^{n-1}+\Lambda+a_2s+a_1$, where s is a complex variable. The resulting regular form is obtained as

$$\begin{aligned}
&\dot{\bar{\mathbf{x}}}(t)=\bar{\mathbf{A}}\bar{\mathbf{x}}(t)+\bar{\mathbf{b}}u(t)+\bar{\mathbf{d}}(t),\ \bar{\mathbf{x}}=\begin{bmatrix}\bar{\mathbf{x}}_r & \bar{x}_n\end{bmatrix}^{\mathrm{T}},\\
&\bar{\mathbf{d}}=\mathbf{P}_1^{-1}\mathbf{d}=\begin{bmatrix}\bar{\mathbf{d}}_r\\ \bar{d}_n\end{bmatrix},\ \bar{\mathbf{A}}=\mathbf{P}_1^{-1}\mathbf{A}\mathbf{P}_1=\begin{bmatrix}\mathbf{A}_{11} & \mathbf{a}_{12}\\ \mathbf{a}_{21} & a_{22}\end{bmatrix},\ \bar{\mathbf{b}}=\mathbf{P}_1^{-1}\mathbf{b}=\begin{bmatrix}\mathbf{0}\\ 1\end{bmatrix}\\
&\bar{\mathbf{x}}_r=[\bar{x}_1\quad \bar{x}_2\quad \Lambda\quad \bar{x}_{n-1}]^{\mathrm{T}},\ \bar{\mathbf{d}}_r=[\bar{d}_1\quad \bar{d}_2\quad \Lambda\quad \bar{d}_{n-1}]^{\mathrm{T}},\\
&\mathbf{A}_{11}=\begin{bmatrix}0&1&0&\Lambda&0\\0&0&1&\Lambda&0\\ \mathrm{M}&\mathrm{M}&\mathrm{M}&&\\0&0&0&\Lambda&1\\0&0&0&\Lambda&0\end{bmatrix}_{n-1\times n-1},\ \mathbf{a}_{12}=\begin{bmatrix}0\\0\\ \mathrm{M}\\0\\1\end{bmatrix},\\
&\mathbf{a}_{21}=\begin{bmatrix}-a_1 & -a_2 & \Lambda & -a_{n-2} & -a_{n-1}\end{bmatrix},\ a_{22}=-a_n.
\end{aligned} \tag{6}$$

The switching function may be rewritten using the regular form as $\sigma=\mathbf{c}\mathbf{x}=\mathbf{c}\mathbf{P}_1\bar{\mathbf{x}}=\bar{\mathbf{c}}\bar{\mathbf{x}}$, $\bar{\mathbf{c}}=\mathbf{c}\mathbf{P}_1$, $\bar{\mathbf{c}}=\begin{bmatrix}\bar{\mathbf{c}}_0 & 1\end{bmatrix}$, with vector $\bar{\mathbf{c}}_0=\begin{bmatrix}\bar{c}_1 & \bar{c}_2 & \Lambda & \bar{c}_{n-1}\end{bmatrix}$. In SM $\sigma(\mathbf{x})=\bar{\mathbf{c}}_0\bar{\mathbf{x}}_r+\bar{x}_n=0$, which gives $\bar{x}_n=-\bar{\mathbf{c}}_0\bar{\mathbf{x}}_r$, meaning that the SM dynamics is described by a reduced order system

$$\begin{aligned}
&\dot{\bar{\mathbf{x}}}_r=(\mathbf{A}_{11}-\mathbf{a}_{12}\bar{\mathbf{c}}_0)\bar{\mathbf{x}}_r+\bar{\mathbf{d}}_r=\mathbf{A}_r\bar{\mathbf{x}}_r+\bar{\mathbf{d}}_r,\\
&\bar{x}_n=-\bar{\mathbf{c}}_0\bar{\mathbf{x}}_r.
\end{aligned} \tag{7}$$

For the controllable pair $(\mathbf{A},\mathbf{b})$, the pair $(\mathbf{A}_{11},\mathbf{a}_{12})$ is also controllable and by selection of vector $\bar{\mathbf{c}}_0$ the eigenvalues of the system matrix $\mathbf{A}_r=(\mathbf{A}_{11}-\mathbf{a}_{12}\bar{\mathbf{c}}_0)$ can be adjusted. Characteristic equation of the system in SM is given by

$$\det(s\mathbf{I}-\mathbf{A}_r)=s^{n-1}+\textstyle\sum_{i=1}^{n-1}\bar{c}_is^{i-1}=0. \tag{8}$$

In systems where matching condition (2) is fulfilled, $\bar{\mathbf{d}}_r=\mathbf{0}$ and the system is invariant to disturbances in SM, confirmed by (7). Its dynamics (8) exclusively depends on the sliding hyperplane parameters. System (7) becomes autonomous and the origin is equilibrium point. In an unmatched case, $\bar{\mathbf{d}}_r\neq\mathbf{0}$ so the SM dynamics (7) is affected by a disturbance. A new equilibrium point on the sliding hyperplane is formed depending on the disturbances, since the SM existence conditions are satisfied by the control signal. Steady state of the system (7) can be easily evaluated using the following relations

$$\begin{aligned}
&\bar{\mathbf{x}}_r(\infty)=\lim_{s\to 0}\{s(s\mathbf{I}-\mathbf{A}_r)^{-1}[\bar{\mathbf{x}}_r(0)+\bar{\mathbf{d}}_r(s)]\},\\
&\bar{x}_n(\infty)=-\bar{\mathbf{c}}_0\bar{\mathbf{x}}_r(\infty),
\end{aligned} \tag{9}$$

where $\bullet(\infty)=\lim_{t\to\infty}\bullet(t)$. Since the sliding hyperplane parameters ensures stable eigenvalues of $\mathbf{A}_r$, the system motion caused only by initial condition $\bar{\mathbf{x}}_r(0)$ will converge into the origin. Therefore, (9) may be expressed as

$$\begin{aligned}
&\bar{\mathbf{x}}_r(\infty)=\Phi_{r0}(\bar{\mathbf{c}}_0)\bar{\mathbf{d}}_r(\infty),\\
&\bar{x}_n(\infty)=-\bar{\mathbf{c}}_0\bar{\mathbf{x}}_r(\infty),\\
&\Phi_{r0}(\bar{\mathbf{c}}_0)=\lim_{s\to 0}(s\mathbf{I}-\mathbf{A}_r)^{-1}=\begin{bmatrix}\frac{\bar{c}_2}{\bar{c}_1}&\frac{\bar{c}_3}{\bar{c}_1}&\Lambda&\frac{\bar{c}_{n-1}}{\bar{c}_1}&\frac{1}{\bar{c}_1}\\ -1&0&\Lambda&0&0\\0&-1&\Lambda&0&0\\ \mathrm{M}&\mathrm{M}&&\mathrm{M}&\mathrm{M}\\0&0&\Lambda&-1&0\end{bmatrix}
\end{aligned} \tag{10}$$

Steady state (10) will exist if $\bar{\mathbf{d}}_r(\infty)$ exists. Based on (10), using (6) and (7), the transformed steady state can be calculated as

$$\bar{\mathbf{x}}(\infty)=[a(\bar{\mathbf{d}}_r(\infty),\bar{\mathbf{c}}_0)\ \vdots\ -\bar{\mathbf{d}}_r(\infty)]^{\mathrm{T}}, \tag{11}$$

$$a(\bar{\mathbf{d}}_r(\infty),\bar{\mathbf{c}}_0)=[\bar{d}_{n-1}(\infty)+\sum_{i=1}^{n-2}\bar{d}_i(\infty)\bar{c}_{i+1}]/\bar{c}_1, \tag{12}$$

showing that it depends on the disturbance steady state and the sliding hyperplane coefficients. The only way to intentionally affect the steady state error is by means of sliding hyperplane selection. Hence, apart from defining the SM dynamics, hyperplane design also has impact on the system sensitivity to unmatched disturbances.

Original steady state is obtained using coordinate transformation $\mathbf{x}(\infty)=\mathbf{P}_1\bar{\mathbf{x}}(\infty)$. If $\mathbf{P}_1$ is expressed in general form

$$\mathbf{P}_1=\begin{bmatrix}p_{11}&\Lambda&p_{1n}\\ \mathrm{M}&&\\ p_{n1}&\Lambda&p_{nn}\end{bmatrix}, \tag{13}$$

original steady state is obtained as

$$\mathbf{x}(\infty)=\begin{bmatrix}a(\bar{\mathbf{d}}_r(\infty),\bar{\mathbf{c}}_0)p_{11}-\sum_{i=1}^{n-1}\bar{d}_i(\infty)p_{1,(i+1)}\\ a(\bar{\mathbf{d}}_r(\infty),\bar{\mathbf{c}}_0)p_{21}-\sum_{i=1}^{n-1}\bar{d}_i(\infty)p_{2,(i+1)}\\ \mathrm{M}\\ a(\bar{\mathbf{d}}_r(\infty),\bar{\mathbf{c}}_0)p_{n1}-\sum_{i=1}^{n-1}\bar{d}_i(\infty)p_{n,(i+1)}\end{bmatrix}. \tag{14}$$

III. Sliding Hyperplane Design

Since SM dynamics is sensitive to unmatched disturbances, a sliding hyperplane design procedure will be derived that provides minimization of the steady state error for the considered class of linear systems with unmatched external disturbance. Euclid vector norm $\| \mathbf{x} \| = (\mathbf{x}^{\mathrm{T}}\mathbf{x})^{1/2}$ may serve as a measure of the steady state distance from the origin. Therefore, the norm of the steady state (14) is obtained as

$$\left\|\mathbf{x}(\infty)\right\|^2 = \sum_{k=1}^{n} (a(\bar{\mathbf{d}}_r(\infty), \bar{\mathbf{c}}_0)\, p_{k1} - \sum_{i=1}^{n-1} \bar{d}_i(\infty)\, p_{k,(i+1)})^2 . \quad (15)$$

The only adjustable parameter $a(\bar{\mathbf{d}}_r(\infty), \bar{\mathbf{c}}_0)$ will be determined according to the condition that the norm (15) has its minimum for $a = a_m$, i.e.

$$\frac{\partial \left\|\mathbf{x}(\infty)\right\|^2}{\partial a}\Big|_{a=a_m} = 0 . \quad (16)$$

By differentiation of (15) with respect to *a*, (16) transforms into the following equation

$$a_m \sum_{k=1}^{n} p_{k1}^2 - \sum_{k=1}^{n} p_{k1} \sum_{i=1}^{n-1} \bar{d}_i(\infty)\, p_{k,(i+1)} = 0 , \quad (17)$$

whose solution is

$$a_m = \left(\sum_{k=1}^{n} p_{k1}^2 \right)^{-1} \sum_{k=1}^{n} p_{k1} \sum_{i=1}^{n-1} \bar{d}_i(\infty)\, p_{k,(i+1)} . \quad (18)$$

Now, vector $\bar{\mathbf{c}}_0$ should be calculated form (12) using the obtained solution a_m. (12) is a single equation with *n*-1 unknowns that cannot be uniquely solved. Solution can be reached only by reducing the number of degrees of freedom to one. By setting $\bar{\mathbf{c}}_0$, and consequently the vector $\mathbf{c}$, $n-1$ eigenvalues of the SM dynamics (8) can be independently chosen. Reduction of degrees of freedom can be done by imposing certain conditions to the eigenvalues. For example, if a single multiple eigenvalue $s = -s_p$, $s_p > 0$ of order $n-1$ is required, number of degrees of freedom is one. The desired SM characteristic equation can be expressed as

$$(s + s_p)^{n-1} = \sum_{i=0}^{n-1} \frac{(n-1)!}{i!(n-1-i)!}\, s^{n-1-i} s_p^i = 0 . \quad (19)$$

Comparison of (19) with (8) gives the needed sliding hyperplane coefficients as functions of s_p

$$\bar{c}_i = \frac{(n-1)!}{(n-i)!(i-1)!}\, s_p^{n-i},\ i = 1, \Lambda\ , n-1,\ \bar{c}_1 = s_p^{n-1} . \quad (20)$$

Using (20), the equation (12) becomes

$$a_m s_p^{n-1} - \sum_{i=1}^{n-2} \frac{(n-1)!}{(n-1-i)!(i-1)!}\, \bar{d}_i(\infty) s_p^{n-1-i} - \bar{d}_{n-1}(\infty) = 0 . \quad (21)$$

Recall that a_m, which minimize the norm (15), has been already determined using (18), hence (21) is a polynomial equation of order $n-1$ with respect to unknown s_p. This equation is now solvable and has $n-1$ solutions. The number of positive real solutions indicates the number of possible different stabile hyperplanes that minimizes the norm (15). Naturally, the largest positive real solution should be selected because it guaranties the fastest SM dynamics. For the selected s_p, the components of vector $\bar{\mathbf{c}}_0$ are calculated according to (20). Finally, the original vector $\mathbf{c}$ of the sliding hyperplane can be determined using the inverse transformation.

$$\mathbf{c} = \left[\bar{\mathbf{c}}_0 \ \vdots\ 1 \right] \mathbf{P}_1^{-1} . \quad (22)$$

IV. A Numerical Example

Consider a stable controllable third order linear system whose model (1) is represented by

$$\mathbf{A} = \begin{bmatrix} -1 & 0 & 1 \\ 1 & -2 & 1 \\ 0 & 0 & -3 \end{bmatrix},\ \mathbf{b} = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix},\ \mathbf{d}(t) = \begin{bmatrix} 2 \cdot h(t-6) \\ -1 \cdot h(t-5) \\ 0 \end{bmatrix} . \quad (23)$$

Note that the disturbance vector is unmatched since condition (2) can never be reached. The transformation matrix and the regular form are defined by

$$\mathbf{P}_1 = \begin{bmatrix} -2 & -1 & 0 \\ 1 & 3 & 1 \\ -2 & -3 & -1 \end{bmatrix}, \bar{\mathbf{A}} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -6 & -11 & -6 \end{bmatrix}, \bar{\mathbf{b}} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$
$$\bar{\mathbf{d}}(t) = \begin{bmatrix} \bar{\mathbf{d}}_r(t) \\ \hline \bar{d}_n \end{bmatrix} = \begin{bmatrix} 1 \cdot h(t-5) \\ -2 \cdot h(t-6) - 2 \cdot h(t-5) \\ \hline 6 \cdot h(t-6) + 4 \cdot h(t-5) \end{bmatrix} . \quad (24)$$

Relation (18) gives $a_m = -0.11$ that according to (14) ensures the steady state $\mathbf{x}(\infty) = \left[1.22 \quad 0.89 \quad -0.78\right]^{\mathrm{T}}$, i.e. the norm $\left\|\mathbf{x}(\infty)\right\| = 1.7$. Two eigenvalues, defined with $s_{p1} = -19.8$ and $s_{p2} = 1.8$, are obtained by solving (21) for $a_m = -0.11$. Positive real value $s_p = 1.8$ is selected in sliding plane determination, which according to (20) and (22) produces $\mathbf{c} = \left[-0.63 \quad -0.03 \quad -1.03\right]$.

A very simple control law (4) is employed in the realization of SM. Switching gain $\alpha = 5$ provides reaching and existing conditions in the presence of the disturbance. The control signal is depicted in Fig. 1a. It can be noticed that the high-frequency switching component is dominant, which would inevitably induce unwanted chattering in a real system. However, the reason for the application of such control algorithm is that the attained SM as much as possible resembles an ideal SM, which is the assumption of the conducted analysis. Fig. 1b shows the switching function,

whose annulment confirms occurrence and existence of SM, even after the action of the unmatched disturbance.

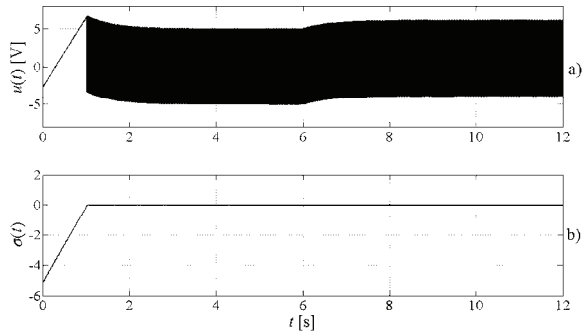


Fig. 1. a) Control signal; b) Switching function.

System trajectory in the phase space is depicted in Fig. 2. From the initial state $\mathbf{x}(0)=\begin{bmatrix}5 & -3 & 2\end{bmatrix}^{\mathrm{T}}$ system trajectory reaches and slides along the plane into the origin. After the action of the unmatched disturbance, phase point is thrown out of the origin along the plane into a new equilibrium point. Time response of the state coordinates for this case is given in Fig. 3, by solid lines. Data analysis has confirmed the analytically predicted steady state and its norm.

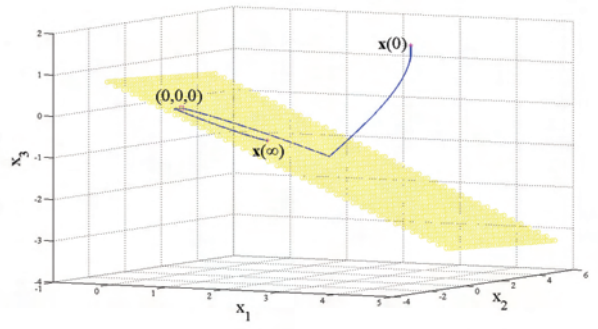


Fig. 2. System trajectory and the sliding plane

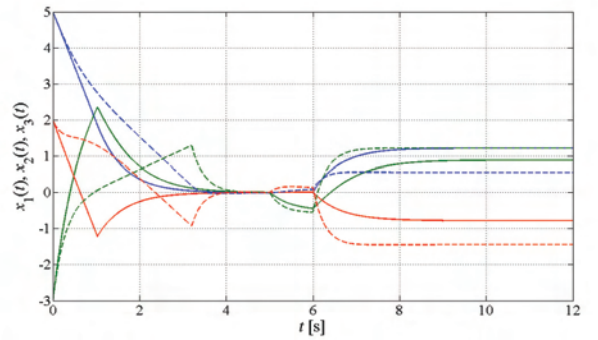


Fig. 3. State coordinates (solid lines for s_p=1.5, dashed lines for s_p=3).

In order to prove that the obtained norm is minimal, some other slope of the sliding plane is selected. Namely, the chosen new sliding plane ensures faster SM dynamics, given by $s_p=3$, which results in $\mathbf{c}=\begin{bmatrix}-3 & -1 & -2\end{bmatrix}$. Recalculated steady state and its norm are respectively $\mathbf{x}(\infty)=\begin{bmatrix}0.55 & 1.22 & -1.44\end{bmatrix}^{\mathrm{T}}$ and $\|\mathbf{x}(\infty)\|=1.97$, which are confirmed by the state coordinate response in Fig. 3, denoted by the dashed lines. The obtained norm is larger then the previous one. Unexpectedly, the faster SM dynamics of prescribed by the new plane, which better rejects matched disturbances, produces worse performance under action of the unmatched disturbances.

V. Conclusion

The paper studies the influence of unmatched disturbances upon SM motion and offers a sliding hyperplane design method that minimizes the system sensitivity against such disturbances, by minimization of the steady state vector norm. It is demonstrated that for a class of linear systems this optimization task can be explicitly solved by renouncing of a certain number of degrees of freedom in SM eigenvalues allocation.

The developed systematic sliding hyperplane design procedure has been demonstrated on an illustrative numerical example. The simulation results confirm the analytically predicted behavior, and the calculated minimum of the steady state norm is achieved. It is also shown that certain SM features in systems satisfying the matching conditions are not necessarily present when unmatched disturbances arise.

References

[1] V. I. Utkin, *Sliding Modes in Control and Optimization*. Berlin, Springer-Verlag, 1992.

[2] B. Draženović, "The invariance conditions in variable structure systems," *Automatica*, no. 5, pp. 287-295, 1969.

[3] S. K. Spurgeon and R. Davies, "A nonlinear control strategy for robust sliding mode performance in the presence of unmatched uncertainty," *Int. J. Contr.*, vol. 57, no. 5, pp. 1107-1123, 1993.

[4] C.-M. Kwan, "Sliding mode control of linear systems with mismatched uncertainties," *Automatica*, vol. 31, no. 2, pp. 303-307, 1995.

[5] S. H. Jang and S. W. Kim, "A new sliding surface design method of linear systems with mismatched uncertainties," *IEICE Trans. Fundamentals*, vol. E-88A, no. (1), pp. 387-391, 2005.

[6] C. C. Wen and C.-C. Cheng, "Design of sliding surface for mismatched uncertain systems to achieve asymptotical stability," *J. Franklin Institute*, no. 345, pp. 926-941, 2008.

[7] H. H. Choi, "An LMI-based switching surface design method for a class of mismatched uncertain systems," *IEEE Trans. Autom. Control*, vol. 48, no. 9, pp. 1634-1638, 2003.

[8] F. Castanos and L. Fridman, "Analysis and design of integral sliding manifolds for systems with unmatched perturbations," *IEEE Trans. Autom. Control*, vol. 51, no. 5, pp. 853-858, 2006.

[9] A. Polyakov and A. Poznyak, "Minimization of the unmatched disturbances in the sliding mode control systems via invariant ellipsoid method," *in Proc. IEEE 18th Int. Conference on Control Applications*, Saint Petersburg, Russia, 2009, pp. 1122-1127.

Proportional Motorized Valve Control Without Real Position Feedback

Mile I. Petkovski[1] and Momcilo G. Spasovski[2]

***Abstract* – In this paper a physical model and implementation device to control of proportional motorized valve is considered. Driving circuit consist of two digital outputs for opening and closing. The control circuitry is performed with PID loop and without the real valve position feedback.**

***Keywords* – PID Control, Programmable Logic Controller.**

I. INTRODUCTION TO PID CONTROL

The PID controller is the most common form of feedback. It was an essential element of early governors and it became the standard tool when process control emerged in the 1940s. In process control today, more than 95% of the control loops are of PID type, most loops are actually PI control. PID controllers are today found in all areas where control is used. The controllers come in many different forms. There are standalone systems in boxes for one or a few loops, which are manufactured by the hundred thousands yearly. PID control is an important ingredient of a distributed control system. The controllers are also embedded in many special purpose control systems. PID control is often combined with logic [1] [2], sequential functions, selectors, and simple function blocks to build the complicated automation systems used for energy production, transportation, and manufacturing. Many sophisticated control strategies, such as model predictive control, are also organized hierarchically. PID control is used at the lowest level; the multivariable controller gives the set points to the controllers at the lower level. The PID controller is an important component in every control engineer's tool box.

PID controllers have survived many changes in technology, from mechanics and pneumatics to microcontrollers and programmable logic controllers. The programmable logic controller has had a dramatic influence on the PID controller. This has given opportunities to provide additional features like automatic tuning [3], gain scheduling, and continuous adaptation.

We will start by summarizing the key features of the PID controller. The "textbook" version of the PID algorithm is described by the next expression:

$$u(t) = K\left(e(t) + \frac{1}{T_i}\int_0^t e(\tau)\,d\tau + T_d\frac{de(t)}{dt} \right) \quad (1.1)$$

Where u is the control signal and e is the control error.

The reference variable is often called the set point. The control signal is thus a sum of three terms: the P-term which is proportional to the error e, the I-term which is proportional to the integral of the error e, and the D-term which is proportional to the derivative of the error e. The controller parameters are proportional gain *K*, integral time *Ti*, and derivative time *Td*. The integral, proportional and derivative part can be interpreted as control actions based on the past, the present and the future. The derivative part can also be interpreted as prediction by linear extrapolation.

In the case of proportional control where integral and derivative parts are equal to zero, a steady state error appeared. The error will decrease with increasing gain, but the tendency towards oscillation will also increase. The steady state error disappears when integral action is used and the strength of integral action increases with decreasing integral time T_i.

The parameters *K* and T_i are chosen so that the closed loop system is oscillatory. Damping increases with increasing derivative time, but decreases again when derivative time becomes too large. Recall that derivative action can be interpreted as providing prediction by linear extrapolation over the time T_d. Using this interpretation it is easy to understand that derivative action does not help if the prediction time *Td* is too large. Also notice that the period of oscillation increases when derivative time is increased.

To obtain a good PID controller it is also necessary to consider

- Noise filtering
- Set point weighting
- Windup
- Tuning
- Computer implementation

In the case of the PID controller these issues emerged organically as the technology developed but they are actually important in the implementation of all controllers. Many of these questions are closely related to fundamental properties of feedback.

The control system presented in this paper is based on implementation of PID control with programmable logic controller for governing of motorized valve but without feedback of real valve position.

The paper is organized in XX section. Section II describe the controlled system. Control circuitry is depicted in section III. Next two section consider the control algorithm and hardware implementation of the control system.

[1]Mile I. Petkovski is with the Faculty of Technical Sciences, I.L.Ribar bb, 7000 Bitola, Macedonia, E-mail: mile.petkovski@uklo.edu.mk

[2]Momcilo Spasovski is with Teri Engineering, 11th October 83/3-2, 1000 Skopje, Macedonia, E-mail: momcilo.spasovski@terieng.com.mk

II. CONTROLLED SYSTEM DESCRIPTION

At the following picture the controlled system configuration is depicted.

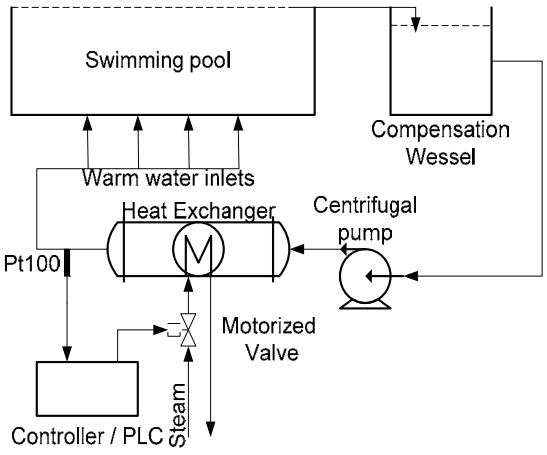


Fig. 1. Controlled system configuration

The control object is the temperature of the water in the swimming pool.

The water temperature has to be controlled around 29 Celsius degrees, with tolerance of 1 degree. This means that the minimum temperature is 28 degrees and maximum temperature should not exceed 30 degrees. The volume of the swimming pool is $60m^3$.

Compensation Wessel is intended to collect the amount of water which exceeds the pool volume. Therefore the filling of swimming pool is performed through this Wessel. Water temperature inside is around the temperature of the pool.

Water is transported from the compensation Wessel by the centrifugal pump through heat exchanger and returned back in the swimming pool through inlets located at the bottom floor. Filtration system is not showed doe to its unimportance for our investigation. The pump flow capacity is dimensioned according to pool volume, to exchange complete volume for period of one hour.

To control of water heating a motorized valve is applied to the system. As mentioned at the title this motorized valve is not equipped with position sensor, and control system has no information of its real position. Feedback is performed by the Pt 100 based transducer.

Control system is based on the programmable logic controller to improve process control flexibility.

III. CONTROL SYSTEM CONFIGURATION

Programmable logic controller equipped with the PID loop control hardware, combined with the logic function are suitable device for building efficient and flexible control hardware.

The principal schematic diagram of control system based on programmable logic controller is presented on Fig. 2.

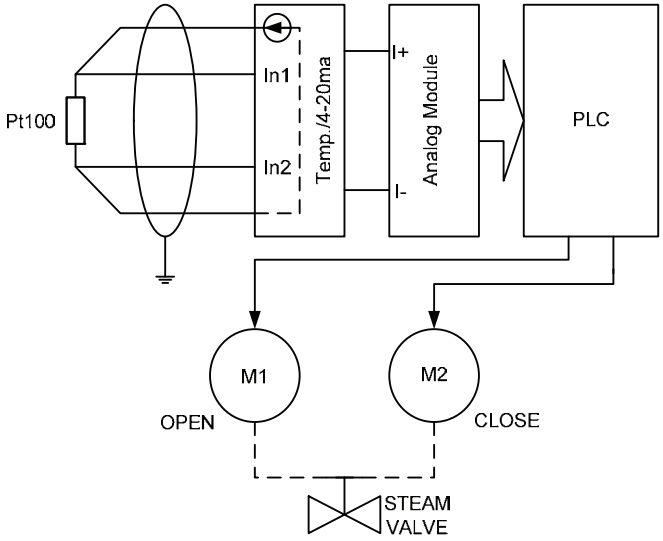


Fig. 2. Control circuitry

Measured temperature by the Pt 100 transducer is converted to analog current standard signal 4-20mA, which allows signal transmission to longer distance. Analog signal is converted in analog module in 12 bit word and sent to controller. Programmable logic controller govern the valve by the two single phase synchronous motor, one for opening and one for valve closing.

It is obvious in fig. 2 that there is not a position feedback between the motorized valve and PLC.

IV. CONTROL ALGORITHM

Programmable logic controller applied here can handle up to 14 PID loops, but in this case only one is required. Standard PID loop utilize three variables:

- PV – Process value
- SP – Set Point
- CV – Control value, the PID output

Because no position feedback between controlled motorized valve and controller, additional variable is involved so-called VVP – Virtual valve position. This variable is purposed to keep the estimation of valve position which is calculated during each valve movement. VVP variable increase there value in the case of valve opening and decrease otherwise. In the inaction case the value is not changed. The amount of VVP changes depend of the control pulse width generated by the PLC to opening/closing valve.

Next figure present the block diagram of implemented control algorithm.

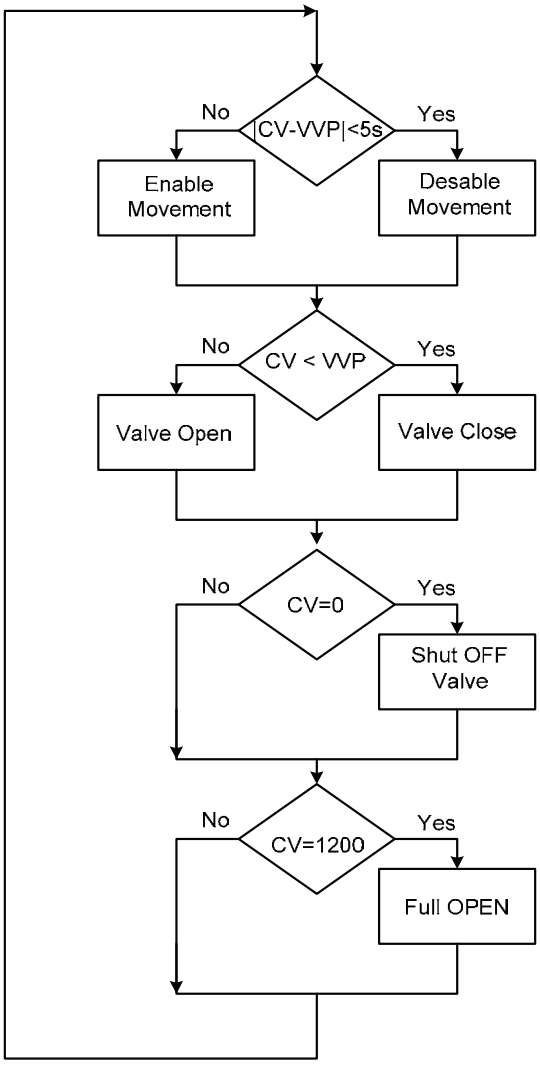


Fig. 3. Control algorithm block diagram

The elapsed time from closed to full open valve is 120 seconds. Sampling period is chosen to be 100 ms. which lead to boundary of the control value in the interval of $[0,1200]$.

VVP, virtual valve position is set to zero during PLC initialization procedure keeping the valve completely closed for the time interval a few seconds longer then interval sufficient to shut off the valve from full opened position.

To avoid oscillation around the control value point, a small hysteresis is involved, which disable movement shorter than 5 seconds.

Control value, CV, is calculated by the PID function loop defined in the PLC structure according to process value, PV and set point, SP, as input parameters, and PID loop parameters, Kp, proportional action constant, Ti, integral time and Td, differential time, defined during PLC programming phase. Programmable logic controller offers a very suitable manner of PID loop parameters adjustments, showed at the next figure:

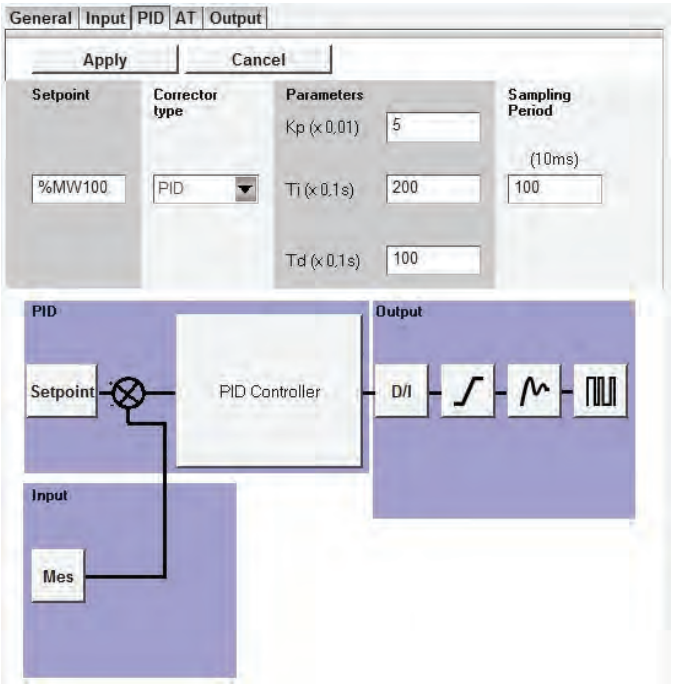


Fig. 4. PID loop configuration windows

The movement action as the output commands are defined in the ladder diagram. Control value obtained as the output result of the PID loop is compared with the VVP, virtual valve position, whish represents the estimation of real valve position. During valve movement, opening or closing VVP is increased, or decreased by 1 every 0.1 seconds. In the initialization phase or in the case of CV=0 the valve is completely closed and VVP set to zero. In the case of CV=1200, which means fool open valve, the valve will be completely opened and VVP is set to 1200.

V. HARDWARE IMPLEMENTATION

Implementation hardware for realization of the control system described above consists of one PLC; model TWIDO TWDLMDA20DRT, one analog module with 12 bit resolution TM2AMI2HT and signal converter Pt100 to 4-20mA. All modules are produces by Schneider Electric, Telemecanique.

TWIDO PLC is equipped wit 20 inputs/output, which are 12 digital inputs and 8 digital outputs, 6 relay outputs and 2 transistors fast outputs. To realize PID function of analog signals an analog module is connected. Analog module consists of two analog inputs configured as 4-20mA range current inputs. Analog signal corresponding to temperature is obtained by the Pt100 transducer and resistance value is converted to current by another converter.

The valve power supply is 230V alternating current and the governing is realized by the first two PLC's digital relay outputs. Valve is controlled by the two synchronous low power electric motors and high ratio gearbox.

Control module is presented at the fig.5.

Fig.5. Temperature control module

To control of opening and closing valve a two relay outputs are used. PWM, pulse width modulated control applied is not with the constant period. This period depend on difference between control value, CV, and virtual valve position, VVP, defined in the algorithm below.

VI. Conclusion

In this paper an algorithm and implementation hardware for temperature control of the swimming pool is presented. The absence of physical feedback of motorized valve position to control system is overrun introducing additional variable VVP – virtual valve position which is estimation of real valve position and is updated during every valve movement.

Applying the programmable logic controller in the control system hardware contribute to its additional flexibility.

PLC controller TWIDO is equipped with the possibilities for auto tuning of PID loop parameters which application is expected in the future implementations.

References

[1] Dinko Vukadinović, Mateo Bašić, “Temperature Control in a Closed Space Environment by using a PLC”, INFOTEH-JAHORINA Vol. 8, Ref. A-4, p. 15-19, March 2009

[2] Mohammad A.K.Alia, “Using PLC for Custom-design of a PID/PWM Program to Control a Heater Temperature”, American Journal of Applied Sciences 4 (5): 307-316, 2007.

[3] Twido Suite V2.0, Programming software, 09/2007.

[4] Twido programmable controllers, Software Reference Guide, TWD USE 10AE eng Version 3.2, 2006.

Comparative analysis on Stock Prices Index Prediction

Cvetko J. Andreski[1]

Abstract – **The identification and forecasting values of market index is challenge for many researchers. The main problem in identification of stock prices and indexes is human behaviour in trade markets and uncertainty of it. In this paper we are dealing with two techniques for identification and forecasting of stock market index: time series analysis and fuzzy time series. We made comparative analysis of these techniques, especially for the forecasting feature.**

Keywords – **Identification, Forecast, Analysis.**

I. INTRODUCTION

Many companies (most of them in financial branch) offer interest to their customers. They need to invest their assets in order to cover the expenses. Insurance companies calculate interest for the insured asset. The stock market is one of the most indispensable sources for companies to raise money. The stock prices are always predicted and studied by each day individual investors, stock fund managers and financial analysts. The forecasting accuracy is main target, because more profit will be achieved if more accurate predictions are given [9].

Recently, fuzzy time series has been widely explored for forecasting data of dynamic and non-linear data. Its application varies from forecasting of university student enrollment [1, 2, 4-7] to forecasting stock index [8-14, 20-21], temperature [15] and financial forecasting [16]. This approach is an implementation fuzzy logic in the time series analysis. Many different methods and models have been proposed for stock index forecasting by researcher into fuzzy time series. Huarng[14] initiated a study on heuristic models of fuzzy time series for forecasting by using stock index data. In addition, Huarng *et.al* [21] continued to analyze a multivariate heuristic model for fuzzy time series forecasting. On the other hand, Yu [8] enhanced of weighted fuzzy time series models for Taiwan Stock Index (TAIEX) forecasting in 2005. It is assigned by the recurrent fuzzy logical relationships (FLRs) in fuzzy logical group (FLG). In establishing fuzzy relationship and forecasting are important step to consider the weight. Furthermore, Cheng *et al* [9, 20] presented the trend-weighted fuzzy time series model for forecasting in 2006 and 2009. In 2010, Chen and Wang [21] also presented the optimal weights of the antecedent fuzzy sets in each fuzzy-trend logical relationship groups. Currently, the advancement of fuzzy time series is still going on.

[1]Cvetko J. Andreeski is with the Faculty of Tourism and hospitality, kej M. Tito 95, 6000 Ohrid, Macedonia, E-mail: cvetko.andreeski@uklo.edu.mk

In this paper we proceed with time series analysis using ARIMA model for analysis. Starting with unit root test and autoregressive and partial autoregressive analysis we made a model for analysis of Macedonian MBI10 index of share trade. We also made analysis on statistics of the model. At the end we made forecast of future values with analysis of validity.

Next part is basics of fuzzy time series and definitions about fuzzy time series. At the end we made identification of the index and forecast with fuzzy time series model.

II. BOX-JENKINS ARIMA METHODOLOGY FOR TIME SERIES

The ARIMA methodology represents a system-theoretic approach for modelling time series, since the specification of the empirical model does not emerge from the findings and concepts provided by economic theory. Instead, the time series is modelled by a combination of two components - the autoregressive and the moving average term i.e. the current value of a variable is represented as a function of lagged values of that variable and past random shocks that hit the variable.

In this part of the paper, we will illustrate the implementation of this methodology by using it to model the dynamics of inflation, as measured by the changes in the Macedonian index of share trade MBI10. Again, we work with daily data and the sample covers the period starting in 18.03.2009 and ending with 06.06.2009. As can be seen, in modelling the index, we do not use the whole series that is available, but we the sample is limited with 01.06.2009. Of course, we are aware that the restricted sample appears to be a sort of handicap in modelling the series, since it reduces the reliability of the estimates as well as the power of some tests. Yet, although we work with a small sample, its size exceeds the minimum number of 50 observations that is usually recommended by the literature on ARIMA modelling. For the property of stationarity we need to make two differentiations to get stationary data series. We made additional ADF test to check the stationarity and to have starting position for the model we should choose.

TABLE I
UNIT ROOT TEST

ADF Test Statistic	-8.288994	1%	Critical Value*	-2.6072
		5%	Critical Value	-1.9470
		10%	Critical Value	-1.6191

*MacKinnon critical values for rejection of hypothesis of a unit root.
Augmented Dickey-Fuller Test Equation
Dependent Variable: D(INDEKS,2)
Method: Least Squares
Date: 06/21/09 Time: 20:17

Sample(adjusted): 3/23/2009 6/02/2009
Included observations: 52 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(INDEKS(-1))	-1.702140	0.205349	-8.288994	0.0000
D(INDEKS(-1),2)	0.358848	0.130303	2.753938	0.0082

R-squared	0.678665	Mean dependent var	-0.957885
Adjusted R-squared	0.672238	S.D. dependent var	54.27995
S.E. of regression	31.07554	Akaike info criterion	9.748421
Sum squared resid	48284.45	Schwarz criterion	9.823469
Log likelihood	-251.4590	F-statistic	105.6008
Durbin-Watson stat	2.211257	Prob(F-statistic)	0.000000

First, although preliminary, information on the possible model can be extracted from the simple visual inspection of the series (See one of the graphs). As can be seen, the graphs clearly show that the series exhibits regular peaks, which over the time appear every fifth day. Thus, this is reflecting the seasonal component in the movement of the inflation rate. Also, it is obvious that a series of upward movements are followed by a series of downward movements and, this suggests the presence of a moving average process. Yet, notwithstanding this obvious information, obtained by the visual inspection of the graph, the most important tool in the stage of model identification is the analysis of the sample autocorrelation (*acf*) and partial autocorrelation (*pacf*) functions along with the accompanying test-statistic (See Table 2). The *acf* cannot determine the precise form of the model, but anyway, it offers a hint that the series can be modelled by a mixed process, which should include both AR and MA components. We are able to gain more information on the possible model, looking at the *pacf*,.

TABLE II

AUTOCORRELATION AND PARTIAL AUTOCORRELATION TEST

	AC	PAC	Q-Stat	Prob
1	-0.485	-0.485	13.172	0.000
2	-0.088	-0.423	13.617	0.001
3	0.019	-0.393	13.638	0.003
4	0.137	-0.186	14.750	0.005
5	-0.248	-0.476	18.479	0.002
6	0.316	-0.142	24.671	0.000
7	-0.146	-0.144	26.024	0.000
8	0.023	0.034	26.060	0.001
9	-0.112	-0.030	26.887	0.001
10	0.105	-0.080	27.640	0.002
11	-0.031	-0.018	27.706	0.004
12	0.059	-0.046	27.956	0.006
13	-0.003	0.150	27.957	0.009
14	-0.093	-0.006	28.602	0.012
15	-0.030	-0.090	28.669	0.018
16	0.132	-0.052	30.047	0.018
17	-0.010	-0.002	30.055	0.026
18	-0.048	0.051	30.250	0.035
19	0.012	0.030	30.262	0.049
20	-0.031	0.041	30.345	0.064

The estimated model along with the accompanying diagnosis tests are given in Table 3. As confirmed by the value of the tests, in all three models, the residuals behave reasonably well i.e. they are distributed normally and do not suffer from serial correlation.

TABLE II

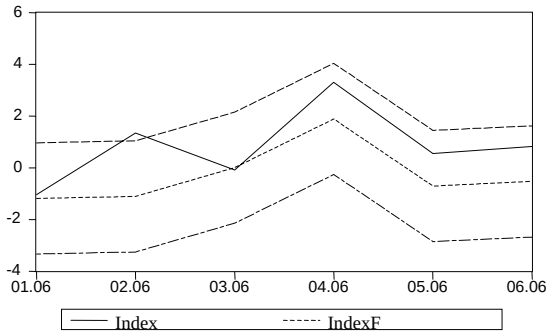
Autocorrelation and partial autocorrelation test

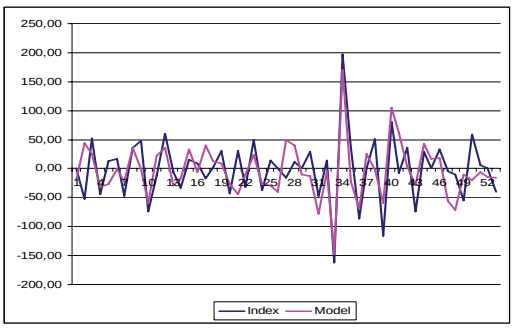
Dependent Variable: INDEKS
Method: Least Squares
Date: 06/21/09 Time: 21:31
Sample(adjusted): 3/18/2009 5/31/2009
Included observations: 53 after adjusting endpoints
Convergence achieved after 37 iterations
Backcast: 3/11/2009 3/17/2009

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(1)	-1.219381	0.065879	-18.50930	0.0000
MA(5)	0.228321	0.060001	3.805304	0.0004

R-squared	0.674157	Mean dependent var	-0.934151
Adjusted R-squared	0.667768	S.D. dependent var	53.75577
S.E. of regression	30.98462	Akaike info criterion	9.741864
Sum squared resid	48962.37	Schwarz criterion	9.816215
Log likelihood	-256.1594	F-statistic	105.5169
Durbin-Watson stat	2.548076	Prob(F-statistic)	0.000000

Graph I





We've used the model to generate out-of-sample forecasts for 01-06.06. As shown by the right panel of Graph 1 through Graph 3, from the point of view of the model's ability for forecasting future inflation, then the first model performs best. Namely, for this model, the forecasted inflation moves relatively close to the actual inflation that lies always within the ± two standard errors confidence interval.

III. DEFINITIONS FOR FUZZY TIME SERIES

There are some basic theories have been defined into fuzzy time series as follows:

Definition 1: Let, Y(t) (t=0,1,2...,) a subset of real numbers, be the universe of discourse on which fuzzy sets ,... $f_i(t)$ (i=1,2,…) are defined and $F(t)$ is a collection of $f_1(t), f_2(t), \ldots$ then $F(t)$ is called a fuzzy time series defined on $Y(t)$ (t = …,0, 1, 2, …). [1]

Definition 2: Suppose $F(t)$ is caused by denoted by $F(t-1) \rightarrow F(t)$, then this relationship can be represented by:

$$F(t) = F(t-1) \cdot R(t, t-1)$$

where $R(t, t-1)$ is a fuzzy relationship between $F(t)$ and and is called the first-order model of $F(t)$ [1].)1

Definition 3: Assumed that $F(t)$ is a fuzzy time series, and is a first-order model of $F(t)$. If $R(t, t-1) = R(t-1, t-2)$ for any time t, the $F(t)$ is called as the time-invariant fuzzy time series. If $R(t, t-1)$ is dependent on time t, that is $R(t, t-1)$ may be different from $R(t-1, t-2)$ for any t, then $F(t)$ is called the time-variant fuzzy time series.

Definition 4: Let U be the universe of discourse, where $U = \{ u_1, u_2, \ldots, u_b \}$. A fuzzy set A_i of U defined as $A_i = f_{Ai}(u_1)/u_1 + f_{Ai}(u_2)/u_2 + \ldots + f_{Ai}(u_b)/u_b$, where f_{Ai} is the membership function of the fuzzy set A_i; $f_{Ai} : U \rightarrow [0, 1]$. Hence, $f_{Ai}(u_1)/u_1$ is the degree of belongingness of u_a to A_i; $f_{Ai}(u_a) \rightarrow [0, 1]$ and $1 \leq a \leq b$ [11].

Definition 5: If there exist a fuzzy relationship), $R(t-1, t)$ such that $F(t) = F(t-1) \times R(t-1, t)$, where x represent an operation, then $F(t)$ is said to be caused by F(t-1). [11]

Definition 6: Let $F(t-1) = A_i$ and $F(t) = A_j$. The relationship between two consecutive data (called a fuzzy logical relationship, FLR), i.e., $F(t)$ and $F(t-1)$, can be denoted by $A_i \rightarrow A_j$, $i, j = 1, 2, \ldots, p$ (where p is interval or subinterval number) is called the left-hand side (LHS), and A_j is the right-hand side (RHS) of the FLR. [11]

IV. LEFT AND RIGHT (LAR) METHOD AND DETERMINING OF WEIGHT

Our reviews on previous studies in related area show that the description of recurrence of fuzzy logical relationship in the fuzzy logical group (FLG) is still main issue. Yu[8], Cheng *et al.* [9, 20] and Chen *et al.* [21] suggested that the weight for fuzzy logical relationships should be considered to improve the forecasting accuracy. In addition, determining of weights was assigned based on domain know-how or their chronological order [8].

In this paper, we do not consider the recurrence of fuzzy relation to obtain the weight. Based on our simulation study is that the number of occurrence of the same FLR can be found more then 3 or 4 times for each FLG. To optimize this condition, we propose left and right method as follows

Let $A_3 \rightarrow A_5, A_3, A_6, A_6, A_3, A_2, A_3, A_5, A_2, A_4$ be an FLG, then the first value (A_2) that appears to left of the original value (A_3), then first value (A_4) that appears to left of the original value (A_3)and has the same value as original. Thus, this FLG can be written simply as

$$A_3 \rightarrow A_3, A_2, A_4 \quad (1)$$

LAR method is generalized by using naïve assumption which the occurrence of present value A_t is very closed to the past value $A_{(t-1)}$.

B.

Computation to Derive the Weight

Suppose be an FLG which $i = j$, $i, j \geq 2$ and $i, j \in Z$. By using a collection of variations of chronological numbers (j-1, j, j+1) in FLG, then computational weight can be assigned as follows: $A_{11}, \ldots A_{jji} A_{jji}$

$\mathbf{W}(t) = [\, w_1 \, w_2 \, w_3 \,]$

$$= \left[\frac{(j-1)}{(j-1)+j+(j+1)} \; \frac{j}{(j-1)+j+(j+1)} \; \frac{j+1}{(j-1)+j+(j+1)} \right]$$

Let j-1 = c_1, j = c_2, and j+1 = c_3, thus

$$= \left[\frac{c_1}{(c_1+c_2+c_3)} \; \frac{c_2}{(c_1+c_2+c_3)} \; \frac{c_3}{(c_1+c_2+c_3)} \right]$$

$$\mathbf{W}(t) = \left[\frac{c_1}{\sum_{h=1}^{3} c_h} \; \frac{c_2}{\sum_{h=1}^{3} c_h} \; \frac{c_3}{\sum_{h=1}^{3} c_h} \right]$$

which $\mathbf{W}(t)$ has satisfied the condition whereby the calculations will equal 1.

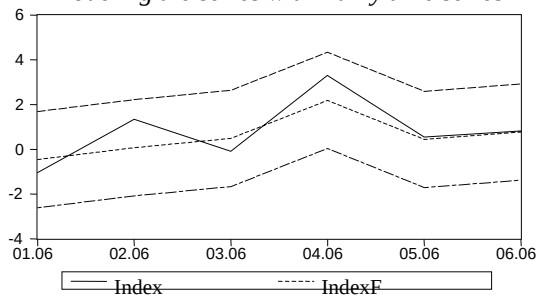
Example 1.

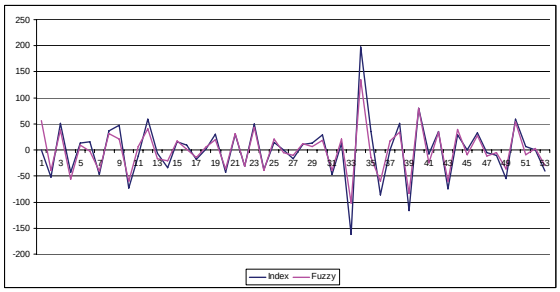
By using (1), the weight can be determined as follows:
Given $A_3 \rightarrow A_3, A_2, A_4$ and $c_1 = 3$, $c_2 = 2$, $c_3 = 4$
then
$w_1 = 3/(3 + 2 + 4)$, $w_2 = 2/(3 + 2 + 4)$, $w_3 = 4/ (3 + 2 + 4)$
thus
$\mathbf{W}(A_3) = \mathbf{W}(t) = [\, w_1 \, w_2 \, w_3 \,] = [\, 0.33 \; 0.22 \; 0.45 \,]$.
Given $A_8 \rightarrow A_9, A_8, A_7$ and $c_1 = 9$, $c_2 = 8$, $c_3 = 7$
then
$w_1 = 9/(9 + 8 + 7)$, $w_2 = 8/(9 + 8 + 7)$, $w_3 = 7/ (9 + 8 + 7)$
thus
$\mathbf{W}(A_8) = \mathbf{W}(t) = [\, w_1 \, w_2 \, w_3 \,] = [\, 0.37 \; 0.33 \; 0.30]$.

Therefore, weight elements do satisfy the condition in both the weight matrices.

Graph II.

Modeling the series with Fuzzy time series





As for ARIMA model we first make the differencing of the time series to get the stationary time series. We made the intervals of the series and each interval is transformed to be a linguistic value time series based on previous equations. We made the fuzzy logical relationships and groups for linguistic value time series. At the end we made forecasting for post processing data based on the linguistic values of previous series. Results are shown in Graph 2.

V. Conclusion

We made identification and forecast on MBI10 index using time series analysis and fuzzy time series analysis. These two techniques have common background but different approach on identification and forecasting. Time series analysis are traditional technique for identification and forecasting and proven as trustable one. The fuzzy time series are still in research and innovative model. On the other hand identification and forecast of stock index is not an easy task to forecast the index prices precisely because many affected factor influence the price of stock and index. We can see that we can get better results in forecasting if we use appropriate fuzzy model.

References

[1]. Q. Song and B. S. Chissom, "Forecasting enrollments with fuzzy time series" – Part I, *Fuzzy Sets and Systems*, **54**(1993) pp. 1-9.
[2]. Q. Song and B. S. Chissom, "Forecasting enrollments with fuzzy time series" – Part II, *Fuzzy Sets and* Systems, **64**(1994) pp. 1-8.
[3]. Q. Song and B. S. Chissom, "Fuzzy time series and its models", *Fuzzy Sets and Systems* **54**(1993) pp. 269-277.
[4]. S. M. Chen, "Forecasting enrollments based on fuzzy time series". *Fuzzy Sets and Systems*, **81**(1996) pp. 311-319.
[5]. S. M. Chen and C. C. Hsu, "A new method to forecast enrollments using fuzzy time series", *International Journal of Applied Science and Engineering*, **3**(2004) pp. 234-244.
[6]. M. Sah and Y. D. Konstantin, "Forecasting enrollment model based on first-order fuzzy time series", *Proceeding of World Academy of Science, Eng and Tech*, **1**(2005) pp. 375-378.
[7]. I. H. Kuo, S. J. Horng, T. W. Kao, T. L. Lin, C. L. Lee and Y. Pan, "An improved method for forecasting enrollments based on fuzzy time series and particle swarm optimization", *Expert Systems with Applications*, **36**(2009) pp. 6108-6117.
[8]. H. K. Yu, "Weighted fuzzy time series models for TAIEX forecasting", *Physica A* **349**(2005) pp. 609-624.
[9]. C. H. Cheng, T. L. Chen, and C. H. Chiang, "Trend-weighted fuzzy time series model for TAIEX forecasting", Proceeding in *ICONIP* (2006), Part III, LNNC 4234, pp. 469-477.
[10]. K. Huarng, "Heuristic models of fuzzy time series for forecasting", *Fuzzy Sets and Systems*, **123** (2001) pp. 369-386.
[11]. K. Huarng, H. K. Tiffany Yu, and W. S. Yu, "A multivariate heuristic model for fuzzy time series forecasting", *IEEE Transactions on Systems, Man, and Cybernetics*, **37**(2007) pp. 263-275.
[12]. T. A Jilani and S. M. A. Burney, "A refined fuzzy time series model for stock market forecasting", *Physica A*, **387**(2008) pp. 2857-2862.
[13]. T. H. K. Yu and K. H. Huarng, "A bivariate fuzzy time series model to forecast the TAIEX", *Expert Systems with Application*, **34**(2008) pp. 2945-2952.
[14]. H. H Chu, T. L. Chen, C. H. Cheng, C. C. Huang, "Fuzzy dual-factor time series for stock index forecasting", *Expert Systems with Applications*, **36**(2009) pp. 165-171.
[15]. S. M. Chen, "Temperature prediction using fuzzy time series". *IEEE Transactions on Systems, Man, and Cybernetics*, **30**(2000) pp. 263-275.
[16]. C. H. L. Lee, A. Liu, and W. S. Chen, "Pattern Discovery of Fuzzy time series for financial prediction", *IEEE Transactions on Knowledge and data Engineering*, **18**(2006) pp. 613-625.
[17]. J.E. Hanke, and D.W. Wichern, "Business forecasting", *Prentice Hall*, 2009.
[18]. H. T. Nguyen and B. Wu, "Chapter; Fuzzy time series analysis and forecasting", *Fundamental of Statistics with Fuzzy Data*, StudFuzz, **198**(2006), pp. 145-182.
[19]. S-T. Li and Y-C. Cheng, "An enhanced deterministic fuzzy time series forecasting model", *Cybernetics and Systems*, **40**(2009), pp. 211-235.
[20]. C.H. Cheng, T.L. Chen, H.J. Teoh and C.H. Chiang, "Fuzzy time series based on adaptive expectation model for TAIEX forecasting", *Expert Systems with Applications*, **34**(2008), pp. 1126-1132.
[21]. C.H. Chen and N.Y. Wang, "Fuzzy forecasting based on fuzzy trend logical relationship groups", IEEE trans. Systems, Man, and Cybernatics-Part B. doi 10.1109/TSMCB.2009.2038358.
[22]. Cvetko J. Andreeski, Pandian M. Vasant, Mile J. Stankovski, and Georgi M. Dimirovski, ELMAN NN AND TIME SERIES IN FORECASTING MODELS FOR DECISION MAKING, WAC 24-26 July 2006, ISSCI 175 (ISBN 1-889335-26-6), IEEE Catalog Number: 06EX1486

SESSION MRS

Metrology and Remote Sensing

A Study of the Performance of an InGaAs CCD Linear Photodiode Array for Fiber-Optic Grating Sensors

Plamen Balzhiev[1], Rumen Arnaudov[2] and Tinko Eftimov[3]

***Abstract* – The paper covers the experimental set-ups, the measurement algorithms, as well as the processing and analysis of results in order to obtain certain optical characteristics of an InGaAs CCD linear photodiode array such as noise levels, dynamic range, linearity response, tolerances of pixel parameters.**

***Keywords* – Fiber-optic grating sensors, InGaAs photodiode, CCD image array.**

I. INTRODUCTION

Fiber optic grating sensing technology has witnessed a fast growth during the last decade because of its inherent capability to multiplex spectrally a large number of both fiber Bragg gratings (FBGs) and long period gratings (LPGs). This attractive opportunity demands sophisticated sensor interrogation devices [1]. One of the most popular techniques is the use of InGaAs CCD photodiode arrays, sensitive in the near infrared region from 800 nm to 1750 nm where most of the fiber gratings' spectra are situated. Since fiber gratings' transmission minima may be very deep – up to 30 dB, and the grating may also be spectrally very narrow (tens of picometers) it essential to determine the parameters of the CCD array and the whole electronic unit for the proper design of the fiber grating interrogation system. The basic characteristics related to their use in these particular applications and which we study in this paper are noise levels, dynamic range, linearity of response, sensitivity.

The accuracy of the measurement crucially depends on the choice of the operating point of the detection unit. Because of the specific applications where high dynamic range is needed for low level signals at grating resonance wavelength, the catalogue data are not sufficient for the proper design of the interrogation unit.

The paper covers the description of the interrogation measurement system, the experimental set-ups, the measurement algorithms, as well as the processing and analysis of results. The interrogation unit thus developed and tested can be used not only for fiber grating interrogation, but for other spectrally encoded fiber sensors in the 800 nm – 1750 nm range, such as SPR (surface plasmon resonance), polarimetric and interferometric sensors [2] with spectrally resolved detection connected in a larger fiber optic sensing network.

[1]Plamen Balzhiev is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: baljiev@gmail.com

[2]Rumen Arnaudov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: ra@tu-sofia.bg

[3]Tinko Eftimov is with the Faculty of Physics at the Plovdiv University, 24 Tsar Assen Str., Plovdiv 4000, Bulgaria, E-mail: teftimov@uni-plovdiv.bg

II. FIBER GRATING INTERROGATION SYSTEM

A. System description

The optoelectronic measurement system is designed to investigate measure and analyze fiber-optic sensors using spectrally multiplexed fiber gratings. This particular system presented on Fig 1 is employed for the experimental measurement and test of the CCD array. It consists of a broadband amplified spontaneous emission (ASE) light source (Joinwit), that emits in the conventional and long-wavelength ($C + L$) communication band from 1525 nm to 1605 nm, an optical fiber, a fiber-optic attenuator, a collimating lens system, a diffraction grating of 600 lines/mm, the CCD photodiode array under study and the associated electronic unit. The latter transform the incident light into digital data which is further transmitted to a personal computer for processing, analysis and visualization.

The optoelectronic unit is specially designed as a part of grating interrogation system. The block diagram of the measurement device (Fig 2) consists of CCD array – G9204-512D (manufactured by Hamamatsu Photonics). It is a 512

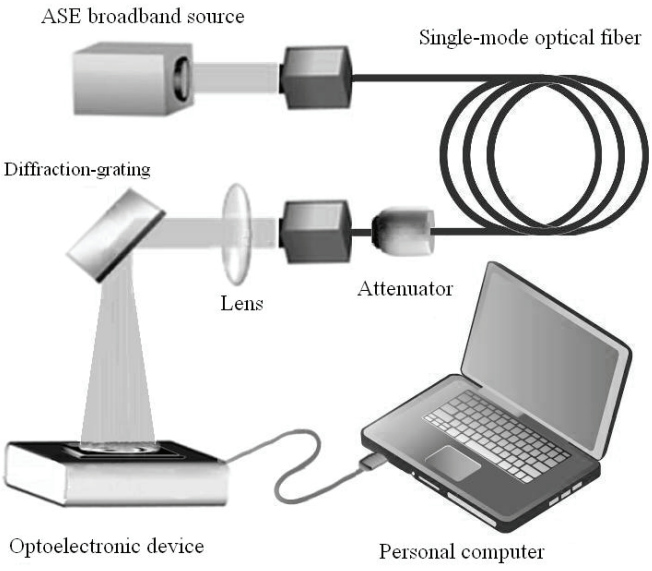


Fig. 1. Diagram of grating interrogation system

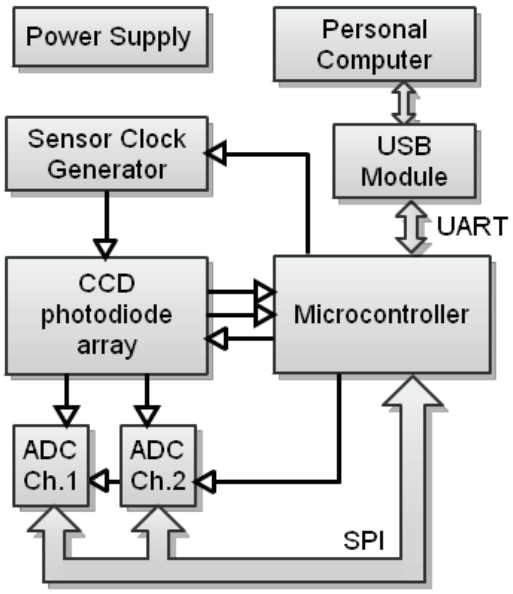


Fig. 2. Block diagram of optoelectronic device

pixel linear InGaAs photodiode array with high sensitivity and stability in 800-1750nm spectral range, low dark current and integrated low noise charge amplifier array comprised of CMOS transistors. Two high-speed capacitive-based SAR analog-digital converters (ADC) transform the analog data from the CCD sensor into 16-bit corresponding digital values. A programmable pulse generator (up to 4MHz) provides the needed clock signals for the proper operation of CCD photodiode array.

The microcontroller in this device performs a set-up of the CCD sensor and ADCs and conducts the measurement procedure [5]. It forms all necessary signals for each individual module to provide accuracy and proper measurement. The obtained data is further transmitted via USB interface to a personal computer. For this purpose serial UART interface to USB converting device is implemented which provides flexibility and high data rates. The designed power supply unit is based on the LDO regulators that produce a low ripple output voltage of 5V and low noise reference voltages of 4,5V and 1,26V for the CCD photodiode array. Some major parameters of the optoelectronic measurement device as part of a grating interrogation system are provided in Table1.

TABLE I
OPTOELECTRONIC DEVICE PARAMETERS

Parameter	Value
Sensor (pixel number)	512
Pixel frequency	500 kHz (up to 4MHz)
Integration time	1ms - 100ms
Pixel Readout time, ms	8,4 ms
ADC Resolution	16 bit
ADC Convertion Rate	500 kSps
ADC Non-linearity, LSB	< ±3
USB transfer speed	500 kbps (up to 2Mbps)
Power consumption(typical @5V)	280mA

B. Measurement algorithm

The optoelectronic device conducts the measurement of all 512 pixels from CCD photodiode array. The operation algorithm presented in Fig. 3 starts with a preliminary initialization of all modules which is followed by configuration and measurement set-up according to previously assigned values from the computer [5]. The parameters that

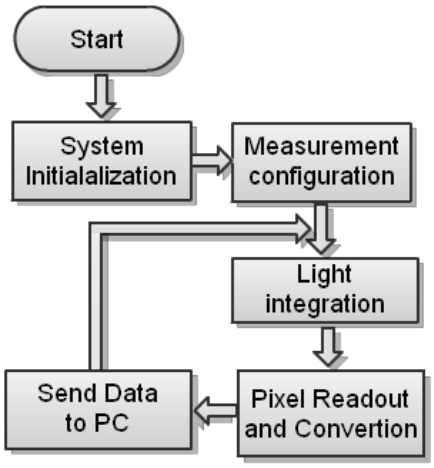


Fig. 3. Block diagram of optoelectronic device

are manually set up are integration time (τ), sensor sensitivity (s), conversion speed and data communication speed (r). The measurement process is performed as the device consecutively reads values from the photodiode array which are converted into 16-bit digital values. After reading all pixels data is further transmitted to a personal computer for processing and visualization.

C. Processing and analysis of results

Data processing algorithms are implemented into the personal computer. A dark current measurement procedure is initially conducted. This data includes the minimal values of measurement process as well as photon and read noise of the CCD sensor, additional sampling noise from ADC and the system noise added from the power supply. This data is necessary to suppress unwanted oscillations. To obtain a reliable and accurate assessment of the dark values an averaging of N-measurements is implemented (1) [4].

$$\hat{S}_d(n) = \frac{1}{N}\sum_{k=0}^{N} S_k(n) \quad (1)$$

The further conducted measurements are presented as relative values to the dark measurement values. In order to avoid high frequency oscillations a low pass filter is applied and the final representative data is obtained from formula (2).

$$\hat{S}(n) = \frac{1}{M}\sum_{m=0}^{M}(S_m(n) - \hat{S}_d(n)) \quad (2)$$

III. PERFORMANCE OF INGAAS CCD LINEAR PHOTODIODE ARRAY

The measurements of the noise levels, the dynamic range and linearity characteristics of the 512 pixel CCD photodiode array were conducted in laboratory conditions with the above described interrogation system. The CCD sensor is tested with the ASE light source and calibrated output power. The

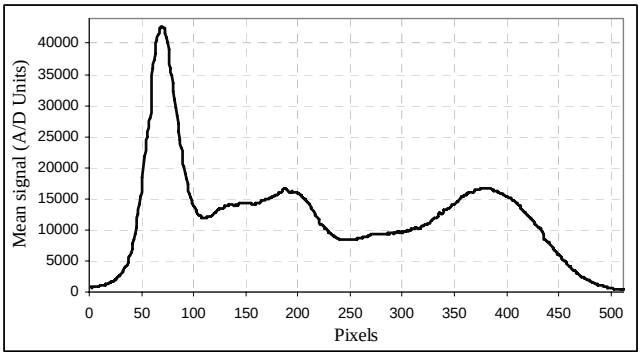


Fig. 4. Measured spectrum of the ASE broad-band (1525-1605 nm) source

minimum total optical power in the 1525-1605 nm range was 9 nW. On Fig.4 the not calibrated spectrum of broadband ASE light source is presented which is measured with the same interrogation system.

A. Noise Levels

The measured noise levels of CCD sensor comprise from many sources – photon noise, dark noise and read noise. Noise levels are calculated as a deviation from the average signal levels (3) [4].

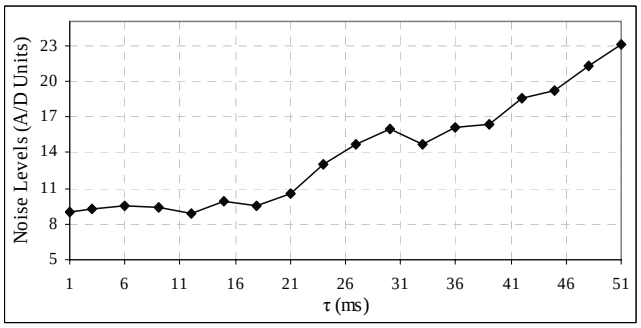


Fig. 5. Noise levels of CCD photodiode array

An important characteristic of noise (NL) is its dependency on integration time - τ (Fig.5). Since read noise is dominant at low light-levels photon noise could be investigated as it increases at longer integration times. On Fig.6 noise levels (NL) of different sensitivities are presented.

$$NL = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(S(n) - S_{mean})^2} \qquad (3)$$

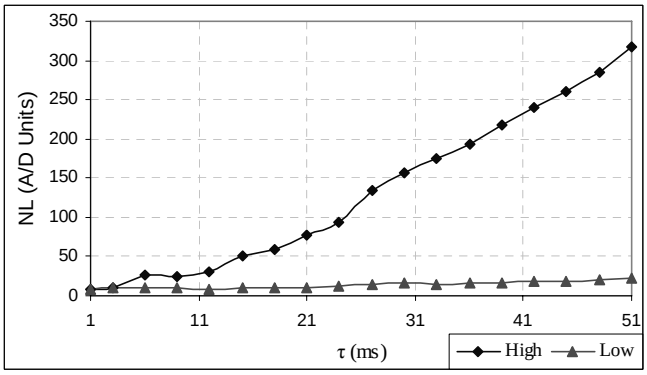


Fig. 6. Noise levels for high and low sensitivity

High sensitivity is achieved with setting different feedback capacitance in the CCD sensor – 0.5pF (high sensitivity) and 10pF (low sensitivity). It is clearly seen that the noise levels increase when larger integration times and a higher sensitivity are applied.

B. Dynamic range

Dynamic range is typically specified as the maximum achievable signal (full well capacity) divided by the CCD sensor noise. It is calculated with formula (4).

$$Dynamic\ Range = 20\lg\left(\frac{S_{sat}(n)}{S_{noise}(n)}\right) \qquad (4)$$

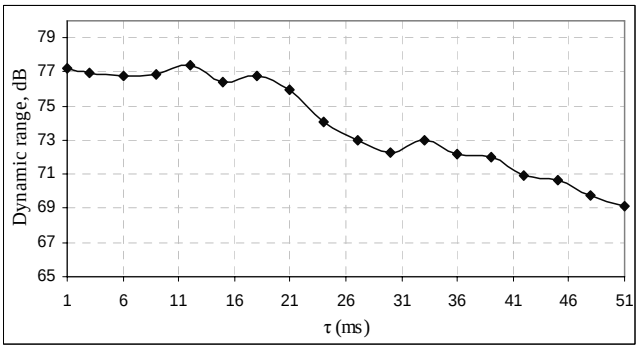


Fig. 7. Dynamic range of the CCD photodiode array

Integration time affects the dynamic range of CCD photodiode array as illustrated on Fig.7. The increase of integration time (τ) produces an increase in dark current and subsequent decrease in dynamic range. At higher sensitivity (lower feedback capacitance) the higher noise levels result also in decrease of dynamic range. The presented graphic on Fig.8 illustrates the characteristics of dynamic range at different sensitivity.

C. Linearity

An important characteristic of fiber-optic interrogation systems is the linearity in response to incident light. Two common techniques are introduced for the measurement and

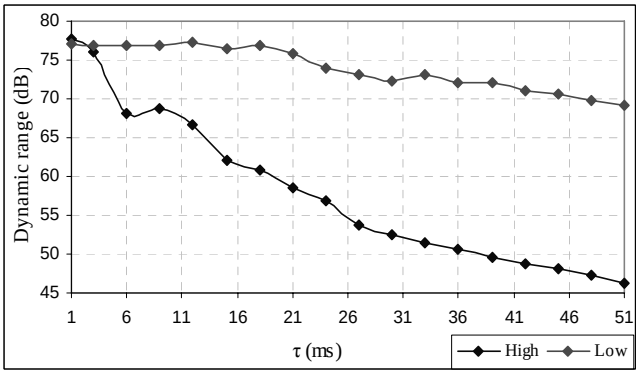


Fig. 8. Dynamic range for two different sensitivities

calculation of the deviation from linearity – measuring output signal as a function of integration time - τ (Fig.9) and measuring output signal as a function of power of incident light - P_i (Fig.10). Nonlinearity is calculated with formula (5).

$$Nonlinearity = \frac{Max.\ Deviation}{Max.\ Signal\ Level} 100, \% \qquad (4)$$

The acquired results are shown in Table II. The relatively higher nonlinearity at high sensitivity results from increased noise levels of measured output signals.

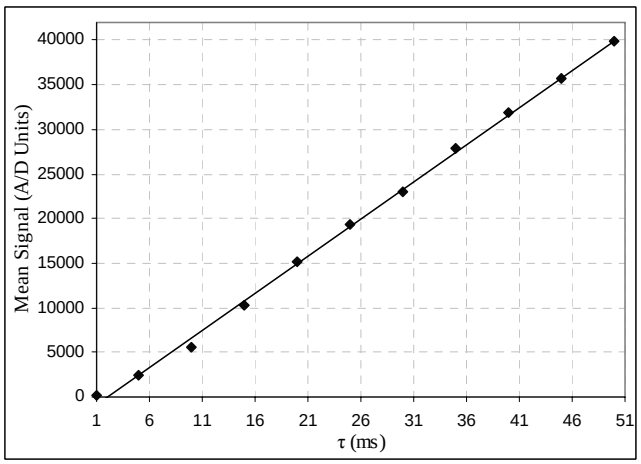


Fig. 9. Linearity as a function of integration time

TABLE II
NONLINEARITY CALCULATION RESULTS

Characteristic	Nonlinearity, %
S(n) = F(Integration time)	
Low sensitivity	1.18
High sensitivity	1.29
S(n) = F(Incident power)	
Low sensitivity	0.83
High sensitivity	1.21

Nonlinear response is usually observed at high illumination intensity. It results in saturation of the pixels and no further change in signal is recorded. Nonlinear response may also result at extremely low illumination levels.

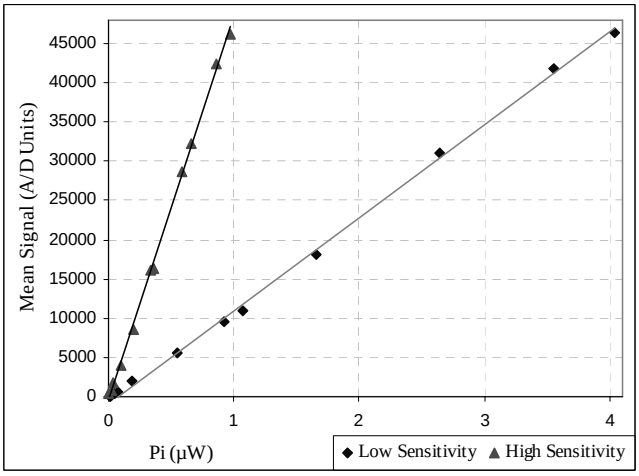


Fig. 10. Linearity as a function of incident power P_i

IV. CONCLUSION

The interrogation system developed and tested here is found to be particularly suitable for spectrally encoded fiber sensors in the 900-1700 nm range [3]. Although the noise levels increase significantly with integration time for the case of high sensitivity, the measured dynamic range even at high sensitivity is better than the typical gratings transmission depth of 30 dB. This is largely due to the far greater increase of the average signal level with integration time. For both high and low sensitivity the sensor array exhibits high linearity in the response.

The investigated CCD linear photodiode array has very good characteristics and is highly appropriate for fiber grating interrogation system. The main advantages of this CCD photodiode array are the increased dynamic range, low noise levels and good linearity of response which are verified with the conducted experimental studies.

REFERENCES

[1] A. D. Kersey, M. A. Davis, H. J. Patrick, M. LeBlanc, K. P. Koo,, C. G. Askins, M. A. Putnam, and E. J. Friebele, Fiber Grating Sensors, *J. Lightwave Techn.*, No. 15 Vol. 8, 1997. pp. 1442-1463.

[2] D. F. Murphy, D. A. Flavin, R. McBride, and J. D. C. Jones, Interferometric Interrogation of In-Fiber Bragg Grating Sensors Without Mechanical Path Length Scanning, *J. Lightwave Techn.*, No. 7 Vol. 19, 2001. pp. 1004-1009.

[3] Alves, J. Santos, J.L. Carvalho, A. Lage, A. "Fiber Bragg sensor interrogation system based on a CCD spectrometer", Proceedings of IEEE, Sensors 2003, pp. 909-913.

[4] A. C. Tamhane, "Statistical Analysis of Designed Experiments: Theory and Applications", Wiley, 2009

[5] T. Wilmshurst, "Designing Embedded Systems with PIC Microcontrollers, Second Edition: Principles and Applications", Newnes, 2007.

Distance Uncertainties of Nodes in Wireless Sensor Networks using Multilateration Algorithm

Mare Srbinovska[1], Cvetan Gavrovski[2] and Vladimir Dimcev[3]

***Abstract* – A multilateration algorithm is presented in this work in order to realize a precise localization of the sensor nodes in wireless sensor networks. Received Signal Strength Indicator (RSSI) is used for distance measurements between sensor nodes. The developed algorithm is implemented in a real sensor network. The test results show that determination of the distances of sensor nodes using this algorithm reaches the intended requirements.**

***Keywords* – Multilateration algorithm, Localization, Wireless sensor networks, RSSI.**

I. INTRODUCTION

Wireless sensor networks (WSNs) have been used in a wide range of applications. They usually consist of a large number of low-cost, low-power, multifunctional sensor nodes that are small in size and communicate in short distances. Their structure and characteristics depend on their electronic, mechanical and communication limitations but also on the requirements of the specific application for example for monitoring the environment, water, air, soil, etc [9]. The position of sensor nodes is usually not pre-determined, although the application can provide some guidelines and insights that can lead to the construction of an optimal design that satisfies application requirements and meets wireless network limitations.

The progress of both sensing and information technology is making the physical world measurement data more available and cheap in terms of acquisition, processing, storage and retrieval. Intelligent sensors, which are possibly arranged in wired or wireless networks, can acquire raw data from the environment for long periods of time and those data can be safely stored virtually forever.

The problem of estimating spatial – coordinates of the node is referred as localization [5-7]. The main objective is to locate each node as accurately as possible with a certain amount of error about the distances between a subset of nodes in wireless sensor networks. Many studies about localization in WSN with different devices and algorithms are analyzed.

[1]Mare Srbinovska is with the Faculty of Electrical Engineering and Information Technologies, Karpos 2 bb POBox 574, 1000 Skopje, Macedonia, E-mail: mares@feit.ukim.edu.mk

[2]Cvetan Gavrovski is with Faculty of Electrical Engineering and Information Technologies, Karpos 2 bb POBox 574, 1000 Skopje, Macedonia, E-mail: cvetang@feit.ukim.edu.mk

[3]Vladimir Dimcev is with Faculty of Electrical Engineering and Information Technologies, Karpos 2 bb POBox 574, 1000 Skopje, Macedonia, E-mail: vladim@feit.ukim.edu.mk

The simplest method for a two dimensional localization is trilateration using distance measurement from an unknown sensor node to three non-linear anchors with known positions. To reduce the influence of distance errors on localization accuracy a multilateration algorithm with more than three anchors is used.

The distances between the sensor nodes can be determined by measuring time of arrival (ToA), time difference of arrival (TDoA) or received signal strength indicator (RSSI).

II. LOCALIZATION ALGORITHMS

A. *Lateration and angulation*

In many applications, sensors have to know their geographical locations. In reality, it is not practical to use GPS in every sensor node because a sensor network consists of thousands of nodes and GPS becomes very costly. Instead of requiring every node to have GPS installed, all localization methods assume only a few nodes equipped with GPS hardware. These nodes are called anchor or beacon nodes and they know their positions without communicating with other nodes.

Most of the existing works focus on increasing the accuracy in position estimation by using different mathematical techniques such as triangulation, multilateration, multidimensional scaling, etc. Lateration and angulation are the two main localization algorithms [8]. Lateration uses distances between the sensor nodes while the angulation determines the position based on angles. Compared with angulation using specific antennas to determine the angles the lateration algorithm is easier to apply due to the simple distance measurements.

Trilateration is a method of determining the relative positions of objects using the geometry of triangles, so to accurately and uniquely determine the relative location of a point on a 2D plane a minimum of 3 reference points are needed (fig.1).

Multilateration localization algorithm with more anchors is used to reduce the influence of the distance errors on localization results.

The process of localization by solving for the mathematical intersection of multiple hyperbolas based on the RSSI, ToA, TDoA is based on the multilateration algorithm (Fig.1). In multilateration when N receivers are used, it results in N-1 hyperbolas. If a large number of receivers are used, N>4, then the localization problem can be posed as an optimization problem that can be solved using, among others, a least square method.

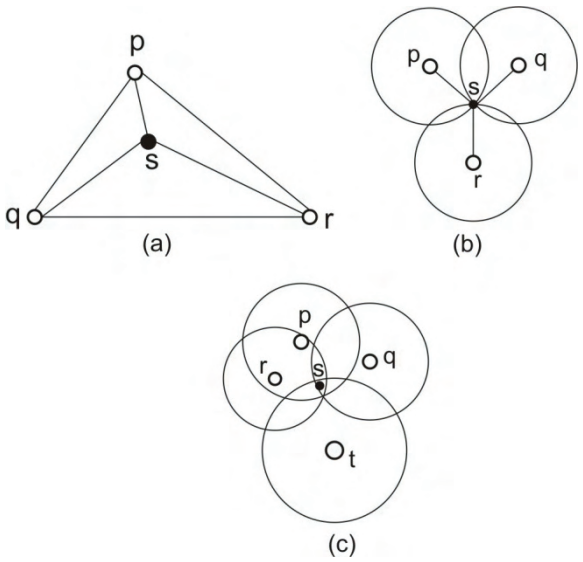


Fig. 1: (a) Triangulation (b)Trilateration (c) Multilateration

B. Received Signal Strength Indicator (RSSI)

Lots of localization algorithms require a distance to estimate the position of unknown devices. In addition to mere connectivity information, the communication between two nodes often allows to extract information about their geometric relationship. Using elementary geometry, this information can be used to derive information about node positions.

The characteristics of wireless communication are partially determined by the distance between sender and receiver, and if these characteristics can be measured at the receiver they can serve as an estimator of distance. The most important characteristics are Received Signal Strength Indicator (RSSI), Time of Arrival (ToA), and Time Difference of Arrival (TdoA).

- Received Signal Strength Indicator (RSSI) [3] techniques measure the power of the signal at the receiver. Based on the known transmit power, the effective propagation loss can be calculated. Theoretical and empirical models are used to translate this loss into a distance estimate. This method has been used mainly for RF signals.

In the case where there is a direct path between a transmitter and receiver. The receiver signal power, P_R is related to distance, d, by the inverse square law.

$P_R \propto d^{-2}$

However this is an ideal case for a point source. In the real world the signal often decays at a faster or slower rate

$P_R \propto d^{-n}$

Where n is the loss exponent. An excepted form of the relation between distance and receive power simplified for the case of a one meter reference distance is:

$$RSSI\ [dBm] = -(10 \cdot n \cdot log_{10}(d) + A) \qquad (1)$$

where d denotes the transmitter-to-receiver distance, n is the propagation path-loss exponent, A represents the RSS value measured by a receiver that is located 1m away from a transmitter, and RSS is the actual measured value at distance d. The signal propagation coefficient n shows the damping of the signal. Both parameters must be determined empirically.

In practical scenarios, the ideal distribution of RSSI is not applicable, because the propagation of the radio signal is interfered with a lot of influencing effects e.g.

- Reflection of objects;
- Superposition of electro-magnetic fields;
- Diffraction at edges;
- Refraction by media with different propagation velocity.

The low complexity and the fast calculation recommend this localization algorithm as very popular and often used in wireless sensor networks.

- Time based methods (ToA,TDoA) record the time-of-arrival (ToA) or time difference-of-arrival (TDoA). The propagation time can be directly translated into distance, based on the known signal propagation speed. These methods can be applied to many different signals, such as RF, acoustic, infrared and ultrasound.

- Angle -of -Arrival (AoA) systems estimate the angle at which signals are received and use simple geometric relationships to calculate node positions.

C. Multilateration

Multilateration [4] is one of the most popular techniques for positioning applied in wireless sensor networks.

The real challenge for triangulation arises when the distance measurements are not perfect, but only estimates $\hat{d}$ with an unknown error ε are known. The intuitive solution to this problem is to use more than three anchors and redundant distant measurements to account into an overdetermined system of equations, written in matrix form:

$$2 \cdot \begin{bmatrix} x_n - x_1 & y_n - y_1 \\ \mathrm{M} & \mathrm{M} \\ x_n - x_{n-1} & y_n - y_{n-1} \end{bmatrix} \cdot \begin{bmatrix} x_u \\ y_u \end{bmatrix} =$$

$$\begin{bmatrix} \left(d_1^2 - d_n^2\right) - \left(x_1^2 - x_n^2\right) - \left(y_1^2 - y_n^2\right) \\ \mathrm{M} \\ \left(d_{n-1}^2 - d_n^2\right) - \left(x_{n-1}^2 - x_n^2\right) - \left(y_{n-1}^2 - y_n^2\right) \end{bmatrix} \qquad (2)$$

For system of linear equation, a solution can be computed that minimizes the mean square error, the solution is the pair (x_u, y_u) that minimizes $\|\mathrm{Ax} - \mathrm{b}\|_2$, where b is the right-hand side from equation (2). Minimizing this expression is equivalent to minimizing the mean square error. As a function of x, its gradient has to be set to zero:

$$A^T Ax = A^T b \qquad (3)$$

Equation (3) has a unique solution under certain conditions

D. Propagation coefficient n

To determine n, RSSI values within 30m are measured and compared with theoretical curves according to equation 1 whereby n is varied from 1 to 4 (figure 2). The majority of the RSSI values are in the area between theoretical curves n=2 and n=3, thus more n values in interval [2.0, 3.0] are researched. The root mean square error between theoretical and measured RSSI values [1], [2] is calculated and compared

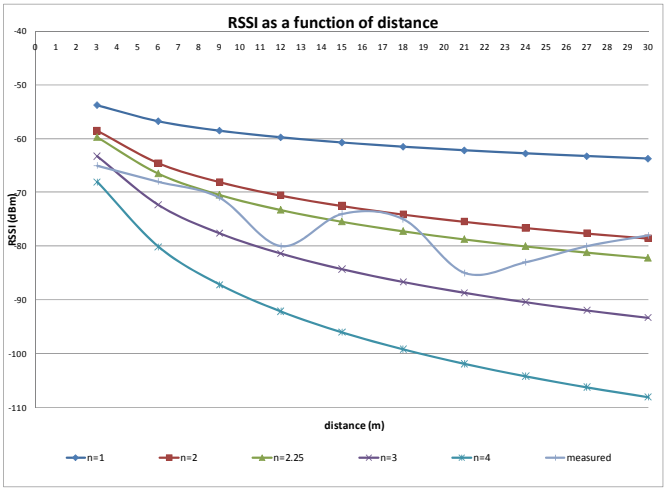


Fig. 2. Measured vs. calculated RSSI with different n

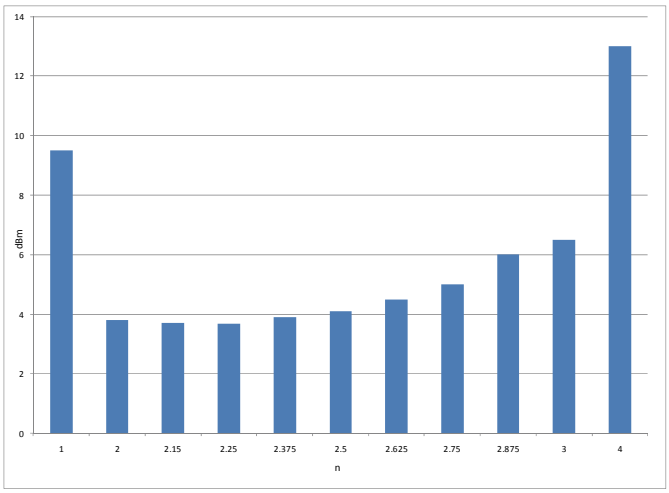


Fig. 3. Root mean square error between theoretical and measured RSSI with different n

with each other. Due to the minimal root mean square error n=2.25 is chosen as the adequate damping factor. Figure 3 shows the typical errors of distance measurements with RSSI.

III. Implementation and Results

The localization algorithm is implemented in a real sensor network which consists of 6 sensor nodes. For the experimental setup a 2.4GHz 802.15.4 Zigbee development kit belonging to the Silicon Laboratories network, is used. The board is shown in Fig. 4.

Each board features a silicon laboratories C8051F121 microcontroller and a Chipcon CC2420 2.4 GHz 802.15.4 transceiver [10]. Support components include a USB interface, JTAG programming interface. The data transfer rate goes up to 250kBps, the radiation power is 1mW (low power mode) or 40mW (high power mode).

The distance of the nodes is estimated by measuring the RSSI from all nodes, where the positions of three random nodes are fixed. The position of the unknown node is determined using the method of multilateration.

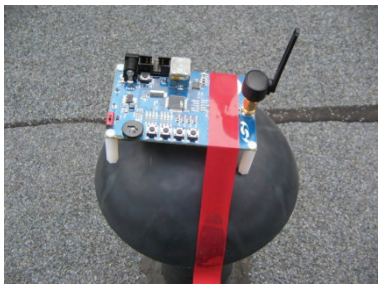


Fig. 4 Sensor node

In order to estimate the RSS-to-distance curve a set of sensor nodes has been positioned as in Fig.5. The positions of three main nodes are known (in the figure named with A_4, A_5 and A_6), the received signal strength to this nodes are known, the location of the three nodes named with X_1, X_2 and X_3 should be determined.

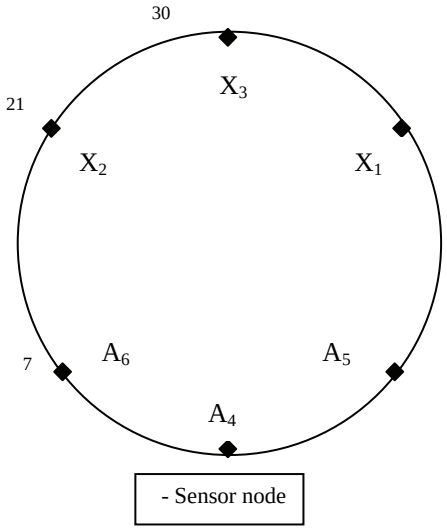


Fig. 5. Position of the sensor nodes in free space

In general, the result of a measurement is only an approximation or estimate of the value of the measurand and thus is complete only when accompanied by a statement of the uncertainty of that estimate [12]. Thus the ideal method for evaluating and expressing measurement uncertainty should be capable of readily providing an interval, in particular, one with a coverage probability or level of confidence that corresponds in a realistic way to that required.

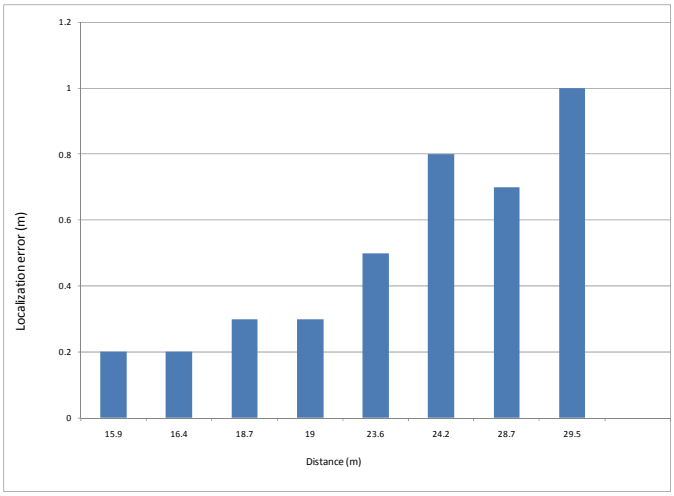


Fig. 6 Distance uncertainties of the sensor nodes

The uncertainties [11] for the distances of the sensor nodes are shown in figure 6. The positions of the nodes are given in figure 5. The localization error is increasing, when the receiving nodes depart from the transmitter. The largest distance between nodes is 30 meters, for this distance the uncertainty is 1m. This inaccuracy is due to the fact that receiver nodes are located very closed to huge buildings, which provides constructive reflected propagation path in addition to the direct one.

IV. Conclusion

Wireless sensor networks are widely applicable to many practical applications including environmental monitoring, military applications, etc. in which sensors may need to know their geographical locations. The multilateration localization algorithm in wireless sensor network was explained. The relationship between RSSI and distance was determined through practical experiments. Distance uncertainties of the sensor nodes through experimental measurements were presented.

References

[1] M.Srbinovska, C. Gavrovski, V.Dimcev, "Localization System using RSSI Measurement of Wireless Sensor Network based on Zigbee Standard", XLIII International Scientific Conference ICEST, Nis, 2008

[2]M.Srbinovska, C.Gavrovski, V.Dimcev, "Optimization Algorithm for Positioning Estimation In Wireless Sensor Networks",XLIV International scientific and applied science conference ICEST 2009, Veliko Trnovo, Bulgaria

[3] Dimitrios Lymberopoulos, Quentin Lindsey and Andreas Savvides, "An Empirical Characterization of Radio Signal Strength Variability in 3-D IEEE 802.15.4 Networks using Monopole Antennas, Yale University, New Haven, USA, 2005

[4] Holger Karl and Andreas Willig, "Protocols and Architectures for Wireless Sensor Networks", John Wiley & Sons, 2005

[5] Marina Petrova, Janne Ruhijärvi, Petri Mähönen and Saverio Labella, "Performance Study of IEEE 802.15.4 Using Measurements and Simulations", RWTH Aachen University, Germany, 2004.

[6] Shashank Tadakamadla, "Indoor Local Positioning System for Zigbee based on RSSI", Mid Sweden University, 2005

[7] E. Callaway, P. Gorday, L. Hester, J. A. Gutierrez, M. Naeve, B. Heile, and V. Bahl, "Home Networking with IEEE 802.15.4: A Developing Standard for Low-Rate Wireless Personal Area Networks," *IEEE Communications Magazine*, vol. 40, no. 8, pp. 70–77, August 2002.

[8] Jason Lester Hill, "System Architecture for Wireless Sensor Networks", PhD Thesis, University of California, Berkeley, 2003

[9] Kazem Sohraby, Daniel Minoli, Taieb Znati, "Wireless Sensor Networks: Technology, Protocols and Applications, 2007

[10] Datasheet for chipcon CC2420 2.4 GHz IEEE 802.15.4/ZigBee RF Transceiver

[11] M.Srbinovska, "*Model for position estimation of smart sensors in distributed measurement systems*", Master Thesis, Ss. Cyril and Methodius University, Skopje, 2008

[12] Guide to the expression of uncertainty in measurement, European committee for standardization, Brussels, 1999

Procedure for Verification of Measuring Transformer Accuracy based on LabVIEW Software

Milan M. Simić[1] and Božidar R. Dimitrijević[2]

***Abstract* – Solution of the experimental measurement and data acquisition system, designed for verification of voltage measuring transformer accuracy in laboratory conditions, is presented in this paper. Functional basis of developed solution is provided by virtual instrumentation software package LabVIEW. Described acquisition system includes standard PC configuration platform with application software, multifunctional calibration instrument Fluke 5100B for generation of reference voltage signals and data acquisition card PCI NI 6251. Software application developed in LabVIEW programming environment, performs measurement, recording, presentation and statistical processing of transformer output voltage basic parameters. Statistical analysis of obtained measuring results, with possibility for graphical presentation of recorded time diagrams for measured voltage parameter values, provides estimation of the mean measured parameter values and measuring uncertainty components, according to the documents for estimation and expression of the uncertainty in measurement.**

***Keywords* – Voltage measuring transformer, LabVIEW virtual instrumentation software, Measuring uncertainty.**

I. INTRODUCTION

With rapid growth of the electrical energy consumption and limitations of natural resources for energy production, a basic demand is increasing of the total efficiency level in electrical energy production, distribution and consumption processes. In globally widespread process of liberalization on the electrical energy world market, within the last ten years are created new technological demands for research and development facilities and institutions over the world. Significant characteristics of this process are involving and integration of wide geographic areas in electrical energy trading operations, which include the increased number of energy market participants, much greater amount of the business transactions and significantly higher number of data and information which must be exchanged and processed. Very important segments in the mentioned process of the energy efficiency level increasing, are measurement and analysis of the basic quality parameters regarding to electrical energy delivered to the individual customers. Measurement of basic voltage, current, power and energy parameters includes measuring transformers, applied in various types of regulation and control energy facilities. Measuring accuracy of the used voltage and current transformers, conclusively affects on total accuracy of the complete measurement process. Especially are important measurements performed for precise determination of the individual customers energy consumption level, which demands measuring transformers of high accuracy class [1].

Procedure for checking of voltage measuring transformers accuracy and condition diagnostic need to be performed inside accredited metrological laboratories or directly on locations of the remote measuring transformer facilities, periodically in the specified time intervals. Such accuracy verification procedure requires sophisticated measurement methods and appropriate measurement and control devices, having high accuracy level. This equipment foremost considers calibration instrument for generation of the reference input voltage parameters, defined by relevant documents, with checking instruments applied for measurement of the transformer output voltage parameters.

Realization of experimental laboratory measurement and data acquisition system, described in this paper, developed for checking of the low voltage measuring transformer accuracy, is functionally based on 16-channel data acquisition PCI card NI 6251 [2], supported by PC programming application inside LabVIEW graphical software environment [3]. As a reference instrument for providing transformer input standard voltage signals with 230V RMS value and 50Hz frequency, is applied multifunctional calibration device Fluke 5100B [4]. Designed programming application performs measurement, continuous recording, graphical presentation and statistical processing of voltage parameter values, measured on controlled transformer voltage outputs. Software processing of the measured values is performed according to relevant documents which prescribe calculation and expression of the measuring uncertainty data.

II. HARDWARE BLOCK CONFIGURATION OF THE MEASUREMENT AND ACQUISITION SYSTEM

Configuration of the experimental laboratory measurement and acquisition system, for low voltage measuring transformer accuracy checking, developed at Department of Measurement on the Faculty of Electronic Engineering in Niš, is presented in the Fig 1. Hardware configuration of the developed solution includes reference calibration instrument Fluke 5100B, data acquisition PCI card NI 6251 and LabVIEW 8.0 application software support, installed on the standard PC programming platform. Calibration device Fluke 5100B is microprocessor based instrument programmable by the users from instrument front panel. Depending on actual requirements, this instrument provides generation of reference AC voltage waveforms with RMS values within the range from 1mV to 1100V and signal frequency from 50Hz to 1KHz. For this specific application, calibration device generates reference standard voltage signals for measuring transformer inputs, having the basic parameters

[1]Milan M. Simić is with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia. E-mail: milan.simic@elfak.ni.ac.rs

[2]Božidar R. Dimitrijević is with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia. E-mail: bozidar.dimitrijevic@elfak.ni.ac.rs

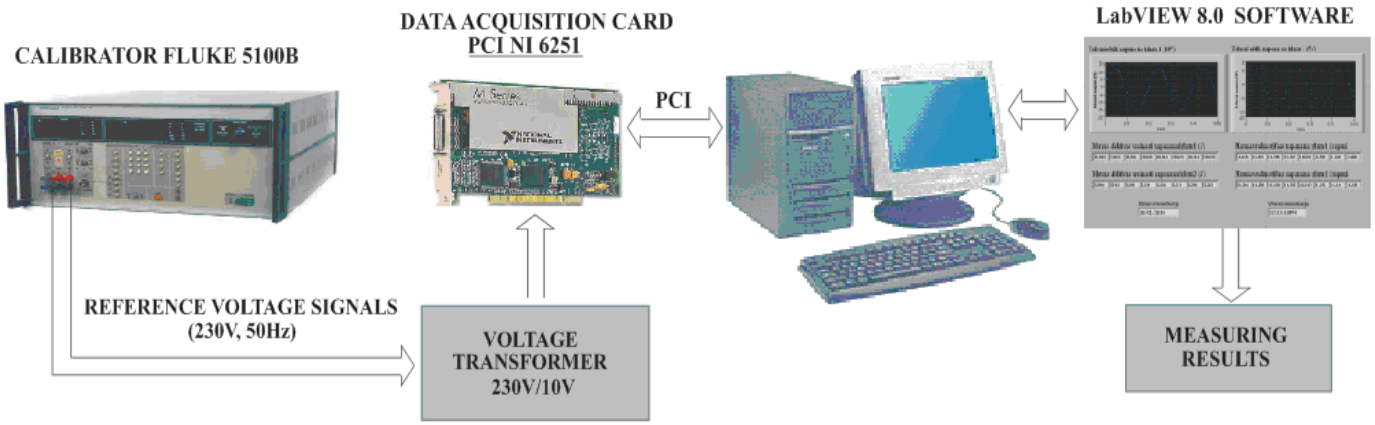


Fig. 1. Simplified block configuration of the developed experimental measurement and data acquisition system

230V RMS values and 50Hz signal frequency, according to European power quality standard EN 50160, which prescribes basic voltage characteristics for electricity supplied using the public distribution systems [5]. Reference voltage waveforms generated on calibration device outputs, are transferring using the standard BNC connectors directly to the inputs of voltage masuring transformer, which accuracy need to be checked. In the next segment of this procedure, voltage signals generated on measuring transformer outputs, within the range of ±10V, must be transferred to the inputs of data acquisition card A/D converter, with 16-bit resolution [2]. Multichannel acquisition card NI 6251, from american manufacturer company National Instruments Corporation, for this specific application purpose uses two analog input channels, which receive voltage signals from measuring transformer outputs having specified 10V and 5V nominal RMS voltage values. Internal communication and two-way data interchange between data acquisition card and PC computer is provided using standard PCI communication interface. Described procedure for verification of the voltage measuring transformer accuracy is software controlled using programming application designed in LabVIEW environment, which will be described in the following segment of the paper.

III. LabVIEW SOFTWARE SUPPORT OF THE SYSTEM

Concept of the virtual measuring instrumentation is method for development of measuring instruments, based on standard PC computers or industrial operational stations, cost effective hardware components applied for measuring data acquisition and programming packages specialized for software supported analysis and graphical presentation of the obtained measuring results. Hardware segment of the virtual measuring instrument is consisting of computer and data acquisition card. Software section of such instruments is programmed depending on user requirements using predesigned functioning blocks, individual elements and instrument front panels from software package databases. A most important advantages of virtual measuring instruments are possibilities for performing simple and fast corrections of the software algorithm sequence which controls execution of the computer supported measuring procedure [3].

Automated procedure for measuring transformer accuracy checking, described in this paper, functionally is supported by control software application developed using LabVIEW 8.0 graphical programming package. Developed control software application performs measurements, chronological recordings,

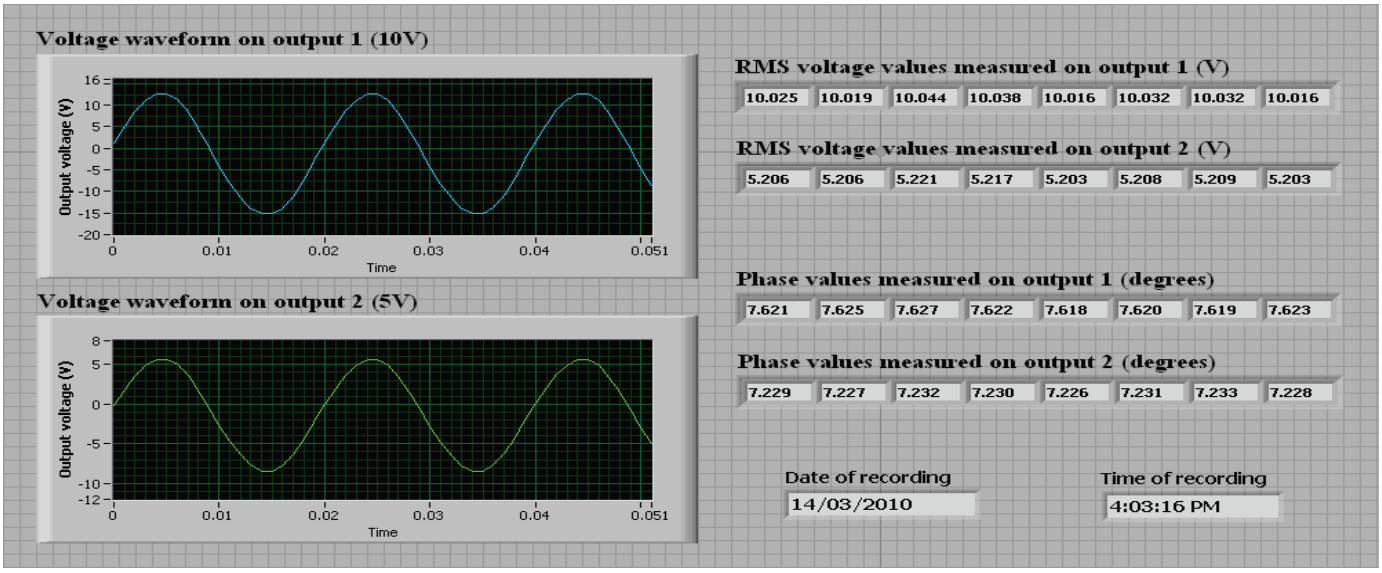


Fig. 2. LabVIEW virtual instrument for graphical presentation of the output voltage waveforms and measured voltage parameters

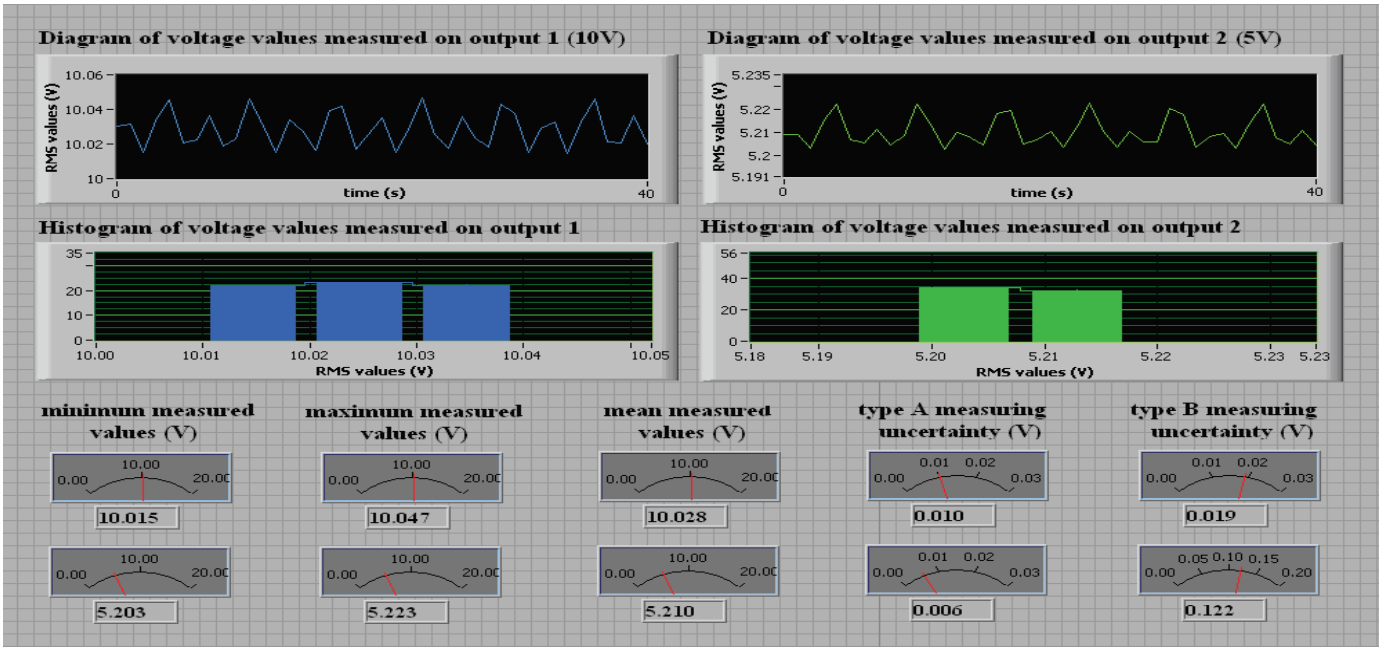


Fig. 3. Statistical processing of the measured RMS output voltage values performed in LabVIEW programming environment

graphical presentations and statistical processing of measuring results concerning RMS voltage values on transformer output, phases, frequency and high-order harmonic components of the output voltage signals. Front panel of the virtual instrument, developed in the LabVIEW programming environment, which provides comparative graphical presentation of voltage signals recorded on measuring transformer outputs, is presented in the Fig 2. Besides graphical presentation of output voltage signal waveforms, on the virtual instrument front panel are indicated measuring results concerning RMS voltage values and phases measured on the transformer 10V and 5V outputs. Additional information regarding to exact date and time of the presented voltage signals recording are also provided on this front panel.

Software supported procedure for statistical analysis of the obtained measuring results regarding to RMS voltage values measured on transformer outputs, performed in the LabVIEW environment, is illustrated in the Fig 3. Developed application software performs comparative graphical presentation of the recorded time diagrams and corresponding statistic histograms for measured RMS values of transformer output waveforms. On presented virtual instrument front panel are also indicated minimum and maximum voltage values obtained during this measuring process on voltage transformer outputs, including calculations and numeric presentation of the mean measured voltage values, with the corresponding measuring uncertainty components for both transformer outputs. The mean measured RMS output voltage values indicated on the virtual instrument front panel, are calculated as an arithmetical mean of obtained measured voltage values, as it shown in the following relation:

$$V_{mean} = \frac{1}{n}\sum_{i=1}^{n} V_i \tag{1}$$

Procedure for calculation of measuring uncertainty values is performed according to recommendations of document: Guide to the Expression of Uncertainty in Measurement [6], defined from the International Organization for Standardization - ISO.

Calculation of type A standard uncertainty component values is performed according to statistical procedures applied on the obtained measuring results, by means of following square root equation for statistical standard deviation of measured values:

$$u_A(V) = \sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}\left(V_i - V_{mean}\right)^2} \tag{2}$$

Standard measuring uncertainty components of the type B are estimated on the basis of data obtained from specifications of the reference calibration instrument Fluke 5100B, provided by instrument manufacturer. According to these specifications for nominal frequency value of 50Hz, reference AC voltage signals are generated with specified nominal relative accuracy value of 0.05% [4]. Considering this specified accuracy value, for type B standard uncertainty components of the measured transformer output voltages are obtained the following results:

$$u_{B1}(V) = \frac{1}{\sqrt{3}}\frac{0.05}{100}10V + \frac{0.028V}{\sqrt{3}} = 0.019V \tag{3}$$

$$u_{B2}(V) = \frac{1}{\sqrt{3}}\frac{0.05}{100}5V + \frac{0.210V}{\sqrt{3}} = 0.122V \tag{4}$$

From previously calculated individual values of measuring uncertainty components of the type A and type B, combined measuring uncertainty components for measured transformer output voltages are calculated using next square root relations:

$$u_{C1}(V) = \sqrt{u_{A1}^2 + u_{B1}^2} = 0.021V \tag{5}$$

$$u_{C2}(V) = \sqrt{u_{A2}^2 + u_{B2}^2} = 0.122V \tag{6}$$

Software processing of the voltage signal frequency values measured on first and second voltage output of the measuring transformer and performed by means of LabVIEW application software, is presented in the Fig 4. Similar with the previously described procedure for statistical processing of the measured RMS output voltage values, on shown virtual instrument front

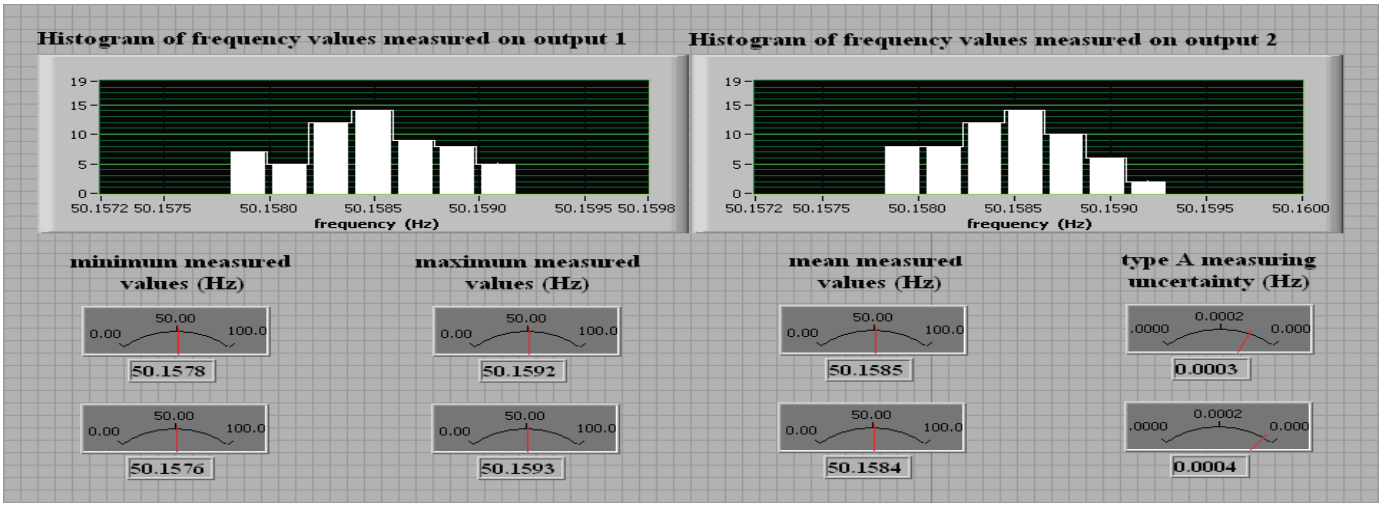


Fig. 4. LabVIEW based statistical processing of the signal frequency values measured on the voltage transformer outputs

panel are illustrated two statistical histograms concerning the signal frequency values measured on 10V and 5V transformer voltage outputs. Presented statistical analysis procedure also includes indication of the minimum and maximum measured frequency values, together with calculation of mean measured frequencies and measuring uncertainty components of type A.

Finally, a form of the final report from previously described procedure applied for voltage measuring transformer accuracy verification, is presented in the Table 1. By this table is given short summary of the measured values concerning transformer output voltage basic parameters, obtained from this procedure.

TABLE I
FINAL SUMMARY OF THE MEASURED OUTPUT VOLTAGE BASIC PARAMETERS

measurement of the RMS voltage values		
measured values	output 1	output 2
minimum values	10.015 V	5.203 V
maximum values	10.047 V	5.223 V
mean values	10.028 V	5.210 V
type A uncertainty	0.010 V	0.006 V
type B uncertainty	0.019 V	0.122 V
combined uncertainty	0.021 V	0.122 V
measurement of the frequency values		
minimum values	50.1578 Hz	50.1576 Hz
maximum values	50.1592 Hz	50.1593 Hz
mean values	50.1585 Hz	50.1584 Hz
type A uncertainty	0.0003 Hz	0.0004 Hz
measurement of the high-order harmonics		
3rd order harmonic	1.583 %	1.589 %
5th order harmonic	0.492 %	0,495 %
7th order harmonic	0.315 %	0,319 %
9th order harmonic	0.194 %	0.197 %

IV. CONCLUSION

Possibilities of using virtual instrumentation software in the procedure for metrological verification of voltage transformer accuracy, based on software processing of measured voltage basic parameter values, are presented in this paper. Reference standard voltage waveforms for measuring transformer inputs are provided using calibration device Fluke 5100B. Hardware section of the developed measurement and acquisition system includes PC computer and data acquisition card PCI NI 6251. Programming application designed in LabVIEW 8.0 software package performs measurement, recording, presentation and statistical processing of measured RMS output voltage values, phase, frequency and high-order harmonic components. This software analysis includes graphical presentation of measured values time diagrams, indication of minimum and maximum measured values, calculation of the mean measured values and standard measuring uncertainties. Presented solution based on the cost-effective acquisition hardware components eliminates demands for manual measurements of output voltage for each point of the input voltage changing, due to providing complete software automation of the metrological verification process.

REFERENCES

[1] T. Chiulan, B. Pantelimon, “Power Transformer Units Condition Assessment Using Virtual Instrumentation”, Jour. of Electronics and Electrical Engineering, Telecommunications Engineering, No. 6(86), 2008.
[2] Data Acquisition Card NI PCI 6251 Specifications, National Instruments Corporation, USA, 2005 (http://www.ni.com).
[3] LabVIEW 8.0 User Manual, National Instruments Corporation, USA, 2007 (http://www.ni.com).
[4] Multifunction Calibrator Fluke 5100 Series B Instruction User Manual, Fluke Corporation, USA.
[5] EN 50160 Power Quality Standard, Power Quality Access Meters and EN50160, Siemens, May, 2003.
[6] Guide to the Expression of Uncertainty in Measurement, ISO organization, Geneva, Switzerland, 1993.

Optimal Charging Time Interval in Direct Sensor to Microcontroller Interface

Zivko D. Kokolanski[1], Cvetan V. Gavrovski[1] and Vladimir I. Dimcev[1]

***Abstract* – The resistive and capacitive modulating sensors can be directly measured with microcontroller by measuring the charging or discharging time of RC circuit. The measurement contains two phases: charging and discharging phase. In the first phase the charging/discharging time interval should be short enough to ensure high speed measurements and in the same time to be long enough to ensure high resolution. Therefore, the optimal charging/discharging time interval that gives the best speed/resolution trade-off has to be determined. In this paper theoretical and experimental analysis for the optimal charging/discharging time interval are presented.**

***Keywords* – Direct sensor-microcontroller interface, Optimal charging time interval, Measurement**

I. INTRODUCTION

Direct sensor to microcontroller interface is an alternative approach for conditioning of modulating resistive and capacitive sensors without the use of an A/D converter. The microcontroller uses the built in timer to measure the charging or discharging time of RC circuit formed by the sensor and reference resistor/capacitor. In this way, the microcontroller and the sensor form a relaxation oscillator causing the modulating sensor to act like a quasi-digital sensor.

Two measurement methods are proposed: a method based on charging [1] or discharging time [2] of the RC circuit. The two methods differentiate by the crossing of the upper or the lower threshold voltage (V_{th} or V_{tl}) of the Schmitt trigger port to create an interrupt. The method based on discharging time gives better measurement results [3] because the lower threshold voltage V_{tl} has better rejection of the power supply interference and because usually the microcontroller ports can sink more current than they can source. In this paper the analysis are restricted to the interfaces based on the measurement of the discharging time but the same methodology can be applied to the interfaces based on the measurement of the charging time.

II. INTERFACE BASED ON DISCHARGING TIME

The most basic direct sensor to microcontroller interface can be realized by using two microcontroller pins, one output and one input pin. Simplified electrical circuit of the direct sensor -microcontroller interface based on measurement of discharging time is shown in Fig. 1.

[1]Authors are with the Faculty of Electrical Engineering and Information Technologies, Karpos 2 bb, 1000 Skopje, Macedonia, E-mail: zivko.kokolanski@feit.ukim.edu.mk

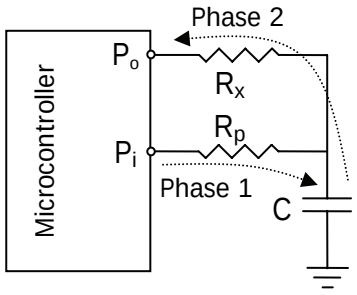


Fig. 1. Direct sensor–microcontroller interface based on measurement of discharging time

The measurement contains two phases: charging phase and discharging phase. The wave shape of the capacitor voltage in the two phases is shown in Fig.2.

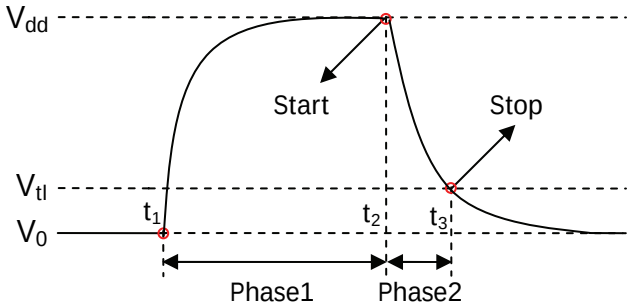


Fig. 2. Wave shape of the capacitor voltage in the two measurement phases

At the beginning the pin P_i is set as output with logical state "1" and the pin P_o is set as input (high impedance state). The capacitor charges through R_p to V_{dd} in a period $t_1 \div t_2$. In the next step the pin P_o is set as output with logical state "0", the timer starts and the pin P_i is set to high impedance state. This time the capacitor discharges through R_x until the voltage reaches the lower threshold voltage V_{tl}. Crossing of the threshold voltage V_{tl} initiates interrupt that stops the timer.

The capacitor voltage in the second (discharging) phase (Fig.2) can be expressed as

$$V_c(t) = V_0 + (V_{dd} - V_0)e^{-\frac{t}{\tau}} \tag{1}$$

where $\tau = R_x C$ is the discharging time constant. Here, the time needed for the capacitor to discharge from V_{dd} to V_0 is

$$t_x = (t_3 - t_2) = \tau \ln\left(\frac{V_0 - V_{dd}}{V_0 - V_{tl}}\right) \tag{2}$$

Having in mind that V_0, V_{dd}, V_{tl} and C are constant, from (2) can be seen that the time interval t_x is proportional to the measuring resistance R_x. This time interval (t_x) is measured with the built in timer in the microcontroller. The result of the time to digital conversion can be expressed as

$$N = kR_x \tag{3}$$

where k is constant dependent on V_0, V_{dd}, V_{tl}, C and the time base of the timer. In practice the input/output resistances and leakage currents of the microcontroller ports cause gain, offset and nonlinearity errors [4]. Additionally the constant (k) in the equation (3) is not very stabile. Therefore, in practice direct sensor-microcontroller interface is realized by using some calibration technique [5] that cancels the contribution of V_0, V_{dd},V_{tl} and C.

III. Optimal Charging Time Interval

The charging time interval in the first phase (phase1) in Fig.2 can be expressed with the equation

$$V_c(t) = V_{dd} + (V_0 - V_{dd})e^{-\frac{t}{\tau}} \tag{4}$$

where $\tau=R_pC$ is the charging time constant. From (4) it can be seen that theoretically the capacitor voltage will reach V_{dd} in infinity. If we take a finite value for the charging time interval in phase1 (Fig.2), than we introduce error in the equation (1) taking $V_c(t_2)=V_{dd}$. This error is smaller for longer charging time intervals, but also longer charging time intervals decrease the speed of the measurement. One possible solution for increasing of the speed of measurement is reduction of the resistance R_p. However this resistance is limited by the maximal current that the microcontroller can source. Also higher value of R_p improves the power supply noise rejection ratio of the measuring system [6].

Very often in the literature that deals with the direct sensor -microcontroller interface, charging time interval higher than 5τ is recommended. However this is not always true and the charging time interval must be chosen depending on the desired resolution. The error of the time interval measurement must be smaller than half of the least significant bit (LSB) for the desired resolution. Therefore the optimal charging interval that gives best speed/resolution trade-off has to be determined.

The relative error of the capacitor voltage in phase1 is expressed with the equation

$$\Delta V_c(t_2) = \frac{V_c(t) - V_c(\infty)}{V_c(\infty)} 100[\%] \tag{5}$$

Considering the Eq. (4) and if we take $V_c(\infty)=V_{dd}$, the Eq. (5) becomes

$$\Delta V_c(t_2) = 100\left(\frac{V_0}{V_{dd}} - 1\right)e^{-\frac{t}{\tau}}[\%] \tag{6}$$

Similarly, considering the Eq. (2) the relative error for the measured time interval is

$$\Delta t_x = \frac{\ln\frac{V_0 - V_c(t_2)}{V_0 - V_{dd}}}{\ln\frac{V_0 - V_{dd}}{V_0 - V_{tl}}} 100[\%] \tag{7}$$

The relative errors of the capacitor voltage (6) and the time interval measurements (7) for charging time intervals of $1\div10\tau$ are given in Table I.

TABLE I
RELATIVE ERRORS OF THE CAPACITOR VOLTAGE AND TIME INTERVAL

t/τ	$V_c(t_2)$[V]	ΔV_c[%]	Δt_x[%]
1	3.22	35.3	/
2	4.33	12.9	10.5
3	4.74	4.7	3.6
4	4.89	1.7	1.3
5	4.94	0.64	0.4
6	4.96	0.23	0.17
7	4.97	0.08	0.06
8	4.97	0.03	0.02
9	4.97	0.01	0.008
10	4.98	0.004	0.003

The relative errors in Table I are calculated by using V_0=0.02, V_{dd}=5V and V_{tl}=1.4V. The relative errors are shown on Fig.3.

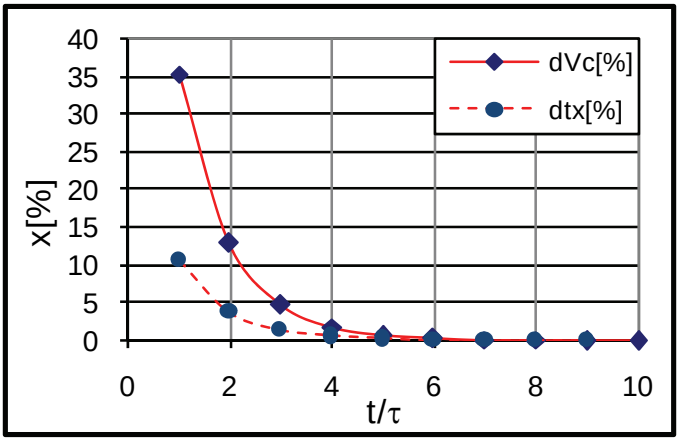


Fig. 3. Relative errors of the capacitor voltage and time interval

From the results reported in Table I and in Fig.3 it can be seen that for charging time interval of 1τ, the high threshold voltage V_{th}=3,6V cannot be reached. Therefore measurement with such charging time is impossible. For charging time interval 5τ, the capacitor voltage reaches 99.5% of the voltage V_{dd}. The relative error of the measured time interval in this case will be 0.4%. This relative error is lower than 0.5LSB for resolution of 6 bits. Therefore if we want to achieve higher resolution, the charging time interval must be proportionally increased. For example if the desired resolution is 12 bits, then the charging time interval must be longer than 9τ. Hence, the optimal charging time should always be determined with respect to the desired resolution. The relative error in that case should be lower than 0.5LSB.

IV. Experimental results and discussion

The theoretical analyses in Section III are experimentally verified for charging time intervals 3τ÷10τ. The experiments are realized by using microcontroller PIC16F877 [7] with clock frequency of 8 MHz, effective instruction cycle speed 2MHz and period 0.5μs. The falling edge of the input signal was registered with the RB0/INT Schmitt trigger pin. This pin initiates interrupt that stops the 16-bit timer - Timer1.

The results of the measurements are sent to personal computer through the serial RS232 port. The MAX232 (TTL/RS232) level translator was supplied from separate power supply to prevent transients (about 170 kHz) from interfering with the power supply rails of the microcontroller. To reduce the noise effects affecting the voltage comparison between V_c and V_{tl} several design solutions were applied:

- Decupling capacitor of 100nF was placed as close as possible to the microcontroller pins as recommended from the manufacturer
- The board ground plane was carefully designed for low electromagnetic interference
- Only the microcontroller was supplied from the power supply to eliminate other interference effects
- The microcontroller didn't execute any other task while waiting for the interrupt
- The program algorithm was not changed while performing the experiments

The passive components were measured with measuring instrument with maximal error of ±0.1%+5 for resistance and ±1%+5 for capacitance measurement. The measured and the nominal values are given in Table II.

Table II
The measured and the nominal values of the passive components

R_x[Ω]	4697.5
R_p[Ω]	1190.9
R_0[Ω]	75
C[μF]	2.284
τ[ms]	2.89

The output resistance of the port R_0 in Table II was measured indirectly by measuring the voltage drop of a resistive divider. The divider was formed by the output port resistance and resistor with nominal value of 470Ω. The time constant given in Table II was calculated as

$$\tau = (R_0 + R_p)C \tag{7}$$

The measurements of the time interval t_x were repeated 100 times for each charging time interval (3τ÷10τ). The standard deviation, the average and the relative error was then calculated for each set of the measurements. The average of the measurements with charging time 10τ was taken as a true value for calculation of the time interval relative error. The results are given in Table III.

Table III
Standard deviation, average and relative error of the time interval measurements

	σ[μs]	$t_{x\text{-}av}$[ms]	Δt_x[%]
3τ	0.8	12.58	3.6
4τ	0.6	12.88	1.3
5τ	0.9	12.99	0.5
6τ	0.8	13.03	0.1
7τ	0.5	13.05	0.04
8τ	0.9	13.06	0.002
9τ	0.4	13.06	0.0005
10τ	0.5	13.06	≈0

The averages of the time interval measurements for different charging time intervals are graphically shown in Fig.4.

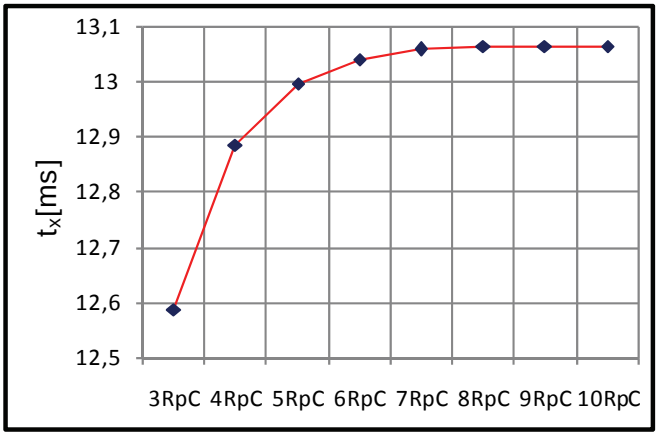


Fig. 4. Time interval average for different charging time

From the results reported in Table III it can be seen that the standard deviation of the measurements was nearly equal for all set of measurements. This was expected having in mind that the noise influence is the same in all measurements. To minimize the differences caused by program related errors, the program algorithm during the measurements was not changed.

The averages of the time interval measurements increased with increasing of the charging time interval. However, for charging interval higher than 8τ the average of the measurements was nearly constant (Fig.4). This is also confirmed with the relative error which is very low (0.002%) for charging time interval of 8τ. Having in mind that the maximal resolution that has been achieved with the direct sensor-microcontroller interface is around 13 bits [8] we can say that charging time interval of 10τ will be always satisfactory. However, for lower resolution and higher speed of the measurements lower charging time intervals should be used. In this way the best speed/resolution trade-off can be achieved.

The comparison of the theoretical (Table I) and the experimental (Table III) results for the relative error of the time interval measurements are shown on Fig.5.

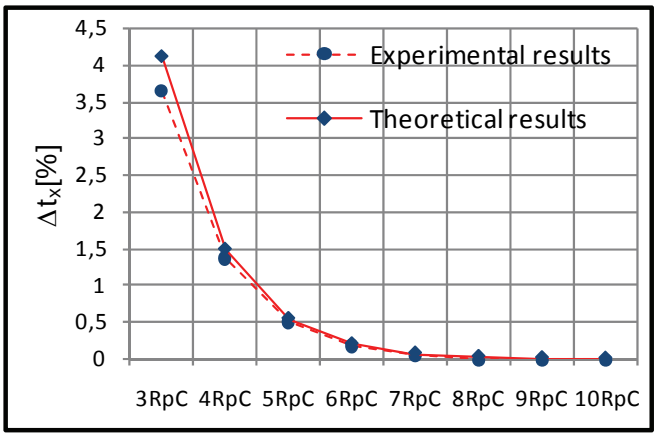


Fig. 5. Comparison of theoretical and experimental results

From the results reported in Fig.5 it can be seen that the theoretical and the experimental results fit very well. The small difference is due to deviations of the output resistance of the microcontroller pin. The applied measurement procedure for determination of this resistance is not very accurate.

The relative error of the time interval measurements with charging time interval 3τ was around 4%. This relative error is very high and such value for the charging time interval cannot be used. When charging interval 5τ was used, the relative error decreased to nearly 0.5%. This charging interval can be applied for measurements where the desired resolution is 6÷7 bits. The relative error of the measurements for charging time intervals over 9τ were very low and the average of the measurements were almost constant. Here, the capacitor voltage at the end of the phase1 (Fig.2) reach 99.99% of the voltage V_{dd}. Therefore charging time intervals higher than 9÷10τ carry no more benefit and should not be used.

V. Conclusion

The direct sensor-microcontroller interface can be used for simple and cost-effective measurements of the resistive and capacitive modulating sensors. The measurement contains two phases, charging and discharging phase. In this paper analysis for the optimal charging time interval that results in best speed/resolution trade-off were presented.

In the paper theoretical and experimental analysis for different charging time intervals from 3τ÷10τ were preformed. The experiments were realized by using PIC16F877 microcontroller. The falling edge of the input signal was registered with the RB0/INT Schmitt trigger pin. The discharging time was measured by using 16-bit timer - Timer1.

The theoretical and the experimental analysis fitted very well. The relative error of the time interval measurements with charging time interval of 3τ was very high and such value for the charging time interval cannot be used. The relative error of the measurements for charging time intervals over 9τ were very low and the average of the measurements remain almost constant. Therefore charging time intervals higher than 9÷10τ carry no more benefit and should not be used.

For measuring systems where the speed of the measurement is not very important, charging time interval of 10τ should be used. For higher speed measurements the optimal charging time interval depends on the desired resolution. The duration of this interval must not result in relative error of the measurements higher than 0.5LSB.

References

[1] D. Cox, “Implementing ohmmeter/temperature sensor”, AN512 Microchip Technology Inc., 1997.

[2] L. Bierl, “Precise measurements with the MSP430”, Report Texas Instrument, 1996.

[3] F. Reverter, R. P. Areny, “Direct Sensor to Microcontroller Interface Circuits”, Marcombo, Barselona, 2005.

[4] A. Custodio, R. P. Areny, R. Bragos, “Error Analysis and Reduction in Simple Sensor-Microcontroller Interface”, IEEE Trans. Instrumentation and Measurement, Vol.50, No.6, Dec. 2001.

[5] Z. Kokolanski, C. Gavrovski, “Improving the Uncertainty of Reconfigurable Microprocessor-Based Measuring Systems”, Master Thesis, Faculty of Electrical Engineering and Information Technologies, Skopje, Mar. 2010.

[6] F. Reverter, M. Gasulla, R. P. Areny, “Analysis of Power-Supply Interference Effects on Direct Sensor-to-Microcontroller Interfaces”, IEEE Trans. Instr. and Meas., vol.56, no.1, Feb. 2007.

[7] Microchip Technology Inc., “PIC16F8XX Data Sheet - 28/40-Pin 8-Bit CMOS FLASH Microcontrollers”, DS30292C, 2001.

[8] F. Reverter, R. P. Areny, “Effective Number of Resolution Bits in Direct Sensor-to-Microcontroller Interfaces”, Meas. Sci. and Tech. 15, pp.2157-2162, 2004.

Applying GIS and SensorWeb for Monitoring of Fire in Protected Areas

Nataša Ž. Veljković[1], Sanja D. Bogdanović-Dinić[2], Dejan L. Pavlović[3] and Leonid V. Stoimenov[4]

***Abstract* – Natural disasters are more often main news topic. Environmental concern is present all over the world and the need for developing efficient monitoring systems for observing natural phenomena' behaviour and collecting field data is increasing. In this paper we will present GINISSENSE system for monitoring environmental phenomena and collecting information regarding features of interest. The system is based on GIS and Sensor Web concept, which is a type of sensor network adapted for environmental monitoring, and it is designed in accordance with Open Geospatial Consortium's (OGC) set of specifications named Sensor Web Enablement (SWE). For the purposes of better demonstration of GINISSENSE system's capabilities we will present the application of the system for monitoring wild fires parameters in protected areas and give a review of existing wild fires monitoring systems.**

***Keywords* – forest fire, sensor web, GIS, environmental protection.**

I. INTRODUCTION

Environmental protection is an important matter that covers a wide variety of issues, including clean air, water quality, chemical security, forest preservation, etc. The current state of the environment is the result of many pressures, which are mostly linked to human activity.

The forest fire protection is especially important in countries where the forest covers more than one third of the country territory. The percentage of forest cover in Serbia accounts for 34.0% [1]. According to a research conducted by the State Enterprise for Forest Management "Srbijašume" [1] in the period 1998-2008, 853 forest fires were registered with the burnt area of 16.357 hectares. The consequences of such events are enormous, both ecological and economical.

There are several methods for fire prediction and preservation of protected areas. They are mostly based on meteorological parameters and fire spread models. Using them together in an integrated and unified system, we can provide monitoring and control of forested regions based on decision-making models and meteorological analysis. Active care of the environment and improvement of its status can be incorporated with modern technologies for enabling continuous care for its state. Many existing systems, as those listed in [2], use new technologies and achievements in various scientific areas and attain high accomplishments. Although the performances of such systems are high, they are also very expensive and therefore hardly affordable.

In this paper we propose a GINISSENSE system for monitoring and control of the wild fire, particularly forest fire, in a way that reduces risks to wildlife and people. The system relies on web based technologies for data gathering and alerting, and meteorological models for calculating forest fire risk. The paper gives an overview of existing solutions along with comparison to our system, emphasizing its advantages.

[1]Nataša Ž. Veljković, [2]Sanja D. Bogdanović-Dinić and [4]Leonid V.Stoimenov are with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: {natasa.veljkovic, sanja.bogdanovic.dinic, leonid.stoimenov}@elfak.ni.ac.rs

[3]Dejan L. Pavlović is with JP "Nacionalni park Đerdap", Kralja Petra I 14a, 19220 Donji Milanovac, Serbia, E-mail: dejan@npdjerdap.org

II. SENSORWEB AND GIS

Key concepts in disaster response are level of preparedness, response times, sustaining the response and coordinating the response [3]. That is why Sensor Web is important as sensing tool, and GIS as coordination tool.

The Sensor Web concept was firstly described at NASA Jet Propulsion Laboratory, as the developmental collection of sensor pods that could be scattered over land or water areas or other regions of interest to gather data on spatial and temporal patterns of relatively slowly changing physical, chemical, or biological phenomena in those regions [4]. The popularity of the concept has been increasing over the years. Sensor Web must be differentiated with the term sensor network. While the term "sensor network" often stands for a computer-accessible network of many spatially distributed sensing devices, a "sensor web" refers to web-accessible sensor assets that can be discovered and accessed using standard web protocols, encodings, and interfaces [5]. Sensor Web represents an intelligent network of connected sensor devices, scattered over sensing area, capable of observing and measuring relevant features of observed phenomena, sharing information with other connected devices and sending data through the network. It can perform intelligent autonomous operations in uncertain environments, respond to changing environmental conditions, and carry out automated diagnosis and recovery through the World Wide Web, enabled by standard encodings and services [6].

Open Geospatial Consortium (OGC) has most contributed to a development of Sensor Web by establishing Sensor Web Enablement working group that has developed a set of specifications named *Sensor Web Enablement (SWE)* [5], which fully describe a Sensor Web system. SWE relays on Web Services and XML technologies and standards. Sensor Web, however, represents an excellent interface for accessing

sensors, but for the system to be complete, it is necessary to display and analyze received sensor data. The most logical approach to this problem is visual.

The Geographic Information System (GIS) as an information technology, which combines geographic locations of natural and artificial objects as well as other types of data in order to generate interactive visual maps and reports, is often used in combination with Sensor Web technology. If used as a data source in a Web-based GIS application, Sensor web gains a visual dimension. Information gained form measured values coming from the sensor is increased greatly by adding the GIS component that contributes to it in a geographical sense.

When monitoring the fire in protected areas, it is important to be aware of methods that can help in predicting the fire. It is essential to identify the physical phenomena that influence the creation of fire, as well as type of vegetation that influence its spread.

III. WILD FIRES MONITORING

A. Meteorological methods for predicting and preventing wild fires

In order to develop efficient systems for prediction and prevention of forest fires, it is necessary to understand parameters that are direct or indirect causes to these events. Even though most of the fires are caused by people, accidentally or by deliberate actions, these are the causes that couldn't be predicted by a software system. That is why we will be focusing on weather conditions and parameters that can be observed and processed.

Several different types of environmental parameters initiate the ignition of forest fires. Meteorological studies have shown that there are several weather indices that could be used for the purposes of calculating the possibility of such events. According to [7] these indices can be classified as structural and dynamic. Structural indices are also called long-term fire risk indices as they are calculated from variables that do not change over short time periods. Dynamic indices, on the other hand, are short-term indices that are used for computing the probability of fire ignition and spread. There are two types of dynamic indices: meteorological and vegetation. Meteorological indices are computed directly from meteorological factors, while vegetation index shows the state of vegetation provoked by those factors. For the purposes of this research, we will be focusing on meteorological factors and indices. Several different types of meteorological indices are in use nowadays for calculating forest fires risks [7]: Behave Index, The Canadian Fire Weather Index (FWI), Portuguese Index, Spanish ICONA Method, Sol Numerical Risk and Italian Fire Danger Index. They are all computed from temperature, humidity and other meteorological parameters' values and used for forest fires prediction. Most commonly used in Europe is The Canadian FWI.

Fire Weather Index is a numerical indicator that shows current state of a forested area and the potential of fire in that area [8]. It is a complex component, comprised of six subcomponents, which determine the effects of fuel moisture and winds on forest fires behavior, based on observations of weather conditions and parameters.

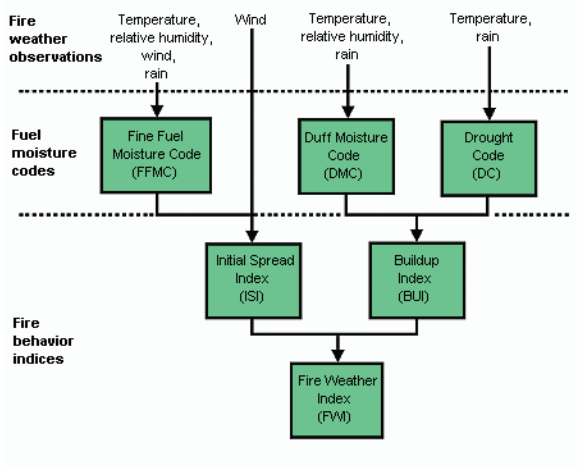


Figure 1. Canadian Forest Fire Weather Index System

FWI is a combination of Initial Spread Index, which represents the combined effect of wind speed and moisture content of fine fuel, and Buildup Index - BUI, which is the indicator of fuel moisture content. The calculation of these values is performed accordingly to WRF-NMM numerical model. Based on its value, there are five FWI categories: very low, low, moderate, high and extreme. Fig. 1 shows the structure of FWI and meteorological factors that are essential for its calculation.

TABLE I
FIRE WEATHER INDEX (RHSS)

	Date					
	1.07.2009.		02.08.2009.		23.08.2009.	
Weather station	**FWI**	**Risk**	**FWI**	**Risk**	**FWI**	**Risk**
Crni Vrh	0	Very low	14	Moderate	10	Moderate
Zlatibor	0	Very low	21	High	8	Moderate
Kopaonik	0	Very low	10	Moderate	5	Low
Nis	8	Moderate	30	High	24	High
Vranje	2	Very low	35	Extreme	34	Extreme
Novi Sad	0	Very low	33	Extreme	27	High

Republic Hydro-meteorological Service of Serbia (RHSS) [9] provides an online service for reviewing FWI values for weather stations in Serbia. Table I shows the RHSS report for three measurements taken during the summer in 2009. The report shows that there is increased risk of forests fires during summer months, especially in the beginning of august, when temperatures are very high, and precipitation and humidity very low. The monitoring of forested areas should be increased in that period.

B. Brief overview of wild fires monitoring systems

The fire prediction systems can be classified as follows:

- Terrestrial systems
- Air systems
- Satellite-based systems

Terrestrial systems use different types of sensors to detect fires: Meteorological sensors, video cameras, infrared cameras (IR) and LIDAR (Light Detecting and ranging) systems that can detect laser light reflected from smoke particles.

Air systems are used to cover large areas (using the aircraft), however a fundamental lack of such systems in terms of reliability is the presence of strong wind and low clouds.

Using satellite-based systems has some advantages and disadvantages. Satellites that are in geostationary orbits are highly effective for the detection of forest fires, but the disadvantages are reflected in the lag-time and in some cases the efficiency of detection of forest fires in the presence of clouds. A detailed overview of all types of prediction systems can be found in [2].

Here we would like to particularly single out one system called DEDICS (Distributed Environmental Disaster Information and Control Systems). DEDICS [10] is an integrated sensor network (of regional character), established for the purpose of improved management and control of arising emergencies in cases of natural disasters. The DEDICS has found the largest use in the early detection and control of the forest-fires spread. An implementation of the DEDICS system, for fire prediction, prevention and control, has been carried out at the University of Seville. The system has three main blocks: the *sensor interface block*, the *image processing block* and the *decision function block*. The sensor interface block is responsible for communication with the sensor stations. The *decision function block* combines the information from images, maps, meteorological data and the database with heuristic knowledge based on rules to produce a forest-fire possibility value. The output of the decision function block is a value in the range [0, 100] that represents the forest-fire possibility and the potential danger of the alarm. If the forest-fire possibility value is above an operator-selected alarm threshold, which is normally 0.5, the alarm will be considered as a forest fire. Otherwise, the alarm will be rejected.

IV. Applying GinisSense for wild fires monitoring

The most important drawbacks of existing automatic forest-fire detection systems are high false alarm rate, low functional coverage and high cost of infrared cameras and maintenance. GinisSense system for monitoring of forest fires is proposed as a simple solution that overcomes those obstacles and achieves high efficiency. GinisSense architecture is based on GIS and Sensor Web technologies, described in the second section of this paper, and it fully applies OGC SWE set of specifications and recommendations. Detailed overview of this architecture is given in [11]. The application of GinisSense architecture for monitoring of forest fires is presented in Fig. 2. The core of the system is comprised of a set of Web Services, responsible for data processing and communicating with sensors. They involve all of the system's functionalities and are hosted on a system server. This enables universal availability of system's capabilities and ease of access. Client applications, responsible for querying and visualizing of collected data, are installed in Control Centers and are used by operators for monitoring of observed parameters. Meteorological sensors are main data sources in the system, providing weather parameters measurements, but the system also has other sensing devices such as video cameras for real-time streaming of observed forests. Each sensing device is equipped with GPRS module and is therefore capable for sending its measurements to the server unit over the network. Sensors perform measurements at defined time intervals and send them to the system for further analyzes. They can communicate with each other and share relevant information regarding observed phenomena and for the purposes of performing some basic data pre-processing.

As elaborated in the third section, Fire Weather Index has proved to be a very efficient method for determining the risk of fire in forested areas. This index is calculated accordingly to meteorological data collected from weather stations, positioned in critical regions. Each weather station consists of temperature, wind, precipitation and humidity sensors, which collect data at specified intervals (e.g. every six hours) and send them through the network to the server component. The data is accepted by a server service; relevant information is extracted and then sent to a DMA component for further analyzes. DMA is an intelligent component, based on a set of reasoning rules, that uses collected meteorological data and terrain information in order to produce meaningful conclusions about the risk of forest fire in a certain area. Terrain data contains information about forest fuels and land topography, both extremely important for determining the forest fire risk. System databases store and preserve previous weather stations' measurements, as well as information regarding characteristics of terrains of forested areas. Combining stored data with actual measurements, DMA can generate accurate conclusions regarding the forest fire risk.

The Fire Weather Index accounts for about 90% of calculated risk value. This data can also be used for predicting fire propagation. Numerous fire spread models exist nowadays [12], and they can be involved in the system as a tool for predicting and preventing the spread of ignited fires. The forest fire risk in the GinisSense system is presented as an integer value, ranging from 0 to 100, where 0 means there is no risk, while 100 means there is extremely high risk of forest fire in observed area. The system will trigger the alarm in case the calculated risk is above 20. There are four alarm groups, based on priorities: low priority (for the risk between 21 and 40), medium priority (for the risk between 41 and 60), high priority (the risk is between 61 and 80) and ultra high priority group (the risk value is over 80). The alarm of low and medium priority will inform only the operator in Control Center about the critical area, while high and ultra high priority alarm will also inform fire fighters units responsible for fire issues in the area.

Beside weather stations, there are also wireless surveillance cameras set in forests, which are used for real-time monitoring of forested area. Operator at Control Center can observe terrain from each camera at any time. A separate service within the system is responsible for receiving a video stream from cameras.

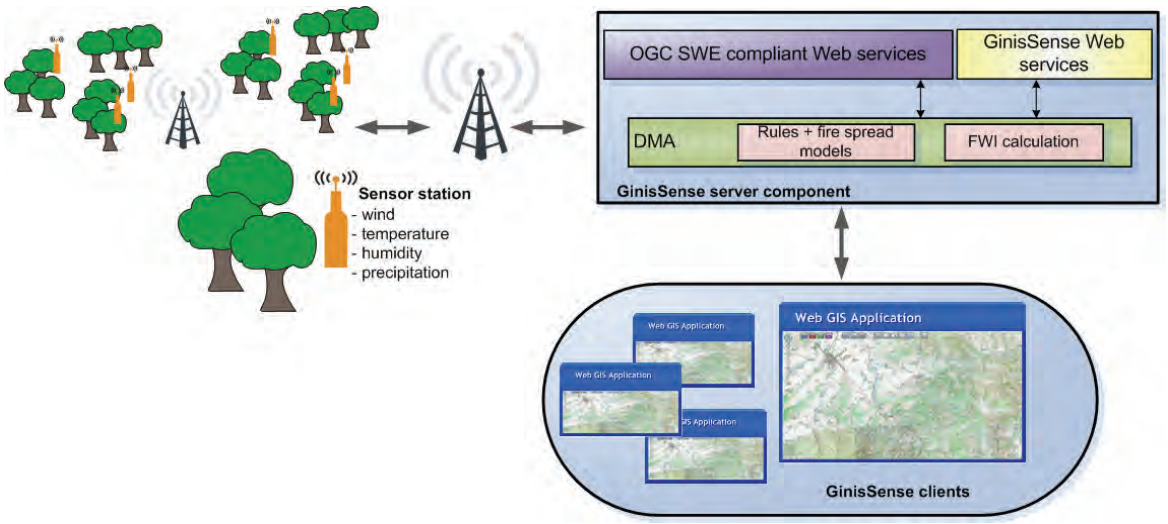


Figure 2. GINISSENSE system for monitoring of forest fires

The operator can trigger the alarm in case he notices a fire on a video, and inform fire fighters about the location of affected areas. Considering the system is intended to be used as a monitoring tool and the calculations based on FWI cannot confirm that there is fire, but only give the risk calculation of fire igniting in a forested area, cameras are the only devices that can confirm existence of fire.

GINISSENSE system has many advantages over existing similar solutions. It is based on popular technologies such as Sensor Web, Web services, XML and GIS, it uses standards for data encoding and communication and relies on sensing devices that come at relatively low cost and are very easy to install and use. The system as a whole provides large area coverage, due to sensors mobility and the possibility of wireless communication between sensing stations and the server component. However, the system cannot confirm if the fire is ignited, it can only give the calculation of the possibility for the fire to be ignited in a forested region.

V. CONCLUSION AND FURTHER WORK

Wild fires issue is increasingly present in Serbia. We have contributed to the solution of this issue, by designing and presenting a GINISSENSE system. The system relies on standards, is easy to implement and uses low cost sensing devices. Despite its sophistication, the system would cost no more than existing, less capable solutions and could reduce total operational costs by providing continual, automated analysis and monitoring. However, the system is only at its beginning and it is expected for this research to continue, evolve and overcome current shortages. Future plans include further Sensor Web exploration for applications in areas of environmental protection.

ACKNOWLEDGEMENT

Research presented in this paper were partially funded by the Ministry of Science of the Republic of Serbia, within the project No. 13003.

REFERENCES

[1] P. Aleksic, M. Krstic, G. Jancic, “Forest fires – ecological and economic problem in Serbia”, Botanica Serbica, vol. 33, no. 2, pp. 169 – 176, 2009.

[2] D. Pavlovic, “Role of integrated systems for early detection and prediction of fire spreading in protected forest areas”, YUINFO 2010. (in Serbian)

[3] A. Annoni, M. Atkinson, R. Denzer, L. Hecht, M. Millot, G. Pichler, A.-M. Sassen, M. Couturier, C. Alegre, H. Sassier, Y. Coene and P.G. Marchetti, "Towards an Open Disaster Risk Management Service Architecture for INSPIRE and GMES", Version No. 9., 2005.

[4] K.A. Delin, S.P. Jackson, and R.R. Some, “Sensor Webs,” NASA Tech Briefs, 23, pg. 80, 1999., http://www.nasatech.com/Briefs//Oct99/NPO20616.html

[5] M. Botts, G. Percival, C. Reed, and J. Davidson, "OGC Sensor Web Enablement: Overview and High Level Architecture," Open Geospatial Consortium Inc., OpenGIS White Paper OGC 07-165, 2007-12-28 2007.

[6] M. Botts, A. Robin, "Bringing the Sensor Web Together", Geosciences. pp. 46-53., October 2007. http://www.brgm.fr/dcenewsFile?ID=473.

[7] J. San-Miguel-Ayanz, P. Barbosa, G. Schmuck, G. Liberta, E. Schulte, ”Towards a coherent forest fire information system in Europe: The European Forest Fire Information System (EFFIS)”

[8] Canadian Forest Fire Weather Index (FWI) System, http://cwfis.cfs.nrcan.gc.ca/en_CA/background/summary/fwi

[9] Republic Hydrometeorological Service of Serbia, http://www.hidmet.gov.rs/index_eng.php

[10] J.R. Martinez-de Dios, B. C. Arrue and A. Ollero, “Distributed Intelligent Automatic Forest-Fire Detection System”, INNOCAP'99, 28th of 29th of April, Grenoble, 1999.

[11] N. Markovic, A. Stanimirovic, L. Stoimenov, “Sensor Web for River Water Pollution Monitoring and Alert System”, 12th AGILE International Conference on Geographic Information Science “Advances in GIScience”, Hannover, Germany, ISSN 2073-8013, 2009.

[12] Lj. Bodrozic, J. Marasovic, D. Stipanicev, “Fire modeling in forest fire management”, Proc.of Int. conf. CEEPUS Spring School - Engineering for the future, Kielce, Poland, pp.7-15, 2005.

Calibration of Electromagnetic Flowmeters Krohne DN 25

Jovan Atanasovski[1], Anastazija Sharevska[2] Dimitar Parnardzhiev[3], Danco Pendovski[4], Hristina Spasevska[5]

Abstract **- Metrology is an experimental science with the ultimate goal - to define more precisely the physical magnitudes such as length, weight, voltage and others. Assessment of physical magnitudes often leads to an analysis of the results of measurements made. In this paper is a calibration of the two electromagnetic flow meters, brand KROHNE IFM 5080K with 25 DN (nominal diameter of the tube which measured the flow of water, which is expressed in mm 27.5 mm). Reference standard (etalon) was PISTON PROVER OT 1500, owned by the Bureau of Metrology of the Republic of Macedonia. The aim is to show the process of calibration, metrological to compare two devices for measuring the flow of fluids from the same manufacturer in terms of the reference standard and to provide analysis of the results obtained. In the end the results of measurements which can be seen in the behavior of electromagnetic flow meters, minimum and maximum values of flow, the average relative error of ten measurements for five different values of flow and uncertainty of these measurements. Constant of the reference standard is confirmed by its internal calibration. Treceability be provided through a set of weights, thermometer and a device for monitoring the ambient conditions that have calibration certificates.**

Keywords **- calibration, etalon, uncertainty, standards, measurement points.**

I. INTRODUCTION

Calibration is a set of procedures which, under certain conditions, establishing relationship between the values of the size measure indicates that the measuring system or values that represent materialized extent or reference material, with appropriate values achieved benchmark-etalons. The result of calibration allows or assignment values of measuring size or determining what adjustments should be applied to indicating. Calibration can also determine other metrological properties such as the impact of influential sizes. The result of calibrations may be registered in a document sometimes called a calibration certificate or calibration report. In this calibration will compare the results, ie behavior of two el.-mag. flow meters from the same manufacturer Krohne.

(1) – (4) - Bureau of Metrology - Blvd. Jane Sandanski Nr.109a, Skopje, Macedonia,E-mail: jovan.atanasovski / anastazija.sarevska / dimitar.parnardziev / pendovski.danco@bom, gov.mk

(5) - Univerzity St. Cyril and Methodius, FEIT - st. Ruger Boskovic bb, Skopje, Macedonia, hristina@feit.ukim.edu.mk

1.1. Benchmark and calibration equipment

In performance of calibration using two electromagnetic flow meters, reference standard - calibrator PISTON PROVER OT 1500, pulse counter, instrument for measuring temperature, humidity and pressure of the environment - TESTO, thermometer for measuring water temperature. Metrological characteristics of these devices are given in the tables that follow.

1.1.1 Electromagnetic flow meters KROHNE IFM 5080K

Tab.1 Metrological characteristics of electromagnetic flow meters KROHNE

Serial no.	xx857	xx858
Pulse output	active (amplitude 15 V)	active (amplitude 15 V)
Output pulse rate	1 kHz for 10 m^3/h	1 kHz for 10 m^3/h
Time constant	ONLY I	ONLY I
Cut - off	OFF	OFF
GK	3.742	3.819
Power supply	(200-260) V AC/50 Hz	(200-260) V AC/50 Hz

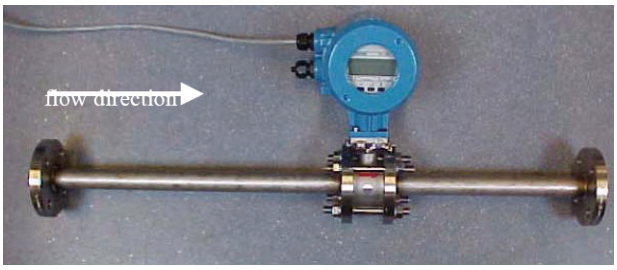


Fig.1 - KROHNE electromagnetic flow meters

1.1.2. Reference standard PISTON PROVER OT 1500

Reference etalon which is used belongs to the BOM with serial number 050418 produced by Trigas FI GmbH. Treceability is to PTB through Trigas FI GmbH. Calibration certificate is WD 5000-L-I17.

Tab.2 Technical Specification OT1500

Flow range	1.5 to 1500 l / min
Temperature range	0-38 oC
Operating pressure	5-12 bar
Linear encoder (length)	100 cm
Input frequency range	0-10 KHz
Analog Input (12 bit resolution)	voltage 0-5 V, current 4-20 mA
Repeatability	+ / -0.03%
Certificate of Calibration	DKD Germany

Fig.2 - PISTON PROVER OT 1500

1.1.2 Counter of electromagnetic pulse (pulses)

Counter who used belongs to the BOM produced by Electronic Control. Treceability is the UME by BOM Laboratory for time and frequency.
Double Chronometry is a technique used in positive displacement calibrators. Reduces uncertainty by ensuring that during the execution of the data targets (complete) pulses are counted and time for both, flow meter which are tested and the linear encoder is used in Calibrators. This technique eliminates the possibility of the involvement of unknown fractions of pulses in calibration interval therefore avoid the uncertainty that is potentially very large, particularly at low flows where pulses are small for both flow meter and linear encoder.

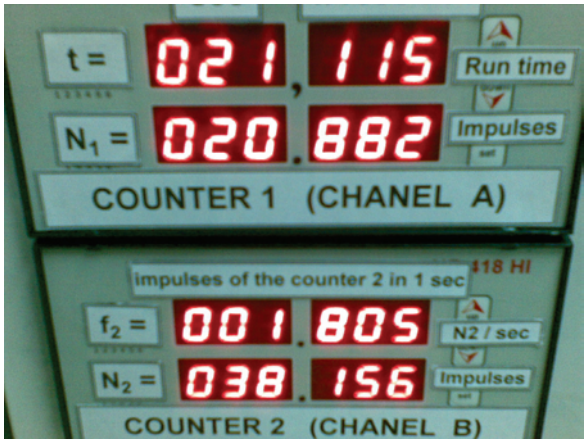


Fig.3 - Counter of electromagnetic impulses

1.1.3 Instruments to measure ambient conditions and watertemperature

Ambient conditions are recorded and recorded continuously with TESTO instrument for measuring temperature, humidity and pressure situated in the same laboratories as well as piston prover. Water temperature is measured with a mercury thermometer.

Tab.3 Technical Specification for thermometer with mercury

Measuring range	**-10...+50 °C**
Resolution	**0,1 °C**
Calibration Certificate	DKD-K-06701 Germany

Tab.4 Technical Specification TESTO
(digital instrument p, h and t)

Type	650
Calibration certificate	UME G3NM-0100 Turkey

II. Ambient Conditions

The range of variation of ambient conditions are recorded and corrections are applied to the measurements.

Reference conditions

- Medium for calibration is water with conductivity ≥ 200 mS / cm
- water temperature (20 ± 5 °C close to 20 °C possibly)
- water pressure downstream of flow meter (2 ± 1) bar
- Ambient temperature range: from 15 °C to 25 °C
- Ambient relative humidity range: from 45% to 75%
- Ambient range of atmospheric pressure: from 86 kPa to 106 kPa (0.86 bar to the 1.06 bar)

III. Measurement procedure

The flow is calculated by the following relation

$$Q = \frac{Vc}{t} = \frac{Vuut}{t} \qquad (1)$$

$$Vc = \frac{Nc}{Kc} \qquad (2)$$

In this relations, Q is the current flow which passes through flow meter (KROHNE), Nc is the number of pulses from the encoder, Kc is a constant of the Calibrator and Vc is the volume of water that passes through the reference standard OT 1500. Volume of water that passes through the Krohne-flow meter can be calculated according

$$Vuut = \frac{Nuut}{Kuut} \qquad (3)$$

For the electromagnetic flow meter with pulse output, Kuut is their output const, Nuut is a number of pulses from Krohne-flow meter.

Constant of calibrator Kc is performed under actual conditions of waterdraw (internal calibration OT1500). This constant must be reduced in normal (standard) conditions of 20 ^{o}C and 1 bar by applying appropriate corrections. The system of relations that define the constant Calibrator under normal conditions can be represented as follows

$$Kc = \left(\frac{N_E}{V_W}\right) * \left(\frac{C_T}{C_E} * C_P\right) \quad (4)$$

Where
C_E = [1 + (TE-20) * CEE] Correction for thermal expansion of the linear encoder

C_T = [1 - (Tt-20) CEt] Correction for thermal expansion of the cylinder Calibrators

Cp = [1 - (ΔP) * CEp] Correction for expansion of the pressure in the cylinder Calibrators

T_E temperature linear encoder at each water draw

Tt water temperature in the cylinder during each water draw

ΔP differential pressure (pressure in the cylinder - Ambient Pressure)

CE_E temperature coefficient of expansion of the linear encoder (published by the manufacturer, 1.1 * 10-5 1/0C for Mitutoyo AT11N)

CEt coefficient of thermal expansion of surface cylinder (3.46 * 10-5 1/0C on 316 steel)

CEp coefficient of surface expansion of the cylinder due to internal pressure. For 1500 this is OT 5.3 * 10-9 1/mbar

Kuut for the electromagnetic flow meter Krone is 360 pul/lit under Technical protocol

1KHz Puls exit 10 m^3/h flow.

Measurements were performed after a period of stabilization of 30min. According to the technical protocol measurements are performed in the following points:

1 m^3/h; 2.5 m^3/h; 5 m^3/h; 7.5 m^3/h; 10 m^3/h;

After collection of values for 10 measurements and that for all measurement points above the reference standard (etalon-calibrator) piston prover and electromagnetic flow meter (UUT) Krone, the volume of water that passed through both devices is calculated according to relations (2) and (3) and relative error as

$$E\% = \frac{\overline{Vuut} - \overline{Vc}}{\overline{Vc}} * 100 \quad (5)$$

Q_E value of flow is calculated with relation (1).

Average, the experimental standard deviation of a value and experimental standard deviation of the mean can be determined from the measurements are performed n = 10 times.

The mean value obtained from

$$\overline{V} = \frac{1}{n} \sum_{i=1}^{n} V_i \quad (6)$$

experimental standard deviation of a value

$$s\left(V_i\right) = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (V_i - \overline{V})^2} \quad (7)$$

experimental standard deviation of the mean is

$$s\left(\overline{V}\right) = \frac{s(Vi)}{\sqrt{n}} \quad (8)$$

Uncertainty is calculated by the following relations

$$u_A^2 = \frac{1}{N(N-1)} \sum_{\alpha=1}^{N} (E_\alpha - E)^2 \quad (9)$$

$$U_A = 2u_A \quad (10)$$

$$U^2 = U_A^2 + U_B^2 \quad (11)$$

Where:
N e number of repetitions (measurements)

E [%] is the average relative error of flow meter (REF. standard), ie Wed Mount. α given in relation (5), αby E N . . . , 1 =

UA [%] e type A is the uncertainty of the confidence level of 95% (k = 2)

UB [%] is the type B relative uncertainty of etalon value of the flow of level of confidence 95%

U is the uncertainty of the combined level of confidence 95%

Results of calibration of electromagnetic flow meters Krohne nr.xx857 and nr.xx858

Electromagnetic
flow meter br.xx857

Q[m3/h]	Q_E[L/h]	Q_{Emin}[L/h]	Q_{Emax}[L/h]	E[%]	N	V_E[m3]	T[°C]	p[bar]	U_A[%]	U_B[%]	U[%]
1	998.84	995.71	1001.48	-0.167	10	22.5240	24.5	2.5	0.014	0.049	0.051
2.5	2504.13	2495.73	2518.48	-0.163	10	25.5308	24.5	2.5	0.006	0.049	0.049
5	5003.53	4984.38	5016.96	-0.066	10	33.5786	24.5	2.5	0.014	0.049	0.051
7.5	7492.56	7472.94	7503.15	-0.068	10	23.2683	24.5	2.5	0.012	0.051	0.052
10	10048.66	10008.66	10086.53	0.083	10	20.0625	24.5	2.5	0.023	0.052	0.057

Electromagnetic
flow meter nr.xx858

Q[m3/h]	Q_E[L/h]	Q_{Emin}[L/h]	Q_{Emax}[L/h]	E[%]	N	V_E[m3]	T[°C]	p[bar]	U_A[%]	U_B[%]	U[%]
1	999.38	993.46	1007.62	-0.075	10	10.1975	24.5	2.5	0.018	0.049	0.052
2.5	2504.72	2496.70	2514.28	-0.097	10	22.8105	24.5	2.5	0.016	0.049	0.052
5	5016.48	5006.56	5050.32	-0.045	10	26.9630	24.5	2.5	0.012	0.049	0.050
7.5	7520.26	7482.02	7562.69	-0.064	10	21.3640	24.5	2.5	0.016	0.051	0.053
10	10009.41	9935.28	10046.08	-0.103	10	24.9470	24.5	2.5	0.017	0.052	0.055

Where:
Q_E is the average flow through the etalon, ie average value of Q_E
Q_{Emin} a minimal amount of flow through the benchmark, ie minimum $Q_{E\alpha}$
Q_{Emax} a maximum value of the flow through the benchmark, ie αmaximum of $Q_{E\alpha}$
Average relative error [%] of flow meter ie average value of Eα is given as

$$E_\alpha = \frac{V_{T\alpha} - V_{E\alpha}}{V_{E\alpha}}.100$$

where:
$V_{T\alpha}$ is the volume of water indicated by flow meter (Krohne)
$V_{E\alpha}$ is the volume of water indicated by the reference standard (OT1500)
N e number of repetitions (measurements)
V_E medium volume is indicated by the reference standard, ie average value from $V_{E\alpha}$
T [°C] average temperature of the water used in calibration.

U_A [%] e type A is the uncertainty of the confidence level of 95% (k = 2)
U_B [%] is the type B relative uncertainty of etalon
value of the flow of level of confidence 95%
U is the uncertainty of the combined level of confidence 95%.

IV. THANKS

Express gratitude for the assistance of Lazar Despotovski, M.Sc Danco Pendovski, as well as collaborators on this paper.

V. LITERATURE

[1] ISO 5168 Measurement of Fluid Flow-Estimation of uncertainty a flow-rate measurement
[2] GUM Guide to the Expression of uncertainty in a measurement
[3] T. Trigas Waterdraw constant determination of OT 1500
[4] OIML Measurement of water flow in closed conduits - Meters for cold portable water
[5] EA-4/02 European co-operation for accreditation

SESSION EM

Electrical Machines

MATLAB/Simulink analysis of no-load start-up regime on a three phase transformer, under different start-up conditions

Mihail Digalovski[1]

***Abstract* – In this paper is present analysis of the three phase power transformer with software package MATLAB/Simulink at various start-up regimes and conditions of the transformer. Inrush current that occur during the start-up regimes of a three phase transformer can reach extremely high values and cause many problems not only in the transformer but also in a power system. For investigation of this problem is developed simulation model of a three phase transformer in MATLAB/Simulink. In the model the effects of saturation, hysteresis and remanent magnetic flux are taken into account.**

***Keywords* – Three phase transformer, MATLAB/Simulink, inrush current, flux, simulation model, transient analysis.**

I. INTRODUCTION

Of all the transitional processes that take place in power transformers, after short circuited terminals on how most dangerous is transitional process of inclusion of the transformer in a power network in no-load regime. Phenomena are even more complex if the level of saturation of magnetic core is greater. The final expression for the change of flux, with inclusion of the network of the transformer in no-load regime is:

$$\Phi = e^{-\frac{t}{T_a}}\Phi_m \cos \omega t_0 - \Phi_m \cos \omega (t_0 + t) \pm \Phi_z e^{-\frac{t}{T_a}}$$

The initial value of start-up current depends on what point in time the primary coil joined the voltage, and the amplitude and polarity of magnetic flux in the magnetic core from the previous exclusion. In time of inclusion (t=0), depending on time t_0 which includes the transformer, there are six typical situations in which you can find a transformer.

1. Starting at t_0=0, zero voltage and without residual magnetism
2. Starting at t_0=0, zero voltage and with residual magnetism with the opposite polarity than that achieved with equivalent voltage conditions
3. Starting at t_0=0, zero voltage and with residual magnetism with the same polarity than that achieved with equivalent voltage conditions
4. Starting at t_0=T/4, maximal voltage and without residual magnetism
5. Starting at t_0=T/4, maximal voltage and with residual magnetism with the opposite polarity than that achieved with equivalent voltage conditions
6. Starting at t_0=T/4, maximal voltage and with residual magnetism with the same polarity than that achieved with equivalent voltage conditions

From the above mentioned cases, the most dangerous cases are under 1, 2 and 3. To obtain the start-up current, you need to know the magnetizing characteristic of the magnetic core B=f(H) and for each value of flux determine the value of the current contract it.

In three phase transformers process of inclusion in individual phases flows different because the magnetic fluxes are displaced 120 ˚ between.

II. OBJECT OF STUDY

The object of study is a three phase power transformer, oil immersed, with wye-wye connection. The rated data for this transformer are: S_n=125 kVA; U_1/U_2=10/0,4 kV; S_i=12 kV; I_1/I_2=7,225/180,63 A; u_{kn}=4 %; f_n=50 Hz; p=±2x2 %; Yy0.

III. MATLAB/SIMULINK MODEL OF THE TRANSFORMER

The short description of transformer model is given in brief. One three phase sinusoidal voltage source is used to supply the transformer model; their voltages are displaced for 120˚, and it represents balanced three-phase power system. The voltage value is setup to the amplitude (peak value) of the transformer primary voltage. Voltage sources are connected with transformer through three-phase circuit breaker which serves as turn-on or turn-off switch. By inclusion of the internal timer in the model, the switch will be used to set-up the time, i.e. the moment for starting the simulation; the only task is to set the breaker to closed position. All outputs and measured signals are converted in per units; this is more convenient when the comparative transient analysis is performed.

First, the primary currents are converted from amperes to per unit value, using the corresponding element in the model. To enable to observe their time changes, all the three signals are connected to the oscilloscope No. 1. In the model the other oscilloscope is used for observing fluxes in the primary windings of the transformer. For the fluxes, the voltage signals are used and converted into corresponding values expressed in per units, too.

The flux in each phase is obtained as the integral of the respective voltage. For this purpose, after the gain is installed in the respective integrator, the flux linkage of the correspondent phase is obtained as an output given in per units; the oscilloscope No. 2 is used to display the transient characteristics of fluxes. It is worth to emphasize another

contribution of the developed transformer simulation model. Namely, in the model is added a multiplexer, at which output are presented separately all three phase flux linkages at the same diagram. This is enabling to carry out the comparative analysis of the time change of the phase fluxes more obvious. Further on, it is possible to compare their values at any arbitrary selected instant of time and to analyze the status of the core for each particular phase.

The input values in the model are the voltage and the frequency rating. In the development of the simulation model it is comprehended to have available also the values of the transformer parameters: there are resistance and inductance of the both windings, as well as the impedance representing the magnetizing branch; usually, per unit values are in use. In a particular database is written the magnetizing curve and transformer parameters, which is obtained by the manufacturer. The curve is used when the transformer core is saturated, i.e. when the non-linear case is studied.

In addition the winding connection is defined; in this case, it is prescribed as $Y_g y_g 0$, meaning both windings are connected in star and grounded, while the voltages are in phase. Next to select is whether hysteresis is included in the simulations it means that respective iron losses have to be taken into consideration. Also, the initial values of fluxes (if any) are set-up. These quantities are depending to the conditions in the particular operating regime that is going to be analyzed.

The developed simulation model is presented in Fig. 1.

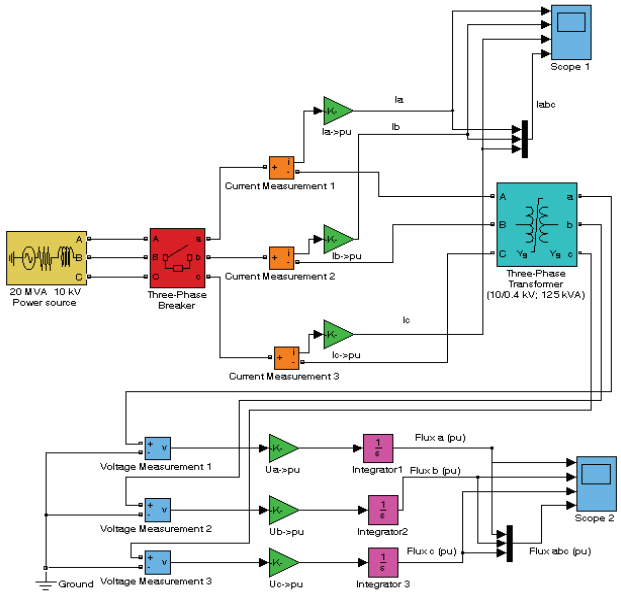


Fig.1. Simulation model of three-phase transformer of no-load regime

IV. MATLAB/SIMULINK RESULTS

A. Starting at t_0=0, zero voltage and without residual magnetism

First case study is a transient process at saturated magnetic core when the initial condition is U_B=0 and without residual magnetism. Transient characteristics of the currents and fluxes are presented in Fig.2 and Fig.3.

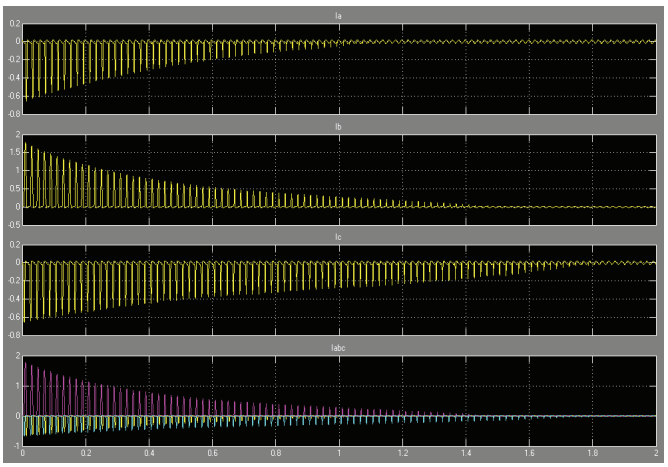


Fig. 2 Currents in the primary winding when U_B=0 and without residual magnetism

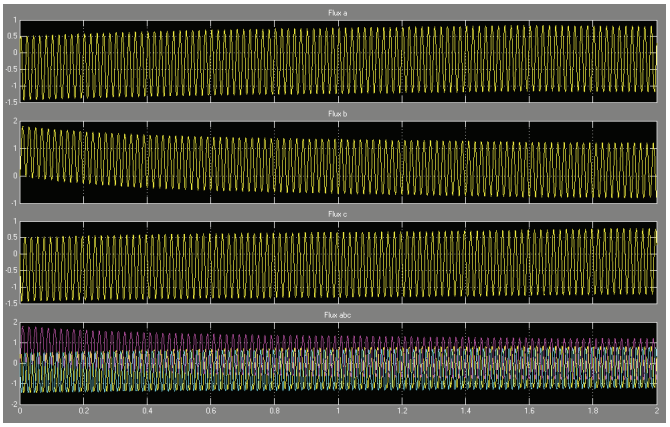


Fig. 3 Fluxes in the primary winding when U_B=0 and without residual magnetism

B. Starting at t_0=0, zero voltage and with residual magnetism with the opposite polarity than that achieved with equivalent voltage conditions

Second case is a transient process at saturated magnetic core when the initial condition is U_B=0 and with residual magnetism with the opposite polarity. Transient characteristics of the currents and fluxes are presented in Fig.4 and Fig.5.

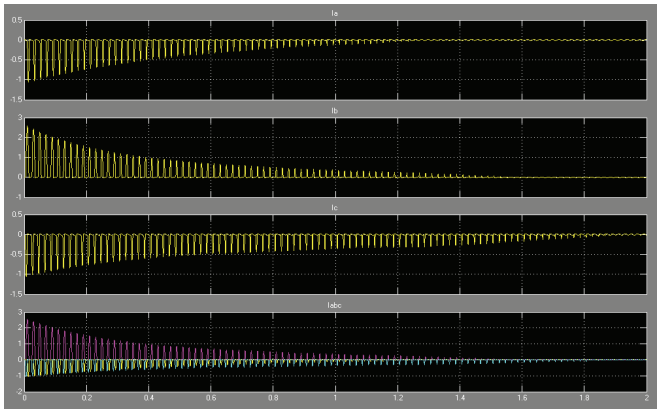


Fig. 4 Currents in the primary winding when U_B=0 and with residual magnetism with opposite polarity

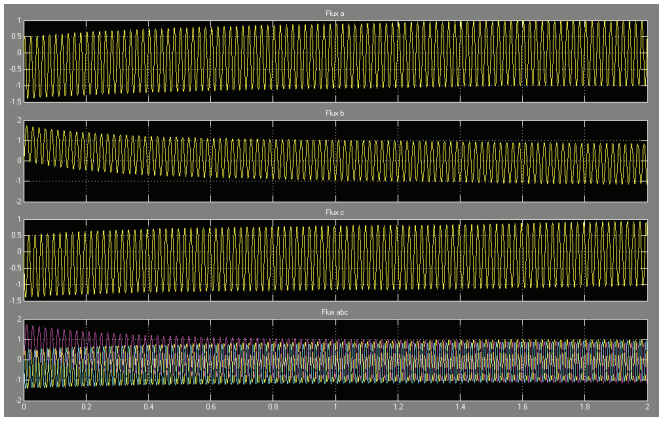


Fig. 5 Fluxes in the primary winding when U_B=0 and with residual magnetism with opposite polarity

C. Starting at t_0=0, zero voltage and with residual magnetism with the same polarity than that achieved with equivalent voltage conditions

Next case is a transient process at saturated magnetic core when the initial condition is U_B=0 and with residual magnetism with the same polarity. Transient characteristics of the currents and fluxes are presented in Fig.6 and Fig.7.

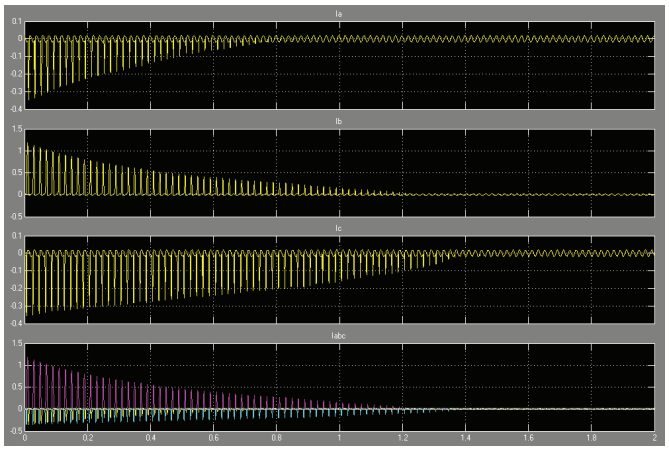


Fig. 6 Currents in the primary winding when U_B=0 and with residual magnetism with same polarity

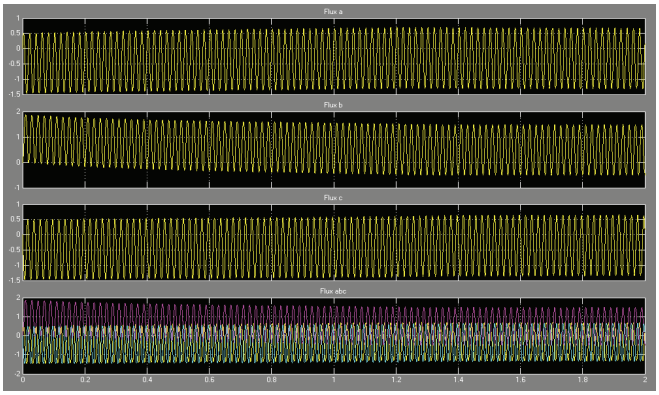


Fig. 7 Fluxes in the primary winding when U_B=0 and with residual magnetism with same polarity

D. Starting at t_0=T/4, maximal voltage and without residual magnetism

First case from the other three cases when U_B=U_{Bmax} is a transient process at saturated magnetic core without residual magnetism. Transient characteristics of the currents and fluxes are presented in Fig.8 and Fig.9.

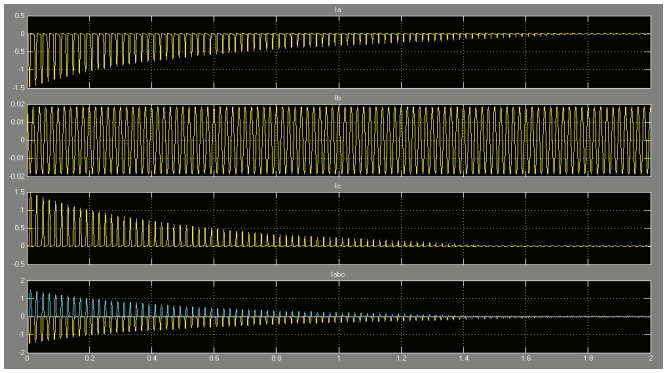


Fig. 8 Currents in the primary winding when U_B=U_{Bmax} and without residual magnetism

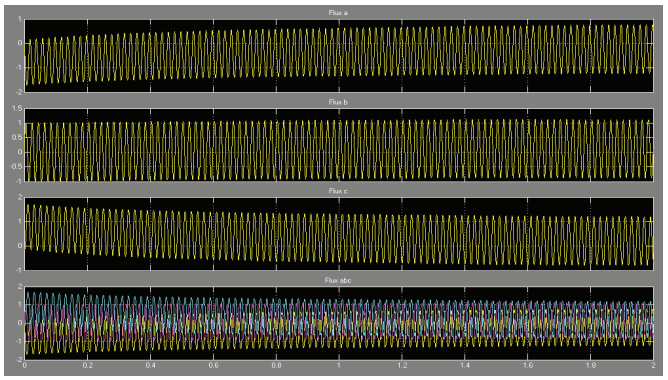


Fig. 9 Fluxes in the primary winding when U_B=U_{Bmax} and without residual magnetism

E. Starting at t_0=T/4, maximal voltage and with residual magnetism with the opposite polarity than that achieved with equivalent voltage conditions

The other case is a transient process at saturated magnetic core when the initial condition is U_B=U_{Bmax} and with residual magnetism with the opposite polarity. Transient characteristics of the currents and fluxes are presented in Fig.10 and Fig.11.

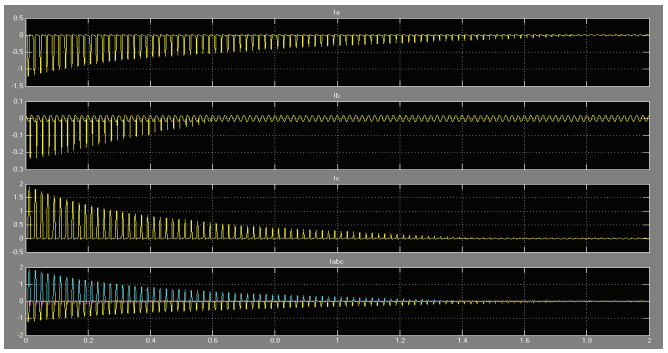


Fig. 10 Currents in the primary winding when U_B=U_{Bmax} and with residual magnetism with opposite polarity

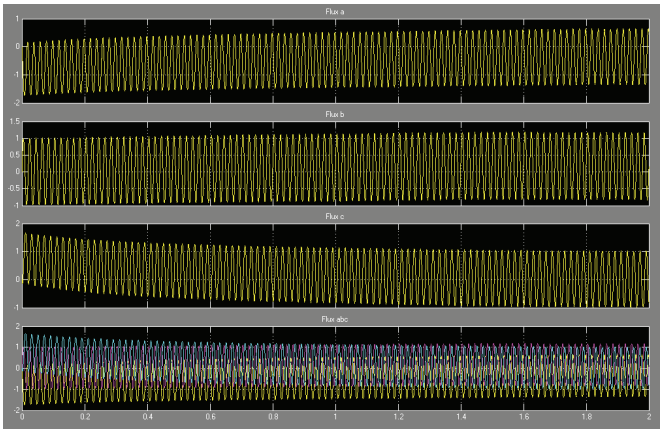


Fig. 11 Fluxes in the primary winding when $U_B=U_{Bmax}$ and with residual magnetism with opposite polarity

F. Starting at $t_0=T/4$, maximal voltage and with residual magnetism with the same polarity than that achieved with equivalent voltage conditions

The last case study is a transient process at saturated magnetic core when the initial condition is $U_B=U_{Bmax}$ and with residual magnetism with the same polarity. Transient characteristics of the currents and fluxes are presented in Fig.12 and Fig.13.

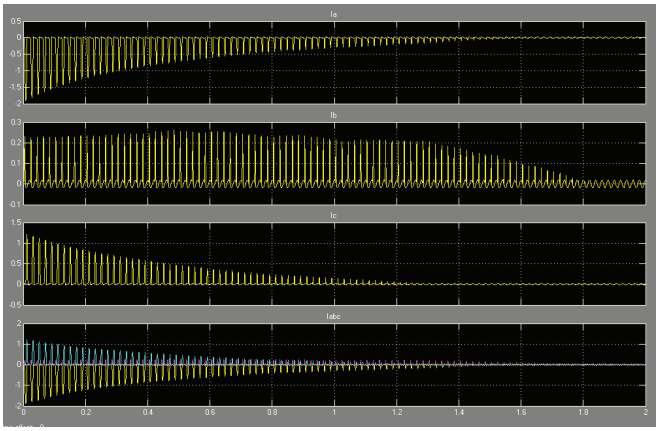


Fig. 12 Currents in the primary winding when $U_B=U_{Bmax}$ and with residual magnetism with same polarity

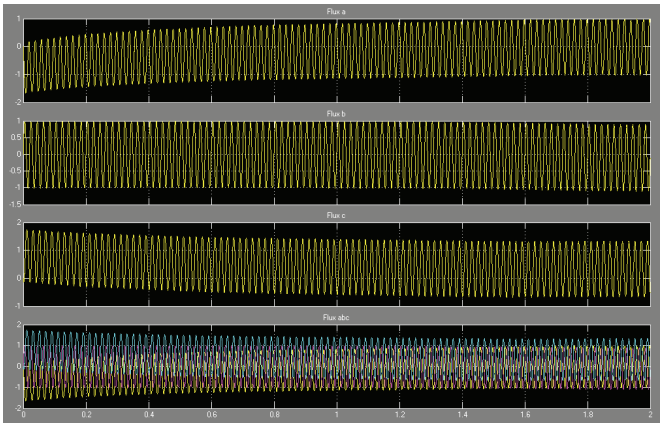


Fig. 13 Fluxes in the primary winding when $U_B=U_{Bmax}$ and with residual magnetism with same polarity

V. Conclusion

In the paper the most typical no-load characteristics at a start up of three phase transformer are presented and analyzed.

The mathematical model of the transformer is implemented in the MATLAB/Simulink environment and transient characteristics are obtained through the particular simulations. For the reference phase is adopted the phase B, which windings are placed on the middle leg of the transformer core. The analysis are carried out under the worst startup conditions, i.e. when the voltage in the analyzed phase passes through zero.

From the detailed observation of the transient characteristics, important conclusion upon the expected quantities of the voltage and currents in the transformer winding are derived. It was shown that under some circumstances it is possible reach that over-currents with values 2-3 times greater than rated value. It can appear incorrect operation of the differential protection of the transformer.

References

[1] L. Petkovska & all: Analyses of transient and nonsymmetrical regimes of electric machines from aspect of power system, Scientific project, pages 212, 1999, September, Skopje-Republic of Macedonia.

[2] L. Petkovska, G. Cvetkovski, V. Sarac. Different Aspects of Magnetic Field Computation in Electrical Machines, Book of Abstracts of 10thn International Symposium IGTE‘2002, p.p. 73; on CD pp. 1-6, Graz, Austria, 2002.

[3] Gregory Washington and Arun Rajagoplan: Simulink tutorial, Spring 2002.

[4] F. Mlakar, I. Kloar. Small Transformers and Inductors, Manual for Electrical Engineers, Vol. 8, Published by Elektrotehniski Vestnik, Ljubljana, Slovenia, 1970.

[5] Saffet Ayasun and Chika O. Nwankapa: Transformer tests using MATLAB/Simulink and their integration into undergraduate electric machinery courses, 2005.

[6] N. Chiesa, A. Avendano, H. K.Hoidalen, B. A. Mork, D. Ishchenko amd A. P. Kunze: On the ring-down transient of transformers, 2008.

[7] M. Popov, L. Van Der Sluis, G. C. Paap, and P. H. Schavemaker: On a hysteresis model for transient analysis, IEEE Power Engineering Review, vol. 20, no. 5, pp. 53 – 55, May 2000.

Critical Speed-Capacitance Requirements for Self-Excited Induction Generator

Milan M. Radić[1], Zoran P. Stajić[2]

Abstract – **A novel method for solving nonlinear algebraic equations describing behaviour of self-excited induction generator, based on the MATLAB Optimization Toolbox routines, is presented in paper. Using this approach, critical rotor speed and excitation capacitance requirements under different loading conditions have been considered. Accuracy of prediction has been improved by including core losses into the calculations.**

Keywords – **Induction generator, Self-excitation, MATLAB, fsolve.**

I. INTRODUCTION

The concept of self-excited induction generator (further: SEIG) has been known since 30's years of last century, when Basset and Potter in [1] reported that an induction machine, whose rotor was externally driven by some prime mover, could act as autonomous generator if capacitor bank of appropriate capacitance was connected across stator terminals. Except a few sporadic responses, there was no serious research of SEIG untill the 80's years of the 20st century. During that long period it has been usual to consider SEIG as something that is possible, but not practically acceptable, in first line due to it's poor regulation characteristics.

After the first signs of the global energetic crisis, scientists and researchers from all over the world increased their interest in SEIG, recognizing it's huge potentials for stand-alone power generation. Recently published research overviews [2], [3] indicate that serious work is still to be done in order to achieve acceptable level of performance to price ratio.

Difficulties related to solving of equations describing the SEIG behaviour were perhaps the greatest problem that researches had faced in earlier years. The main obstacle emerges from the fact that when machine operates as a SEIG, both voltage and frequency are independent variables. Different methods and formulas for approximate analysis have been proposed, but common point in the most of approaches has been to simplify equivalent circuit of the machine, neglecting core losses for example. Having in mind the physics of self-excitation process [4], [5] which implies that machine's operating point has to be in the saturated region of the magnetizing curve, it is clear that such simplifications affect accuracy of prediction. On the other hand, they appeared to be necessary, due to great mathematical difficulties that emerged even in trying to solve nonlinear equations obtained from simplified circuits.

[1]Milan M. Radić is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia
E-mail: milan.radic@elfak.ni.ac.rs

[2]Zoran P. Stajić is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia
E-mail: zoran.stajic@elfak.ni.ac.rs

There are numerous phenomena that can be analysed when research of SEIG is performed. Determination of critical speed or capacitance necessary for self-excitation is of great importance, and there was significant work in the past considering this type of problems [6], [7]. Besides mentioned simplifications in analysis, the usual approach in literature is to consider critical conditions for self-excitation of an unloaded machine, which is operating state with no practical significance. Also, the analysis is usually performed across one dimension, meaning that critical speed for exact capacitance or critical capacitance for exact speed is calculated.

In this paper, an unified approach is used to calculate contours defining critical speed-capacitance combinations under different values of symmetrical three-phase load connected to the stator terminals. Calculation has been simplified using a method for solving nonlinear algebraic equations based on the MATLAB Optimization Toolbox routines [8]. This method also allowed core losses to be taken into account easily.

II. MATHEMATICAL MODEL

Equivalent circuit of an SEIG with R-L load connected to stator terminals is shown on Fig. 1. Core losses are modeled by adding resistor R_c in parallel with magnetizing inductance X_m, as it was proposed in [8].

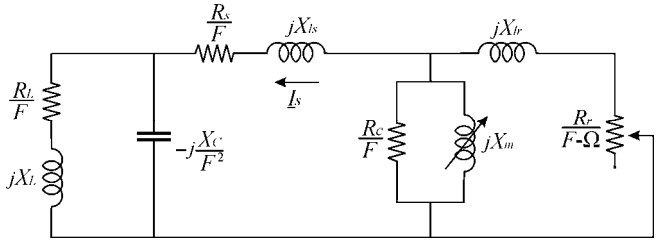


Fig.1. Per-phase equivalent circuit of a three-phase SEIG

The meaning of used symbols is:

- R_s, R_r, R_L - stator, rotor and load resistance;
- R_c - magnetizing resistance;
- X_m - magnetizing reactance at rated frequency ;
- X_{ls}, X_{lr}, X_L - stator leakage reactance, rotor leakage reactance and load reactance at rated frequency;
- X_C - capacitor reactance at rated frequency;
- $F = f_s / f_{sn}$ - actual to rated stator frequency ratio [p.u.];
- $\Omega = n/n_{sn}$ - actual rotor speed to rated synchronous speed ratio [p.u.].

Values of all machine parameters are expressed in ohms. Also, parameters describing the rotor part of the equivalent circuit are refered to stator, although there is no explicit notification neither in Fig. 1, nor in the text.

In general, both R_c and X_m are variables, depending on the actual conditions of operation, but in this analysis they will be considered as constants. Assumption that R_c has a constant value is questionable, because it neglects the fact that core losses depend on frequency. On the contrary, assumption that magnetizing reactance is constant, when critical conditions for self-excitation are considered, absolutely matches the truth.

From the theory of self-excitation it is known that machine's operating point have to be in saturated zone of the magnetizing curve, what can be expressed as $X_m < X_{muns}$, where X_{muns} is the value of magnetizing reactance in unsaturated state. In other words, limit of stability is determined by relation $X_m = X_{muns}$, which explains previous statement.

From the equivalent circuit shown on Fig. 1, using impedance loop method, can be written:

$$\underline{I}_s(\underline{Z}_s + \underline{Z}_1 + \underline{Z}_2) = \underline{I}_s \underline{Z}_{tot} = 0 \tag{1}$$

where

$$\underline{Z}_s = \frac{R_s}{F} + jX_{ls} \tag{2}$$

$$\underline{Z}_1 = (\frac{R_L}{F} + jX_L) \,\|\, \frac{-jX_C}{F^2} \tag{3}$$

$$\underline{Z}_2 = \frac{R_c}{F} \,\|\, jX_m \,\|\, (\frac{R_r}{F-\Omega} + jX_{lr}) \tag{4}$$

Since in steady-state stator current I_s differs from zero, it follows that $\underline{Z}_{tot}$ must be zero, and that can be written as:

$$\mathrm{Re}\{\underline{Z}_{tot}\} = 0 \tag{5}$$

$$\mathrm{Im}\{\underline{Z}_{tot}\} = 0 \tag{6}$$

Eqs. (5) and (6) are basic equations of an SEIG, and they have to be satisfied in any stable operating condition.

After some mathematical operations, Eqs. (5) and (6) can be written as:

$$\mathrm{Re}\{\underline{Z}_{tot}\} = \frac{\frac{R_L X_C^2}{F^3}}{R_L^2 + \left(FX_L - \frac{X_C}{F}\right)^2} + \frac{R_s}{F} + \frac{\frac{R_c X_m^2}{F}\left[\frac{R_r^2}{(F-\Omega)^2} + \frac{R_c R_r}{F(F-\Omega)} + X_{lr}^2\right]}{\frac{R_c^2}{F^2}\left[\frac{R_r^2}{(F-\Omega)^2} + (X_m + X_{lr})^2\right] + X_m^2\left[\frac{R_r^2}{(F-\Omega)^2} + \frac{2R_c R_r}{F(F-\Omega)} + X_{lr}^2\right]} = 0 \tag{7}$$

and

$$\mathrm{Im}\{\underline{Z}_{tot}\} = \frac{\frac{(X_L X_C - R_L^2)X_C}{F^2} + X_L^2 X_C}{R_L^2 + \left(FX_L - \frac{X_C}{F}\right)^2} + X_{ls} + \frac{\frac{R_c^2 X_m}{F^2}\left[\frac{R_r^2}{(F-\Omega)^2} + X_m X_{lr} + X_{lr}^2\right]}{\frac{R_c^2}{F^2}\left[\frac{R_r^2}{(F-\Omega)^2} + (X_m + X_{lr})^2\right] + X_m^2\left[\frac{R_r^2}{(F-\Omega)^2} + \frac{2R_c R_r}{F(F-\Omega)} + X_{lr}^2\right]} = 0 \tag{8}$$

If all parameters of the equivalent circuit, including reactance of the capacitor X_C are known, it is possible to determine relative rotor speed Ω and relative frequency F by solving Eqs. (7) and (8). However, it is necessary to use some numerical method, since Eqs. (7) and (8) are high-degree nonlinear equations. If any of standard methods is selected for this purpose (e.g. Newton-Raphson iterative method), it demands long-lasting mathematical transformations in order to get appropriate form of equations. Even if this work is properly done, serious problems related to convergency of the process will occur in the next phase, which makes such approach almost useless.

In this paper, system of two nonlinear equations is solved using user-friendly numerical routine from MATLAB Optimization Toolbox, named *fsolve*. This routine uses nonlinear least-squares algorithm that employs Levenberg-Marquardt method.

The first important feature of the *fsolve* routine is that further mathematical derivations are not needed, because it can easily operate with Eqs. (7) and (8). Also, it has extremely good convergence, and if no serious mistake is made during definition of starting vector values, zeros of the system are easily calculated.

During analysis whose results are presented here, reactance of capacitor has been varied across wide range, using program loop and considering that all other parameters in the equivalent circuit are constant. For each different value of capacitor reactance (i.e. capacitance), two different values of relative speed Ω and frequency F have been calculated. The first solution can be obtained if starting vector $[F \quad \Omega]^+$ is defined as $[0.98\Omega \quad \Omega]^+$, and the second one if starting assumption is approximatelly $[0.6\Omega \quad \Omega]^+$. As a result, this calculations finally give pairs of values (C, Ω) that define closed contour in the $C-\Omega$ plane. This contour presents critical speed-capacitance conditions that have to be fulfilled in order to sustain self-excitation under specific loading conditions.

III. Numerical Results

Presented approach has been used to calculate critical speed-capacitance contours for real three-phase squirrel-cage induction machine operated as an SEIG. Stator of the machine is star-connected, with rated values for motor mode of

operation $380V$, $50Hz$, $3.2A$, 2860 rpm. Parameters of the machine are given in Table I.

TABLE I
MACHINE PARAMETERS

R_s	4.05 Ω
R_r	2.75 Ω
X_{ls}	4.34 Ω
X_{lr}	2.77 Ω
X_{muns}	226 Ω
R_c	1200 Ω

Load impedance connected across stator terminals can be expressed as $\underline{Z}_L = Z_L e^{j\varphi_L}$, where φ_L is angle of impedance at rated frequency. Created mathematical model allows investigation of performance of an SEIG under any type of symmetrical R-L load. In order to identify influence of load impedance variation to the shape of critical $C-\Omega$ contours, in this analysis pure resistive load has been assumed. Results can be seen in Fig. 2. Log scale is used for the sake of better visibility. The actual load resistance used in calculation is assigned to it's resulting contour. Values of load resistance per phase are expressed in p.u., where $1\text{p.u.} = 68.5\Omega$.

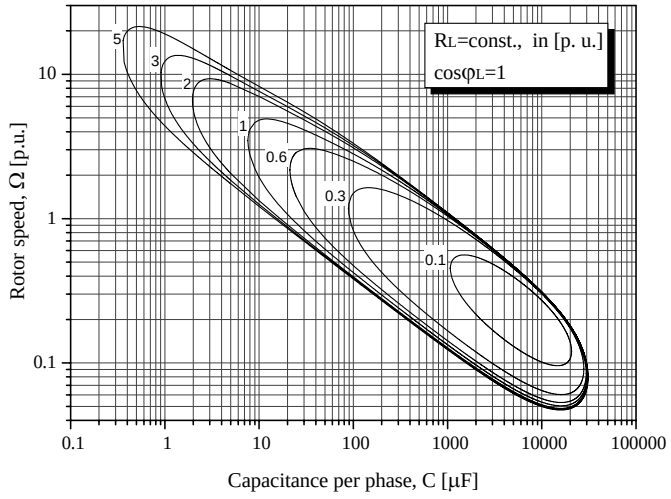


Fig. 2. Critical $C-\Omega$ contours for different load resistances R_L

From Fig. 2 it is obvious that if load resistance is set to the constant value, self-excitation can be achieved for many different combinations of rotor speed and excitation capacitors. Pairs of values (C,Ω) lie on the closed contour of irregular shape. Identification of such contours is of great importance because they represent critical conditions that have to be fulfilled in order to sustain self-excitation of symetrically loaded SEIG. For a specific value of load resistance R_L, point whose coordinates are (C,Ω) must be inside relevant contour. Only in this case SEIG will be able to generate power in steady state. Otherwise, if the point defined by (C,Ω) lies outside the critical contour, self-excitation will fail.

There is the point defined by minimum possible capacitance $C_{\min}$, and also the point defined by maximum possible capacitance $C_{\max}$, on each contour. At this points, self excitation can exist only if relative rotor speed has the exact value. For any other value of capacitance that lies between $C_{\min}$ and $C_{\max}$, relative rotor speed can take random value between boundry values defined by contour. Similiar analysis could be done if relative rotor speed is observed instead of capacitance.

Important conclusion that issues from the Fig. 2 is that dimensions of critical $C-\Omega$ contours expand as load resistance R_L takes higher values. The contour will be largest when generator is not loaded, since $R_L \to \infty$, but this operating state is not of great practical importance. This is the reason why critical contour of SEIG under no load has not been calculated. If Fig. 2 is carefully studied, it could be noticed that dimensions of contours rapidly shrink when load resistance falls under 1 [p.u.]. Intuitive conclusion can be made that there is certain minimum of load resistance $R_{L\min}$, at which critical contour does not exist anymore. Instead of contour, there is only point defined by the exact pair of values (C,Ω). It is obvious that load resistance must not take value lower than $R_{L\min}$, because SEIG will not be able to sustain self-excitation under such circumstances. This critical value of load resistance is function of machine's equivalent circuit parameters, and can not be changed.

However, previous considerations have mostly theoretical meaning. Practical application imposes limitations in rotor speed and capacitance of the capacitors connected to stator. It is not safe to run rotor of a conventional induction machine at speeds that are several times higher than rated synchronous speed, since mechanical stress and vibrations may cause serious damage. Also, from the aspect of economy, it is not reasonable to use large capacitor units, whose prices are exceeding price of induction machine. Having this in mind, it is enough to define only those parts of critical $C-\Omega$ contours that are situated in the region of practically acceptable speeds and capacitances, as shown on Fig. 3.

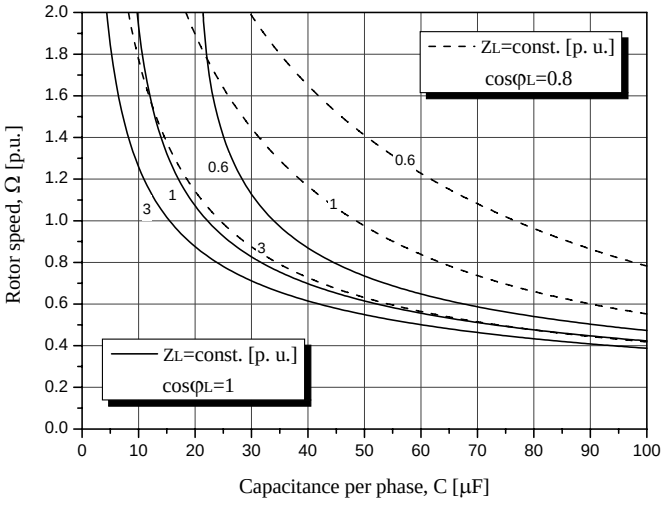


Fig. 3. Practically significant segments of $C-\Omega$ contours

Influence of combined resistive-inductive load can be observed from the Fig. 3. Curves calculated with assumption that load is purely resistive ($\cos\varphi_L = 1$) have been drawn with solid lines. Dashed lines represent critical $C - \Omega$ curves calculated for $\cos\varphi_L = 0.8$. It is obvious that presence of inductivity in load impedance leads to increased demands in applied capacitance, if rotor speed remains constant. In the case that capacitance remains constant, rotor has to run at higher speed to enable self-excitation, if combined resistive-inductive load is connected to the stator.

Curves presented in the Fig. 3 can also be used for quick estimation of possible consequences of desired loading, according to the actual operating state. For example, if machine whose parameters are given in Table I operates as an SEIG with excitation capacitance $C = 20\,\mu F$ and at rotor speed $\Omega = 1.2\,\text{p.u.}$, it will be possible to decrease load resistance from $R_{L1} = 3\,\text{p.u.}$ to $R_{L2} = 1\,\text{p.u.}$, without loss of self-excitation. The previous statement is clear if we notice that the point with coordinates $(C,\Omega) = (20\,\mu F, 1.2\,\text{p.u.})$ is placed above both critical $C - \Omega$ curves. On the contrary, if rotor speed is constant, but has the value $\Omega = 1\,\text{p.u.}$, described manipulation would lead to loss of self-excitation because the point $(C,\Omega) = (20\,\mu F, 1\,\text{p.u.})$ lies below critical curve for $R_{L1} = 1\,\text{p.u.}$

IV. Conclusions

Critical rotor speeds and excitation capacitances required for sustainable operation of an SEIG loaded with symmetrical three-phase load of arbitrary value and type, have been investigated in this paper. Mathematical difficulties that had been following any intention to solve nonlinear equations describing an SEIG in the past, have been succesfuly eliminated by using *fsolve* routine from MATLAB Optimization Toolbox. Based on this approach, full calculations have been performed, without need to simplify exact equivalent circuit of an SEIG by neglecting core losses.

It has been shown that, for arbitrary load impedance that is greater than minimum allowed value, exact closed contour in $C - \Omega$ plane can be defined. This contour can be considered as a set of critical rotor speed and capacitance combinations that allow operation of an SEIG at the limit of stability. Due to it's nature, it could be named as „critical $C - \Omega$ contour". For a given load impedance $\underline{Z}_L$, sustainable self-excitation will exist only if the point whose coordinates are (C,Ω) is placed inside the critical contour. Dimensions of critical contour expand as load impedance tends to get higher values. Shape of the contour is not uniform, and for the exact value of $|\underline{Z}_L|$ it will be affected by actual value of load power factor at rated frequency.

References

[1] E. D. Basset, F. M. Potter, „Capacitive excitation of induction generators", AIEE Trans., vol. 54, pp. 540-545, 1935.

[2] G. K. Singh, "Self-excited induction generator research – a survey", Electric Power Systems Research, vol. 69, pp. 107-114, 2004.

[3] R. C. Bansal, "Three-Phase Self-Excited Induction Generators: An Overview", IEEE Trans. Energy Conversion, vol. 20, no. 2, pp. 292-299, 2005.

[4] M. G. Say, "*Alternating Current Machines*", London, Pitman Publishing, 1977.

[5] I. Boldea, S. Nasar, *The Induction Machine Handbook*, Boca Raton - London - New York - Washington D.C, CRC Press, 2002.

[6] L. Wang, C. H. Lee, "A Novel Analysis on the Performance of an Isolated Self-Excited Induction Generator", IEEE Trans. Energy Conversion, vol. 12, no. 2, pp. 109-117, 1997.

[7] C. Chakraborty, S. N. Bhadra and A. K. Chattopadhyay, "Excitation Requirements for Stand-Alone Three-Phase Induction Generator", IEEE Trans. Energy Conversion, vol. 13, no. 4, pp. 358-365, 1998.

[8] M. H. Haque, "A Novel Method of Evaluating Performance Characteristics of a Self-Excited Induction Generator", IEEE Trans. Energy Conversion, vol. 24, no. 2, pp. 358-365, 2009.

Variable Speed Constant Capacitance Operation of Self-Excited Induction Generator

Milan M. Radić[1], Zoran P. Stajić[2]

Abstract – **This paper presents theoretical and experimental results of self-excited induction generator under varying rotor speed operation research. Three-phase squirrel-cage 1.5 kW induction machine, excited with symmetrical capacitor bank and loaded with symmetrical three-phase resistive load, was the subject of investigation. Experimentally obtained results have been compared with calculated performance curves and very good agreement between them has been achieved.**
Keywords – **Induction generator, Self-excitation, Variable speed.**

I. INTRODUCTION

Although the basics of self-excited induction generator (further: SEIG) theory have been established almost one century ago [1], SEIG still occupies attention of researches from all over the world. For a long period of time, it has been neglected as a device suitable for electric power generation, due to great difficulties related to voltage and frequency regulation.

Growing demands in electric power generated from renewable energy resources aroused researchers' interest in SEIG during last several decades. Three-phase squirrel-cage induction machine operated as a SEIG has many advantages compared to conventional electric generators. The most important are: low starting investments and maintenance costs, simplicity, robustness and extremely good behaviour in the case of sudden short circuit. Intensive research that is actual nowadays [2]-[6], combined with development of power electronics should overcome weakness of voltage and frequency regulation in the future. The great work is still to be done in order to expand commercial use of SEIGs [7].

One of SEIG's possible applications is stand-alone wind power generation on the sites with considerable wind potential [4], [5], [8]. In specific applications, it is even unnecessary to achieve full voltage and frequency regulation. For example, resistive consumers (boilers, heaters, etc.) are not significantly affected by actual voltage and frequency value. Such appliances could be supplied from unregulated SEIG driven by small wind turbine, whenever there is enough wind power on one side and need for energy on the other.

The motivation for work presented in this paper has been to identify how does variation of rotor speed affect voltage, frequency, stator current, generated power and shaft torque of the SEIG. Theoretical and experimental research was done assuming that SEIG was loaded with three-phase symmetrical resistive load and excited with symmetrical capacitor bank connected to the stator.

[1]Milan M. Radić is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia
E-mail: milan.radic@elfak.ni.ac.rs

[2]Zoran P. Stajić is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia
E-mail: zoran.stajic@elfak.ni.ac.rs

II. FORMULATION OF THE PROBLEM

When an induction machine operates as a SEIG, there is no external power grid that defines voltage and frequency on the stator terminals. Thus, both of them are unknown variables whose values change independently, being affected by rotor speed, capacitance of excitation capacitors and loading conditions. Saturation level of the magnetic circuit is also variable, which means that magnetizing inductance can not be considered as constant. In such circumstances, standard equivalent circuit of an induction machine is not suitable for analysis, and specific modifications have to be made. Several different variants of SEIG's equivalent circuit can be found in literature [9]-[12], but common point for all of them is that they neglect power losses in the magnetic core of the machine. Since SEIG always operates in the saturated region of the magnetizing curve, it is clear that such simplification can diminish accuracy of prediction. On the other hand, simplifications of this kind were necessary, due to great mathematical difficulties that follow attempts to solve high-degree nonlinear equations obtained from equivalent circuit.

Recently, Haque [13] proposed a new method for solving of these equations, based on MATLAB Optimization Toolbox routine named *fsolve*. In distinction from standard iterative methods that demand long-lasting and fatiguing mathematical transformations, this approach is very simple and fast. It also allows use of exact equivalent circuit that takes core losses into account. Such equivalent circuit is shown on Fig. 1.

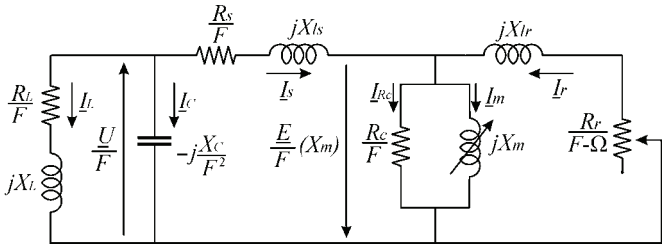


Fig. 1. Per-phase equivalent circuit of a three-phase SEIG

Symbols used in Fig. 1 have following meaning:

- R_L, X_L - load resistance and reactance;
- R_s, X_{ls} - stator resistance and leakage reactance;
- R_r, X_{lr} - rotor resistance and leakage reactance;
- R_c, X_m - magnetizing resistance and reactance;
- X_C - capacitor reactance;

U - stator voltage;

$\frac{E}{F}$ - normalized air-gap voltage;

F - per-unit value of stator frequency;

Ω - per-unit value of rotor speed.

Parameters and variables describing the rotor part of the equivalent circuit are referred to stator. All reactances in equivalent circuit are valid for rated frequency. Per-unit stator frequency F and per-unit rotor speed Ω are expressed in terms of rated frequency of the machine and rated synchronous speed, respectively. Magnetizing reactance at rated frequency is considered as variable, what is of essential importance in order to take variation of saturation level into account. Finally, normalized air-gap voltage is represented as function dependant on actual value of X_m. Identification of that function can be performed by processing results obtained from ideal no-load test at rated frequency and several different voltages.

Parameters of the equivalent circuit that are available for external control are load impedance at rated frequency $\underline{Z}_L = R_L + jX_L$ and capacitor reactance at rated frequency X_C. Rotor speed Ω is also defined externally, by the appropriate prime mover (e.g. wind turbine). When operates in steady state, SEIG can be treated as stable oscillator with stator current of constant effective value. Total impedance seen by stator current $\underline{I}_s$ can be described as

$$\underline{Z}_{tot} = \underline{Z}_s + \underline{Z}_1 + \underline{Z}_2 =$$

$$= \frac{R_s}{F} + jX_{ls} + \left(\frac{R_L}{F} + jX_L\right) \| \frac{-jX_C}{F^2} +$$

$$+ \frac{R_c}{F} \| jX_m \| \left(\frac{R_r}{F-\Omega} + jX_{lr}\right), \qquad (1)$$

and have to be equal with zero, otherwise self-excitation will not exist. After necessary mathematical transformations, total impedance can be separated into real and imaginary part, which leads to basic equations of a SEIG:

$$\mathrm{Re}\{\underline{Z}_{tot}\} = 0 \qquad (2)$$

$$\mathrm{Im}\{\underline{Z}_{tot}\} = 0\,. \qquad (3)$$

If all parameters of equivalent circuit are known, system defined by Eqs. (2) and (3) can be solved for desired rotor speed Ω. Calculated roots of the system, F and X_m, have to be real values and have to fulfill criterion

$$(F > 0) \wedge (0 < X_m < X_{muns})\,, \qquad (4)$$

where X_{muns} is magnetizing reactance of unsaturated machine at rated frequency. Otherwise, self-excitation is not possible for the selected combination of values $(\underline{Z}_L, X_C, \Omega)$.

When F and X_m are known, it is easy to solve equivalent circuit and to calculate all other quantities that are of interest. If Ω is being varied in small steps, starting with minimum value that allows self-excitation for assumed $\underline{Z}_L$ and X_C, simultaneous calculations of stator voltage U, stator current I_s, load power P_L and torque on the shaft T_{sh} lead to identification of desired performance curves. The upper limit of rotor speed, at which calculation should be ended is usually determined by mechanical limitations, or by the fact that stator voltage or stator current significantly exceeds rated value. It is important to notice that modified equivalent circuit shown on Fig. 1 can be successfully used for calculations of currents and voltages, but power and torque have to be calculated regarding the original circuit in which resistive elements are not divided by F.

III. Experimental Results

In order to validate explained theoretical approach, series of experiments was performed in laboratory, using a real three-phase squirrel cage, Y-connected induction machine as SEIG. Machine's rated values for motor mode of operation are: $1.5\,kW$, $380\,V$, $50\,Hz$, $3.2\,A$, 2860 rpm. This machine was externally driven by separately excited DC motor, what allowed arbitrary variation of rotor speed. On the stator side, machine was loaded by symmetrical, three-phase, Y-connected resistive load of constant value R_L. Symmetrical three-phase capacitor bank of constant per-phase capacitance C was connected in parallel with load and stator, in order to enable self-excitation. Varying rotor speed from the lowest value that allows existence of self-excitation to the upper limit that has been set to 120 % of rated synchronous speed, a large number of experimental points were recorded. Finally, recorded values were compared with calculated performance curves. Parameters of the machine that was used in experiments are given in Table I.

TABLE I
MACHINE PARAMETERS

R_s	4.05 Ω
R_r	2.75 Ω
X_{ls}	4.34 Ω
X_{lr}	2.77 Ω
X_{muns}	226 Ω
R_c	1200 Ω

From the results of ideal no-load test performed at rated frequency and with different voltages applied to the stator of the machine, pairs of values $(E/F, X_m)$ that define the shape of the $E/F = f(X_m)$ curve have been identified. Inside the acceptable range of X_m values, which is defined by Eq. (4), experimentally obtained pairs of values $(E/F, X_m)$ can be approximated by third-degree polynomial defined as:

$$\frac{E}{F} = A_0 + A_1 X_m + A_2 X_m^2 + A_3 X_m^3 \qquad (5)$$

Experimentally obtained points can be seen on Fig. 2.

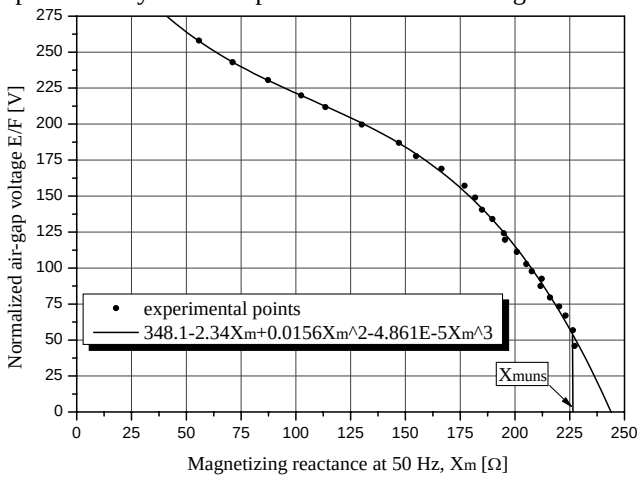


Fig. 2. Normalized air-gap voltage versus X_m

Experiments were performed for two different values of load resistance $R_{L1} = 517\,\Omega$ and $R_{L2} = 104\,\Omega$. Two different values of excitation capacitance per phase, $C_1 = 20\,\mu F$ and $C_2 = 30\,\mu F$, were also used. Experimentally obtained results are shown in Figs. 3-6, along with calculated curves. Since shaft torque has not been measured, Fig. 7 shows only calculated curves.

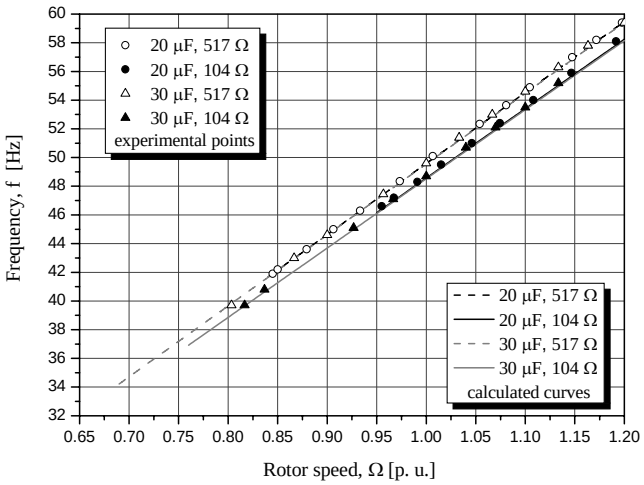


Fig. 3. Frequency f versus Ω, for $R_L = \text{const.}$, $C = \text{const.}$

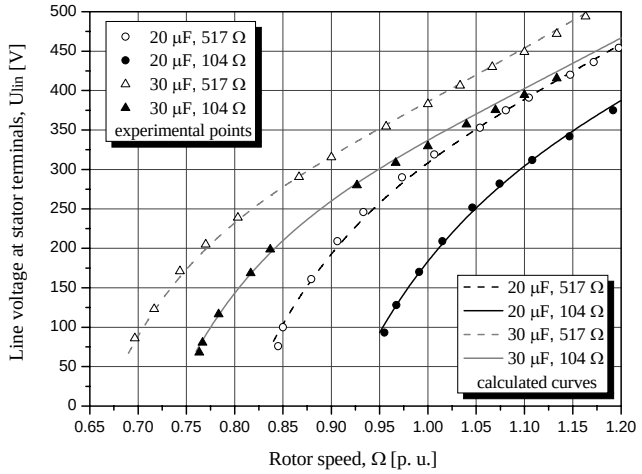


Fig. 4. Line voltage U_l versus Ω, for $R_L = \text{const.}$, $C = \text{const.}$

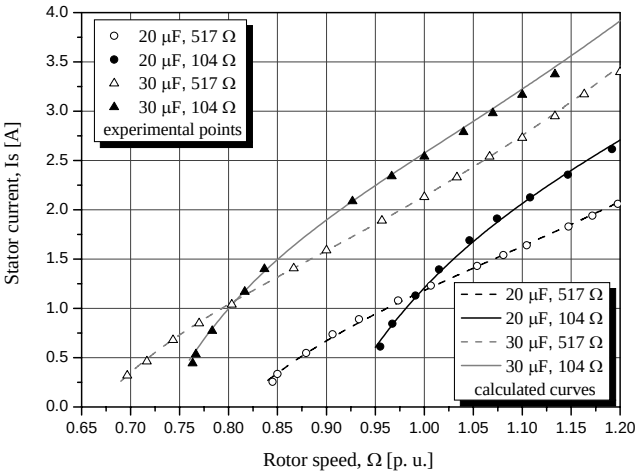


Fig. 5. Stator current I_s versus Ω, for $R_L = \text{const.}$, $C = \text{const.}$

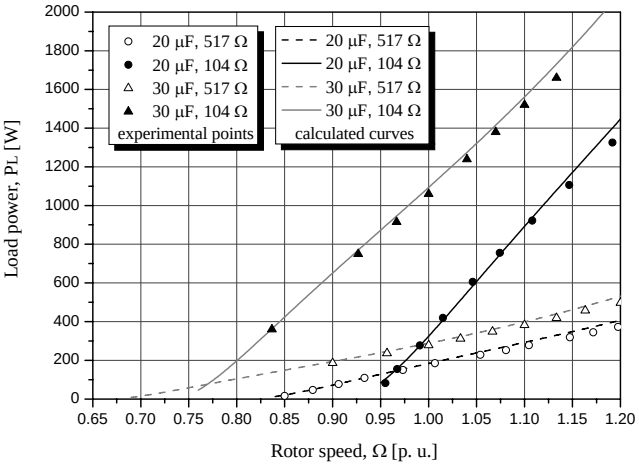


Fig. 6. Load power P_L versus Ω, for $R_L = \text{const.}$, $C = \text{const.}$

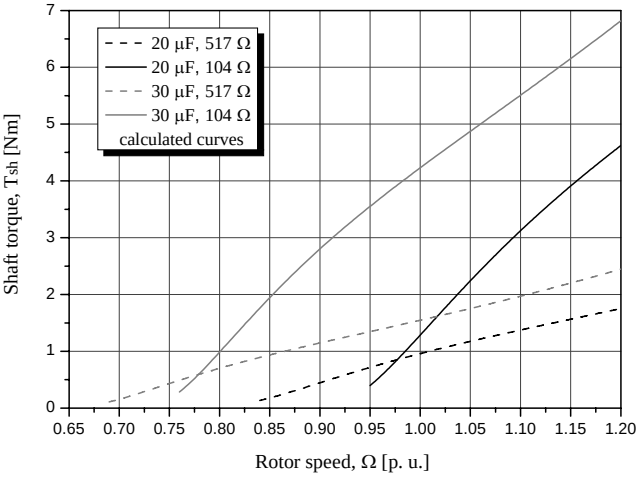


Fig. 7. Shaft torque T_{sh} versus Ω, for $R_L = \text{const.}$, $C = \text{const.}$

IV. Discussion

Very good agreement between experimentally measured values and calculated performance curves has been achieved. This fact confirms validity of exploited calculation method

and enables more serious research in future, when model will be expanded by inclusion of equations that describe a real turbine. Slight disagreement between calculated and measured values can be noticed at higher rotor speeds, which could be explained by the fact that all parameters of the equivalent circuit except magnetizing reactance have been considered as constant. In reality, that is not true because stator and rotor resistances depend on actual temperature of windings. Magnetizing resistance R_c is also depends on actual frequency and saturation level. Further, leakage reactances may be treated as functions of stator current [10]. Finally, tolerance of excitation capacitors that were used in experimental work is 5%. As a consequence, capacitor bank was not perfectly symmetrical.

Inside the normal operating region, where rotor speed is close to rated synchronous speed and excitation capacitance is reasonably small, dependence of SEIG's frequency upon rotor speed is almost perfectly linear. It is obvious from the Fig. 3 that value of excitation capacitance has no important influence to the frequency. As for the value of rotor resistance, situation is slightly different. It is obvious that lower value of R_L causes notable drop of stator frequency for the same rotor speed. This can be explained by the fact that at lower values of load resistance more power is transferred from rotor to stator, which causes greater difference between angular velocity of rotor and rotating magnetic field.

Generally spoken, higher values of load resistance and excitation capacitance allow sustain of self-excitation at lower speeds. After self-excitation has been established at minimum speed for selected combination of R_L and C, voltage, current, load power and shaft torque constantly grow as speed reaches higher values.

If both load resistance and excitation capacitance are high, voltage at the stator terminals can easily exceed rated value (Fig. 4). On the other hand, if high value of capacitance is combined with low load resistance, stator voltage will remain inside acceptable boundaries, but stator current exceeds rated value (Fig. 5). It is obvious that exact definition of practically acceptable operating area of a SEIG is the problem that can not be easily solved. Thus, it deserves serious consideration in future research.

Fig. 6 shows that for constant values of R_L and C, generated power is almost proportional to the rotor speed. The angle between Ω-axis and function $P_L = f(\Omega)$ is greater when applied capacitors have larger capacitance. Similar rule can also be observed on Fig. 7, which shows calculated performance curves $T_{sh} = f(\Omega)$. In distinction from $P_L = f(\Omega)$ curves, torque on the shaft of the machine becomes slightly saturated at higher speeds.

V. CONCLUSIONS

Specific operating mode of SEIG, characterized by constant load resistance and constant excitation capacitance, but variable rotor speed, has been considered in this paper. Such operation is very interesting from the aspect of exploitation in small, stand-alone wind power generation units. Identification of performance curves that show how does variation of rotor speed affect voltage, frequency, stator current, generated power and shaft torque of the SEIG was successful. Proposed mathematical approach was verified through experimental investigation. Comparison of calculated performance curves to measured data shows that good agreement between them has been achieved.

In future research, accuracy of prediction can be taken to a higher level by considering of several nonlinearities that are neglected at this moment. Also, mathematical model will be expanded by including of equations that describe actual wind turbine. That will be of essential importance in study of practical applications in real environment.

REFERENCES

[1] E. D. Basset, F. M. Potter, „Capacitive excitation of induction generators“, AIEE Trans., vol. 54, pp. 540-545, 1935.

[2] D. Joshi, K. S. Sandhu, M. K. Soni, “Constant Voltage Constant Frequency Operation for a Self-Excited Induction Generator”, *IEEE Trans. Energy Conversion*, Vol. 21, No. 1, pp. 228-234, 2006.

[3] B. Singh, S. S. Murthy and S. Gupta, “Analysis and Design of Electronic Load Controller for Self-Excited Induction Generators”, *IEEE Trans. Energy Conversion*, Vol. 21, No. 1, pp. 285-293, 2006.

[4] L. A. C. Lopes and R. G. Almeida, “Wind-Driven Self-Excited Induction Generator With Voltage and Frequency Regulated by a Reduced-Rating Voltage Source Inverter”, *IEEE Trans. Energy Conversion*, Vol. 21, No. 2, pp. 297-304, 2006.

[5] S. Singaravelu and S. Velusami, “Capacitive VAr requirements for wind driven self-excited induction generators”, *Energy Conversion and Management*, Vol. 48, pp. 1367-1382, 2007.

[6] J. M. Ramirez and E. Torres, “An Electronic Load Controller for the Self-Excited Induction Generator”, *IEEE Trans. Energy Conversion*, Vol. 22, No. 2, pp. 546-548, June 2007.

[7] R. C. Bansal, “Three-Phase Self-Excited Induction Generators: An Overview”, IEEE Trans. Energy Conversion, vol. 20, no. 2, pp. 292-299, 2005.

[8] G. Raina, O. P. Malik, „Wind energy conversion using a self excited induction generator“, *IEEE Trans. Power Apparatus and Systems*, Vol. 102. No. 12, pp. 3933-3936, 1983.

[9] M. G. Say, “*Alternating Current Machines*”, London, Pitman Publishing, 1977.

[10] I. Boldea, S. Nasar, *The Induction Machine Handbook*, Boca Raton - London - New York - Washington D.C, CRC Press, 2002.

[11] C. Chakraborty, S. N. Bhadra and A. K. Chattopadhyay, “Excitation Requirements for Stand-Alone Three-Phase Induction Generator”, IEEE Trans. Energy Conversion, vol. 13, no. 4, pp. 358-365, 1998.

[12] S. M. Alghuwainem, „Steady-state analysis of an isolated self-excited induction generator driven by regulated and unregulated turbine“, *IEEE Trans. Energy Conversion*, vol. 14, No. 3, pp. 718-723, 1999.

[13] M. H. Haque, “A Novel Method of Evaluating Performance Characteristics of a Self-Excited Induction Generator“, IEEE Trans. Energy Conversion, vol. 24, no. 2, pp. 358-365, 2009.

Electromagnetic and Thermal Analysis of Power Distribution Transformer with Method of Finite Elements

Blagoja S. Arapinoski[1] and Mirka I. Popnikolova Radevska[2]

***Abstract*: The paper will be presented the way of obtaining the electromagnetic characteristics of power distribution transformer type ETN 630, a product of EMO Ohrid – Republic of Macedonia, with software that performs simulation calculations based on the method of finite elements. The magnetic field is described by well known system of Maxwell's equations. Because the core materials in the transformer structure are nonlinearity, equations have to be solved iteratively and simultaneously. All computational results presented in the paper are obtained by using the Finite Element Method (FEM).**

Also over the same transformer made him a model for calculating the temperatures in the separate component parts, such as: Core, Coils, refrigeration fluid (the transformer oil) and the transformer container.

***Keywords* – Power transformer, ETN 630, Finite Element Method, Electromagnetic characteristics, magnetic field, flux density, calculation of the temperatures.**

I. INTRODUCTION

Object of study is a three phase core – type power transformer with rated data: Sn=630kVA, U_1 / U_2 =10/0.4 kV, f = 50 Hz, Yy0.

In The paper will be given a model of the transformer, its geometry and characteristics of the materials from which it is derived, and obtained distributions of electromagnetic fields in separate regions, at rated load. Also given will be the model for the calculation of temperatures and heat fields in separate regions of the transformer. As tool that made electromagnetic and temperatures analysis in this paper is applicable software package FEM 4.2 which is specialized software for analysis of this type. The algorithm which is basis of the software performs calculations using the method of finite elements.

Once a suitable model of the transformer may to various distinct approximations and test models of which the manufacturer may be offer the results of calculations that would improve some of the basic characteristics. Such analysis is applied in a contemporary production of electrical machinery and equipment for providing the wide range of optimizations without additional which cost would increase the cost of the device. Initial nominal sizes and data distribution transformer ETN 630, which is under consideration in this paper taken are from the test protocol have produced by this type transformer. In making transformer model of the software packages will be observed that all dimensions given in test protocol list of already produced transformer in order to obtain model that will later be analyzed and the results of calculations can be compared with those declared by the manufacturer.

[1]Blagoja S. Arapinoski is with the Faculty of Technical Sciences, Ivo.Lola .Ribar bb, 7000 Bitola, Macedonia,
E-mail: b.arapinoski@gmail.com

[2]Mirka I. Popnikolova Radevska is with the Faculty of Technical Sciences, Ivo .Lola .Ribar bb, 7000 Bitola, Macedonia, E-mail: mirka.radevska@uklo.edu.mk

II. ELECTROMAGNETIC ANALYSIS OF THE POWER TRANSFORMER ETN 630 IN FEM 4.2.

II. 1. MATHEMATICAL MODEL FOR ELECTROMAGNETIC ANALYSIS OF POWER TRANSFORMER ETN 630 IN FEM 4.2

FEM 4.2 software package is a specialized program for designing, modeling and analysis of electric machines and appliances. The program was developed by David Maker, and is the last version that stands out in the previous three versions with that, only the FEM 4.2 is possible despite the Electrostatic and electromagnetic analysis to makes electrical circuit and thermal analysis. FEM 4.2 has possibility to solve magnetic vector potential and consequently magnetic flux density by solving relevant set of Maxwell equations for magnetostatic case as well as for time harmonic case. In magnetostatic case field intensity **H** and flux density **B** must obey:

$$\nabla \mathbf{H} = \mathbf{J} \tag{1}$$

$$\nabla \mathbf{B} = 0 \tag{2}$$

subject to a constitute relation between **B** and **H** for each material:

$$\mathbf{B} = \mu \mathbf{H} = \frac{1}{v} \cdot \mathbf{H} \tag{3}$$

and for nonlinear material permeability μ is actually function of **B** .

FEM goes about finding a field that satisfies Eq. 1-Eq. 3 via a magnetic vector potential. Flux density is written in terms of the vector potential **A** , as:

$$\mathbf{B} = \nabla x \mathbf{A} \tag{4}$$

This definition of **B** always satisfies Eq. 2. Then Eq.1 can be redefined as:

$$\nabla x \left(\frac{1}{\mu(\mathbf{B})} \nabla \mathbf{A} \right) = \mathbf{J} \tag{5}$$

II. 2. CHARACTERISTICS OF THE MATERIAL COVERED IN THE MODEL OF THE TRANSFORMER ETN 630.

Magnetic circuit of transformer is made of cold rolled transformer sheet type M-27, its magnetic characteristics is given on Figure 1. Each sheet has a thickness of 0.635mm^2 filling factor is 0.98. Sheet are each insulated with paper thickness 0.25mm, and the total weight of the magnetic circuit is 701.56kg.

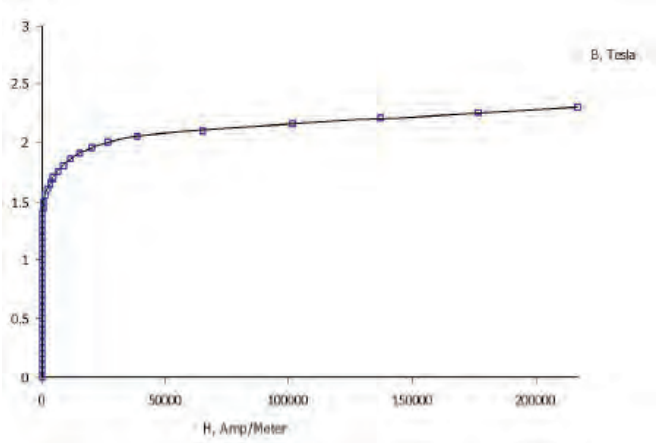


Fig. 1. Magnetic characteristic of transformer sheet M-27

The primary winding of high voltage on transformer is composed of eight layers (on Fig. 2. are presented as rectangular blocks) each of the layers has a 120 turn. Conductor which is made from a copper coil with a cross-section of 7.7931 mm^2. The modeling of the Coils current density recorded in the appropriate phase of the transformer, the value of current density on high voltage winding is J = ± 3.410 (MA/m^2).

The secondary coil of the low voltage is composed of two layers. The two layers are identical and consist of 11 turn. The value of current density on low voltage winding is J = ± 2.903 (MA/m^2).

This model is made ready for electromagnetic analysis in FEM 4.2. after being given, the appropriate boundary conditions.

II. 3. PROCESOR PART – GENERATING MASH OF FINITE ELEMENTS AND SOLOVING EQUATIONS WHICH DESCRIBE THE TRANSFORMER ETN 630

After defining the geometry of the transformer and the complete transfer the materials with their characteristics are accessing the second mode of FEM 4.2. This is the so-called solver mod that automatically draw the mesh of finite elements over which the geometry was introduced [2]. Check the validity is automatically. So if there is any part that is not defined network elements will not be final be drawn.

On Figure 2. is shown finite element mesh in cross section of the power three phase transformer. It contains 33,427 nodes and 66,780 triangular finite elements.

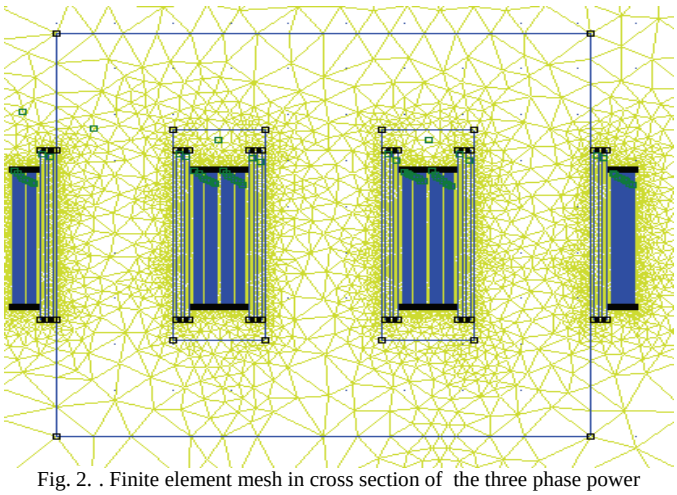


Fig. 2. . Finite element mesh in cross section of the three phase power transformer ETN 630.

The activation of fkern.exe, calculations are carried out equations that describe the transformer. Iterative calculations are made and more repetitions each subsequent faster than the previous, they may begin previous solution is approximately true. The actual analysis is carried out 12 repetitions that takes time is approximately 100 seconds on a PC Pentium 4th. Time that will make the calculation is depend with the number of nodes and triangular finite elements, and the dimensions of the triangle finite elements.

II.4. DESCRIPTION OF ELECTROAMGNETIC ANALYSIS AND PRESENTATION OF RESULTS

Starting with electromagnetic analysis on the model of the transformer, one of ways to get a good representation of the solution for the distribution of magnetic flux with finite element method, is drawing the lines of magnetic flux. They are lines that magnetic flux is established in the geometry of the transformer. When the magnetic flux lines are close to each other, then the flux density have greater value. The solution is initially shown with 19 lines of magnetic flux (which is also selected and it is possible to select and greater or lesser number of lines) - shown in Figure 3.

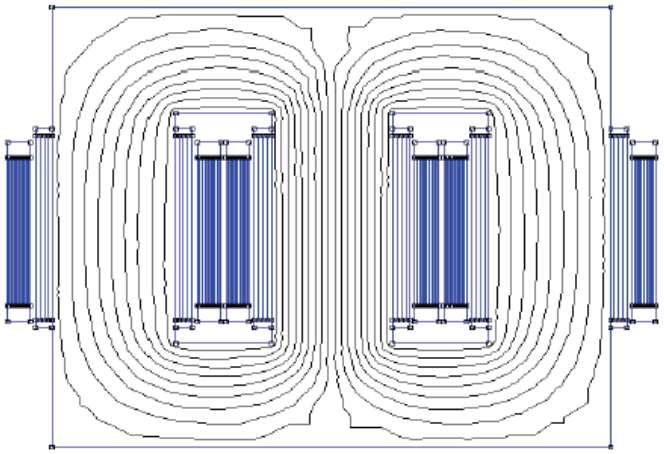


Fig.3. Distribution of magnetic flux in the transformer

In addition on figures 4 and 5, respectively show magnetic field density, and magnetic strength of magnetic field in the separate regions of transformer.

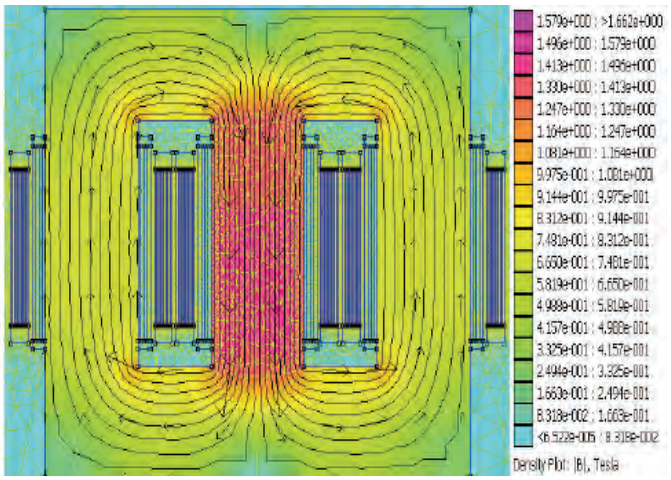


Fig. 4. Magnetic field density in in cross section of the transformer

It may be noted that in the mid-point value of the module magnetic induction is B=1.398 T, and the size of the magnitude of the magnetic field is H=472.535 A/m.

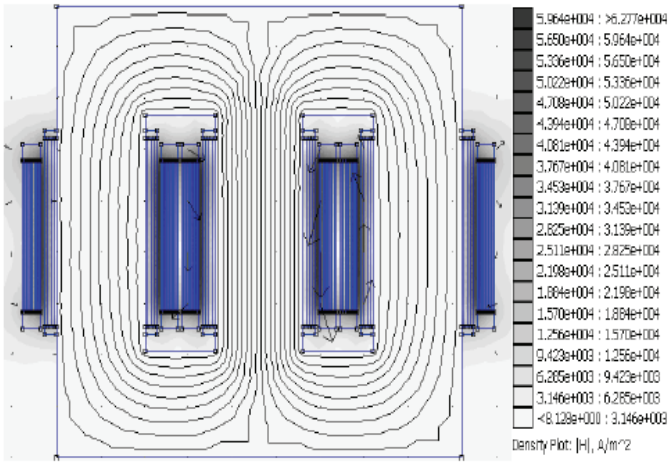


Fig. 5. Distribution of magnetic field in separate regions of the transformer

III. THERMAL ANALYSIS OF POWER TRANSFORMER TYPE ETN 630, IN FEM 4.2

Similarly as in electromagnetic calculations by the method of finite elements and the heat calculations, the system of differential equations are replaced by algebraic equations corresponding to a finite number of variables. The relevance of this method comes to terms with the sudden development of computer technology because only using fast computers it is possible to quickly resolve the large number of equations. Finite element method is very adaptable and powerful tool that enables you to obtain information on models with very complicated structures and arbitrary loads.

III.1 THERMAL ANALYSIS OF THE SIMULATION MODEL OF THE POWER TRANSFORMER ETN 630 WITH FEM 4.2

The temperature analysis is done in several consecutive steps:

1. Initial step in the temperature analysis is defining the geometry of the transformer. In the program package FEM 4.2 enabled directly draw the geometry to be analyzed. Because the transformer is made electromagnetic analysis can be used geometry of that analysis should be further noted that to enter transformatorkiot container with cooling ribs so as to meet the volume of fluid refrigeration. Fig. 6. given intersection pattern ETN transformer 630, the xy plane, ie the appearance of the geometry of the problem is analyzed.

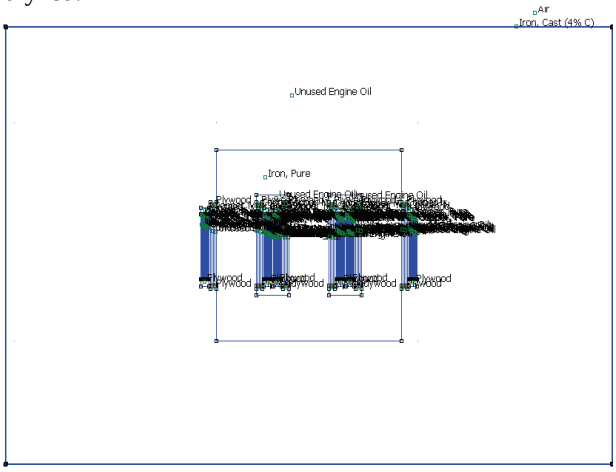


Fig. 6. Geometry of the model for thermal analysis in FEM 4.2

2. Defining characteristics of the material from which the transformer is made.

3. Forming a mesh of finite element - After defining the geometry of the model and characteristics of the materials from which the transformer is made, the following phase in which the model is divided into a number of triangular finite elements. Mesh of finite elements covering the transformer model contains 47,334 nodes and 94,604 triangular finite elements and is presented on figure 7.

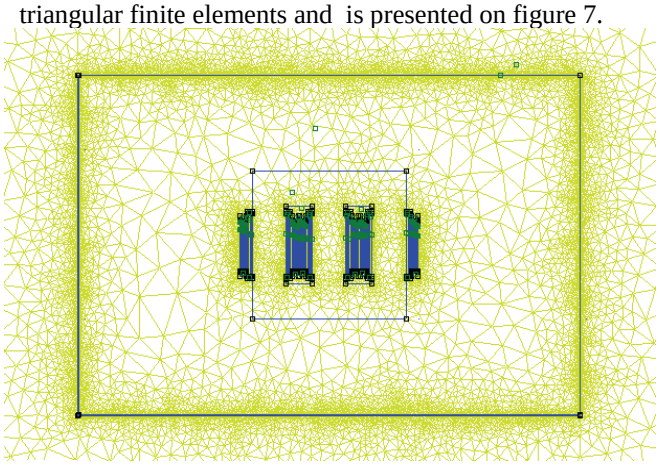


Fig. 7. Finite element mesh in cross section of the transformer ETN 630.

4. Defining the boundary conditions and loads - In order to obtain a solution to the algebraic system of equations necessary for the transformer model to define the boundary conditions and external loads. As is known, the power losses in the transformer primary source of heat that is generated as a thermal load is given heat rate generated per unit volume.
The total losses amounted to 7535.2(W) which are distributed as follows: magnetic core losses in 1300 (W), losses in high voltage coil is 2995.2 (W) and losses in low voltage coil is 3240 (W).
5. Solver, which generates results-Once a definition of all the sizes are approached to solving the problem of thermal analysis system composed of algebraic equations.
6. Post processing part - In this mode, the programming package of the solution obtained data are processed for further calculations of certain characteristic sizes. The results obtained are shown in different ways: graphically, in tables, visual, etc.
At Figure 8. heat distribution is shown in the pattern of distribution power transformer which is subject to this analysis. The legend on the right image shows the values of temperatures in separate regions, stating that temperature program calculated in Kelvin (K), which means that 273.15 (C) correspond to 0 (° C).

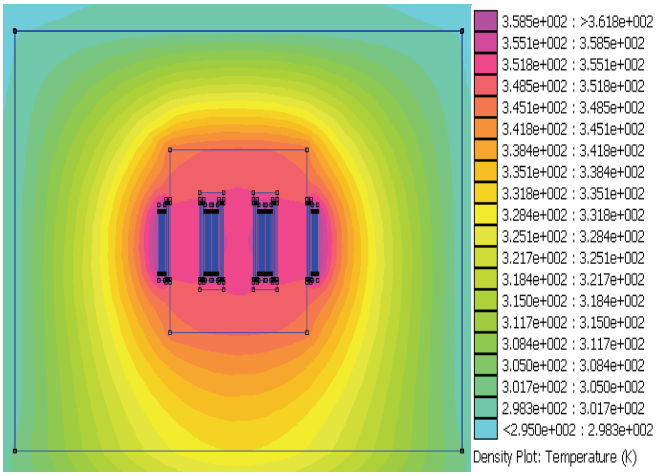


Fig. 8. Distribution of head field in separate regions of the transformer

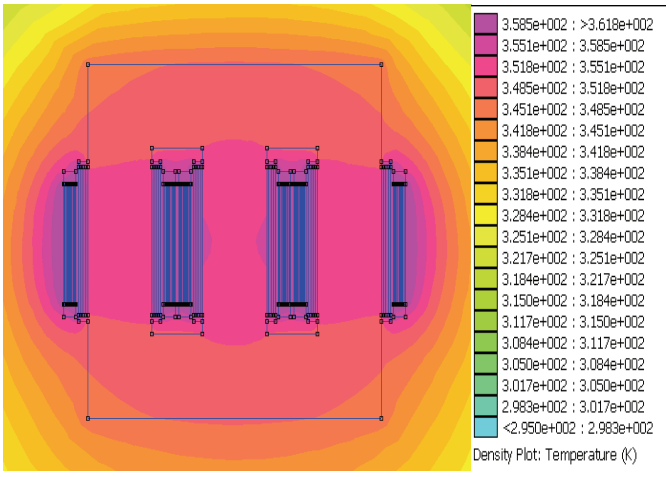


Fig. 9. Distribution of head field in separate regions of the transformer

For better visibility of the Figure 9. is given the distribution of temperatures in the magnetic circuit of the transformer and Coils because they are parts which are active and they are sources of heat.
From the analysis of concrete-made model of transformer, the nominal load, obtained the following average values of temperature: HVC 89,66 (° C), LVC 88,08 (° C), magnetic core 78,11 (° C), transformer oil 48 (° C), pot 32,651 (° C).

IV. Conclusion

From the performed research and made electromagnetic and heat analysis of the model of the energy transformer type ETN 630, may be adopted the following conclusions:
1. Made and honored model of the transformer in the paper is a real picture because the values of sizes that characterize the transformer declared in the test protocol sheet coincide with those obtained by using modified accounting software FEM
2. This type of analysis gives good results and offers the opportunity to conduct further investigations in the future which would be in favor of improving the characteristics of power transformers of this type without additional investment, since the use of the software, do not require additional costs.
3. From the analysis of thermal-made model of transformer in FEM 4.2. can be concluded that this method of calculating the temperatures in the transformer, contribute to optimize the solution for feasibility since the design phase can be made to control heating.

References

[1] M. Popnikolova Radevska: “Calculation of Electromechanical Characteristics on Overband Magnetic Separator with Finite Elements”, ICEST 2006, p.p. 367-370, Sofia, Bulgaria 2006.
[2] D. Meeker, “Finite Element Method Magnetics Version 4.2”, User’s Manual, 2006 .
[3] STEINERT Betriesanweisungen fur Uberbandmagnetscheider und Aushebemagnete, Technische Daten TD UME P, 2002.
[4] John Wiley & Sons, “ELECTROMAGNETIC DEVICES” New York-London-Sydney.
[5] M. R. Popnikolova, M. Cundev, L.Petkovska, “Nonlinear Electromagnetic Field Calculation in Solid Salient Poles synchronous Motor”, Proc. Of EPNC 96, Poznan, Poland, 1996.
[6] John Wiley & Sons, “ELECTROMAGNETIC DEVICES” New York-London-Sydney.
[7] Z. Radakovic, “Numerical Determination of Characteristic Temperatures in Directly Loaded Power Oil Transformer”
[8] CIGRE WG 12.09 (Thermal Aspect of Transformers), Survey of Power Transformer Overload field practics, CIGRE 147 - 1995
[9] B. D. Lahoti, D. E. Flowers, Evaluation of transformer loading above nameplate rating, IEEE Trans. Power Aparaturs System PAS-100 (4) (1981) 1989-1998
[10] Danny Snyders: “*Maxwell*”, an interative multimedia program, project of the katholike Universitet of Leoven the Provinciale Industriele Hogeschool of West Vlanderen.

POSTER SESSION PO I

PO I - Control Systems

Legendre Orthogonal Functional Network Applied in Modeling of Dynamical Systems

Dragan Antić[1], Bratislav Danković[2], Zoran Jovanović[3] and Staniša Perić[4]

***Abstract* – Functional networks represent extension of neural networks, which deal with general functional models. In this paper, we present Legendre-type orthogonal functional network and their application in the modeling of dynamical systems. As an example of the proposed modeling method, we considered the modeling of the hydraulic multi tank system.**

***Keywords* – Legendre polynomials, Orthogonal functional networks, Genetic algorithm, Modeling.**

I. INTRODUCTION

Functional networks were first introduced in [1, 2]. They represent extension of the standard neural networks and unlike neural networks, they deal with general functional models. Functional network have many advantages, so the problem that can be solved by the neural network, also can be formulated by functional network [1]. There are also many examples that cannot be solved by the neural network but can [1-3] be naturally formulated using the functional network. Neuron functions may be multivariate, multi-argument, and different. Functional network is a very useful general framework for solving a wide range of problems: the solving of differential functional and difference equation [2], nonlinear time series and prediction modeling [3], factorization model of multivariate polynomials [4], the identification of nonlinear system [5], linear and nonlinear regression [4], approximation functions [6] etc. Some of these applications have been developed in [5-7]. In this paper, an orthogonal functional network model is presented whose neuron functions are approximated by Legendre orthogonal basis functions.

The Legendre polynomials and their orthogonal properties were established during eighteenth century. One of the most important applications of Legendre orthogonal functions is designing orthogonal filters [8-11] which are useful for forming orthogonal signal generators, least square approximations and practical realizations of optimal and adaptive systems. This paper describes a method for obtaining classical orthogonal Legendre functions using a new transform. The obtained functions have been used to design orthogonal neurons that enable generating the known orthogonal functions in the form of real physical signals. The parameters and weights of the Legendre orthogonal functional network are determined by genetic algorithm as the learning algorithm and known optimization technique. As a case study, an experimental three tank hydraulic system was considered. Experiments were performed to approve theoretical results and demonstrate that the method described in the paper is very suitable for modeling dynamical systems in the sense of accuracy and algorithm speed.

[1]Dragan Antić is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: dragan.antic@ elfak.ni.ac.rs

[2]Bratislav Danković is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: bratislav.dankovic@elfak.ni.ac.rs

[3]Zoran Jovanović is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: zoran.jovanovic@ elfak.ni.ac.rs

[4]Staniša Perić is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: stanisa.peric@ elfak.ni.ac.rs

II. LEGENDRE ORTHOGONAL FUNCTIONS

Consider the orthogonal, shifted Legendre polynomials in their explicit form [11-12]:

$$P_n(x)=\frac{1}{n!}\sum_{j=0}^{n}(-1)^{n-j}\binom{n}{j}\frac{(n+j)!}{j!}x^j \tag{1}$$

These polynomials are orthogonal over interval (0, 1). From Eq. (1) we obtain the sequence of Legendre type orthogonal polynomials:

$$\begin{aligned} &P_0(x)=1,\\ &P_1(x)=2x-1,\\ &P_2(x)=6x^2-6x+1,\\ &P_3(x)=20x^3-30x^2+12x-1,.... \end{aligned} \tag{2}$$

After applying the substitution $x=e^{-t}$ into Eq. (1) and Laplace transform, the following rational function can be obtained:

$$W_n(s)=\frac{1}{s}\frac{\prod_{i=1}^{n}(s-i)}{\prod_{i=1}^{n}(s+i)} \tag{3}$$

Denote for $\varphi_n(t)=L^{-1}\left[W_n(s)\right]$ series of Legendre exponential functions orthogonal over interval $(0,\infty)$ with weight $w(t)=e^{-t}$ i.e.:

$$\int_0^\infty \varphi_i(t)\varphi_j(t)e^{-t}dt = \begin{cases} 0, & i \neq j \\ N_i, & i = j \end{cases} \quad (4)$$

Rational functions given with (3) can be factorized and realized in the form of the filter. Each filter section with transfer function $(s\text{-}i)/(s+i)$ can be easily practically realized (Fig. 1).

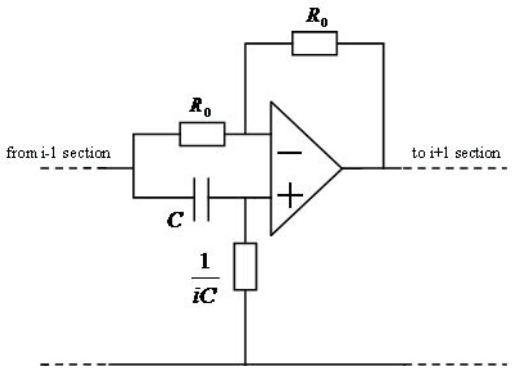


Fig. 1. Practical realization of the single filter section

III. LEGENDRE ORTHOGONAL FUNCTIONAL NETWORKS

Typical architecture of a functional network [6] is given in Fig. 2.

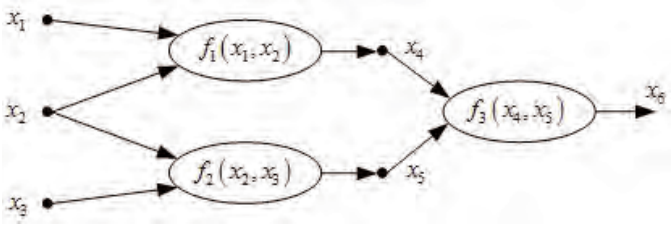


Fig. 2. A functional network model

Besides of the input (x_1, x_2, x_3) and output (x_6) layers, a functional network consists of one or more layers of intermediate storing units (x_4, x_5), which store information produced by neuron units, and one or more layers of processing units (f_1, f_2, f_3). A neuron unit evaluates a set of input values, coming from the previous layer (or input units) and delivers a set of output values to the next layer (or output units). Each neuron has associated neuron function that can be multivariate and can have as many arguments as inputs. Once the input values are given, the output is determined by the neuron type and its function.

If we use orthogonal basis functions for neurons, we can obtain general orthogonal functional network [13] (Fig. 3) with the following output:

$$y = \sum_{i=1}^{n} w_i f_i(x) \quad (5)$$

If we use Legendre orthogonal series described in Section 2 as orthogonal basis, we can design Legendre orthogonal functional networks. Function f_i in this case will represent the first i sections connected in series $f_i(s) = \prod_{j=1}^{i} \frac{s-j}{s+j}$. Output signals of these sections are orthogonal in time domain. These networks can approximate any given function and model any system by adjusting variable weights w_i. These weights are adjusted in such a way that we minimize the error $e = y - f(x)$ where $f(x)$ represents either the function to be approximated or the response to the applied input signal x of the system to be modeled. In our experiments, optimal weights will be determined by genetic algorithm [14-15].

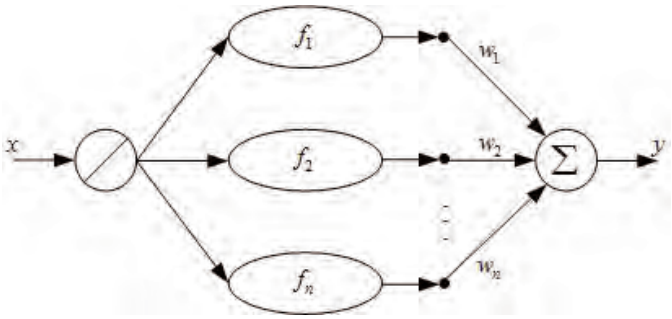


Fig. 3. A general orthogonal function network

IV. EXPERIMENTAL RESULTS

For the purpose of verification of our modeling method, we consider a multi tank hydraulic system manufactured by "Inteco", Poland [16] (Fig. 4). The multitank system relates to liquid level control problems commonly occurring in industrial storage tanks. It comprises three separate tanks fitted with drain valves. The separate tank mounted in the base of the set-up acts as a water reservoir for the system. Some of the tanks have a constant cross section, while others are spherical or conical, so having variable cross section (this creates main nonlinearities of the system). Several issues have been recognized as potential impediments to high accuracy modeling and control of level or flow in the tanks: nonlinearities caused by shapes of tanks, saturation-type nonlinearities (introduced by maximum or minimum allowed level in tanks), valve geometry and flow dynamics, pump and valves input/output characteristic curve.

Fig. 4. Multitank hydraulic system

In order to obtain a model for the given hydraulic system, method described in previous section is applied. The same step input signal (x=1) is applied to both the unknown system and Legendre orthogonal functional network with four neurons in the middle layer (four function in orthogonal basis). The only known data about the system is the measured output - tank liquid level H_1 ($f(x)$), given in Fig. 5.

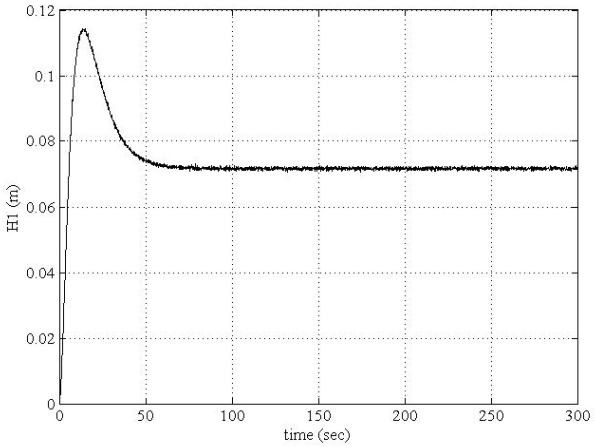


Fig. 5. Step response of unknown hydraulic system ($f(x)$)

The next step is to form difference of measured system output and functional network output as well as to calculate the error. Optimal parameters values for the best model of unknown system are determined by using genetic algorithm. Genetic algorithm used in experiment was with the following parameters: initial population of 200, number of generations 100, stochastic uniform selection, reproduction with 10 elite individuals, Gaussian mutation with shrinking and scattered crossover. The goal of the experiment was to make a error (difference between the outputs from the system and from the functional network) as small as possible for a choosen input, i.e., to obtain the best model of the unknown system. So, we used error as the fitness function for the genetic algorithm. Experimental time was 300 seconds.

After the experiment, following weights for the Legendre orthogonal functional network were obtained: w_1=0.0691, w_2=-0.0602, w_3=-0.0332 and w_4=0.0243. With these weights, final functional network output (y) is given in Fig. 6. We can see from the Figs. 5 and 6 that our functional network now represents accurate model of considered three tank hydraulic system.

V. CONCLUSION

In this paper we present a new method for obtaining models of dynamical systems based on Legendre orthogonal functional networks. Theoretical background on this method is given with full description of Legendre functions. Using these functions, we designed Legendre orthogonal functional network. Filters are described in details and illustrated. The idea for a new method for systems modeling is based on the fact that functions generated by Legendre network are orthogonal in the left semi plane of the complex plane, so this network is convenient for the modeling of continuous systems.

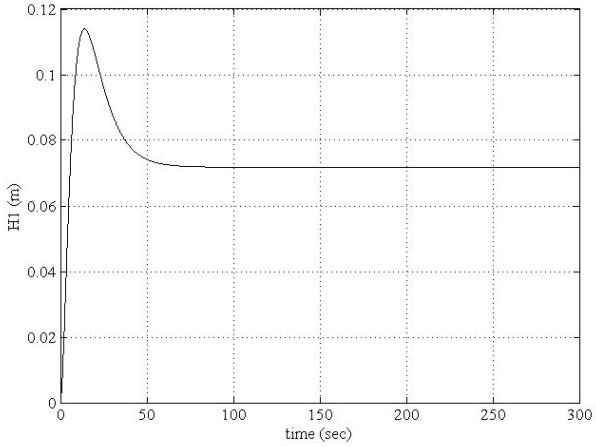


Fig. 6. Step response of Legendre orthogonal functional network (y)

Adjustable network which can be used for modeling of arbitrary systems can be designed on the basis of these functions. Specific models are obtained using adjustable weights. During the modeling of concrete unknown system, these weights are being adjusted in order to obtain the best model of unknown system in the sense of mean square error. Optimal adjustment of the weights is accomplished using genetic algorithms, which have demonstrated very good performances as global optimizers in many types of applications.

In order to verify the obtained theoretical results, a modeling of a multitank hydraulic system is considered. Model of this system is obtained using a proposed modeling method. Experiments have demonstrated that the method described in the paper is very suitable for systems modeling and it achieves excellent results in the sense of modeling algorithm simplicity and speed as well as model accuracy.

REFERENCES

[1] E. Castillo, “Functional Networks”, Neural Processing Letters, vol.7, pp. 151-159, 1998.

[2] E. Castillo, A Cobo, J.M. Gutierez and E. Pruneda, “Working with Differntial, Functional and Equations Using Functional Networks”, Appl. Math. Model., vol. 23, pp. 89-107, 1999.

[3] E. Castillo and J.M. Gutierez, “Nonlinear Time Series Modeling and Prediction Using Functional Networks, Exctracting Information Masked by Chaos”, Phys. Lett. A, vol. 244, pp. 71-84, 1998.

[4] Y. Zhou, “An Approximate Factorization Model of Multivariate Polynomials Based on Algebra Neural Networks and Leraning Algorithm”, Journal of Computation Research&Development, vol. 36, no. 6, pp. 668-674, 1999.

[5] E. Castillo, A Cobo, J.M. Gutierez and E. Pruneda, *An Introduction to Functional Networks with Applications*, New York, Kluwer Academic Publisher, 1998.

[6] Y. Zhou, X. Lu, Z. Xie and B. Liu, An Orthogonal Functional Network for Function Approximation”, Inetrnational Conference on Intelligent Cimputation Technology and Automation, Changsha, Hunan, China2008, vol. 2, pp. 962-966.

[7] E. Castillo, J.M. Gutierez, A.S. Hadi and B. Lacruz, “Some Applications of Functional Networks in Statistics and Engineering”, Technometrics, vol. 43, no. 1, pp. 10-24, 2001.
[8] P.C. McCarthy, J.E. Sayre and B.L.R. Shawyer, “Generalized Legendre Polynomials”, Journal of Mathematical Analysis and Applications, vol. 177, pp. 530-537, 1993.
[9] B. Danković, S. Nikolić, M. Milojković and Z. Jovanović, “A Class of Almost Orthogonal Filters”, Journal of Circuits, Systems, and Computers, vol. 18, no. 5, pp. 923-931, 2009.
[10] L.J. Karam and J.H. McClellan, “Complex Chebyshev Approximation for IIR Filter Design”, IEEE Transactions of Circuits and Systems II, vol. 42, pp. 207-216, 1995.
[11] M. Milojković, S. Nikolić, B. Danković, D. Antić and Z. Jovanović, “Modelling of Dynamical Systems Based on Almost Orthogonal Polynomials”, Mathematical and Computer Modelling of Dynamical Systems, in press.
[12] M. Abramovic and I. Stegun, *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Table*, National Bureau of Standards Applied Mathematics Series, Washington, 1972.
[13] Y. Zhang, T. Zhong, W. Li, X. Xiao and C. Yi, “Growing Algorithm of Laguerre Orthogonal Basis Neural Network with Weights Directly Determined”, in *Lecture Notes in Computer Science, Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence 5227*, eds. D.S. Huang, Berlin, Springer-Verlag, pp. 60-67, 2008.
[14] J.H. Holland, *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, 1975.
[15] M. Mitchel, *An Introduction to Genetic Algorithms*, MIT Press, Cambridge, 1996.
[16] Inteco, *Modular Servo System-User’s Manual*, (2008) Available at www.inteco.com.pl

Identification of the Multitank System Using Genetic Algorithm

Darko Mitić[1], Dragan Antić[2], Saša Nikolić[3] and Marko Milojković[4]

***Abstract* – The paper considers the implementation of the genetic algorithms in parameter identification of three tank hydraulic system, defined by highly nonlinear mathematical model. The output orifice resistances and the flow coefficients are estimated for all tanks, based on the experimental measured data. The validation of the model and parameters are performed by digital simulation and comparison with the experimental results.**

***Keywords* – Identification, Genetic algorithm, Multitank System.**

I. INTRODUCTION

Genetic Algorithms (GAs) are well known evolutionary search algorithms implemented in many control engineering problems [1, 2]. Since GAs represent optimization procedures, they can be applied to identification problems.

There are two inter-related problems existing in system identification: the choice of a suitable model structure and the estimation of model parameters. Techniques for the selection of structure and for non-linear-in-the-parameters estimation are still the subject of ongoing research [3]. There are many classical methods for system identification, such as least-squares and maximum-likelihood. Most of these are for linear or linear-in-the-parameters non-linear systems, and they often fails in the search for a global optimum if the search space is not differentiable or linear-in-parameters. Furthermore, for these methods, initial information on the system parameters is needed for convergence; estimated parameters may be biased if the noise is correlated; and they cannot easily be applied to nonlinear systems.

GAs can be applied to continuous- and discrete-time system, both on-line and off-line and both time domain and frequency domain systems, and can directly identify physical parameters or poles and zeroes of the system. In this paper, one GA is implemented in the parameter identification of the three tank hydraulic system, which represents the highly nonlinear plants.

The paper is organized as follows. The problem formulation is given in Section 2 including the statement of the plant nonlinear model. Section 3 describes the used GA briefly. Experimental framework is presented in Section 4 both with the results of the identification. Section 5 summarizes the results.

[1]Darko Mitić is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: darko.mitic@ elfak.ni.ac.rs

[2]Dragan Antić is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: dragan.antic@ elfak.ni.ac.rs

[3]Saša Nikolić is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: sasa.s.nikolic@ elfak.ni.ac.rs

[4]Marko Milojković is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: marko.milojkovic@ elfak.ni.ac.rs

II. PROBLEM STATEMENT

A. Three Tank System Description

Three-tank system has been proposed as a benchmark system for system modeling, identification, control, fault detection and diagnosis, as well as for fault-tolerant control. The system exhibits typical characteristics of a constrained hybrid system and has been proven useful to serve as a test environment for algorithms concerning state estimation, parameter identification, and control systems. Such a system can be viewed as a prototype of many industrial applications in process industry, such as (petro) chemical plants and oil, and gas systems.

The plant corresponds to the multitank systems manufactured by Inteco, Poland [4]. The multitank system (Fig. 1) comprises a number of separate tanks fitted with drain valves. The separate tank mounted in the base of the set-up acts as a water reservoir for the system. Some of the tanks have a constant cross section, while others are spherical or conical, so having variable cross section. This creates main nonlinearities of the system. A variable speed pump is used to fill the upper tank. The liquid outflows the tanks due to gravity. The tank valves act as flow resistors. The area ratio of the valves is controlled and can be used to vary the outflow characteristic. Each tank is equipped with a level sensor based on hydraulic pressure measurement.

The multitank system relates to liquid level control problems commonly occurring in industrial storage tanks. For example, steel producing companies around the world have repeatedly confirmed that substantial benefits are gained from accurate mould level control in continuous bloom casting. Mould level oscillations tend to stir foreign particles and flux powder into molten metal, resulting in surface defects in the final product. The Multitank System has been designed to operate with an external, PC-based digital controller. The control computer communicates with the level sensors, valves and pump by a dedicated I/O board and the power interface. The I/O board is controlled by the real-time software which operates in MATLAB®/Simulink RTW/RTWT® rapid prototyping environment.

Fig. 1. The Multitank system by Inteco

B. *The Mathematical Model*

The three tank system depicted in Fig. 2. can be decribed using the well known "mass balance" equations:

$$\frac{dH_1}{dt} = \frac{1}{\beta_1(H_1)} q - \frac{1}{\beta_1(H_1)} C_1 H_1^{\alpha_1} \tag{1}$$

$$\frac{dH_2}{dt} = \frac{1}{\beta_2(H_2)} C_1 H_1^{\alpha_1} - \frac{1}{\beta_2(H_2)} C_2 H_2^{\alpha_2} \tag{2}$$

$$\frac{dH_3}{dt} = \frac{1}{\beta_3(H_3)} C_2 H_2^{\alpha_2} - \frac{1}{\beta_3(H_3)} C_3 H_3^{\alpha_3} \tag{3}$$

where q represents inflow to the upper tank, H_i is the fluid level in the *i*-th tank (*i*=1,2,3), C_i is the resistance of the output orifice of *i*-th tank, α_i represents the flow coefficient for the *i*-th tank.

$\beta_i(H_i)$ represents the cross sectional area of *i*-th tank at the level H_i. These values for the single tanks are the following: $\beta_1(H_1) = aw$ is the constant cross sectional area of the upper tank; $\beta_2(H_2) = cw + \frac{H_2}{H_{2\max}} bw$ is the variable cross sectional area for the middle tank, and $\beta_3(H_3) = w\sqrt{R^2 - (R - H_3)^2}$ is the variable cross sectional area of the lower tank.

The specified parameter values are the following: $a = 0.25m,\ b = 0.345m,\ c = 0.1m,\ w = 0.035m,\ R = 0.364m$, and $H_{1\max} = H_{2\max} = H_{3\max} = 0.35m$.

Liquid levels in the tanks H_1,H_2,H_3 are the state variables of the system. There are four controlled inputs: liquid inflow q and valves settings C_1,C_2,C_3. The goal of the identification is to determine values of parameters C_i and α_i (*i*=1,2,3).

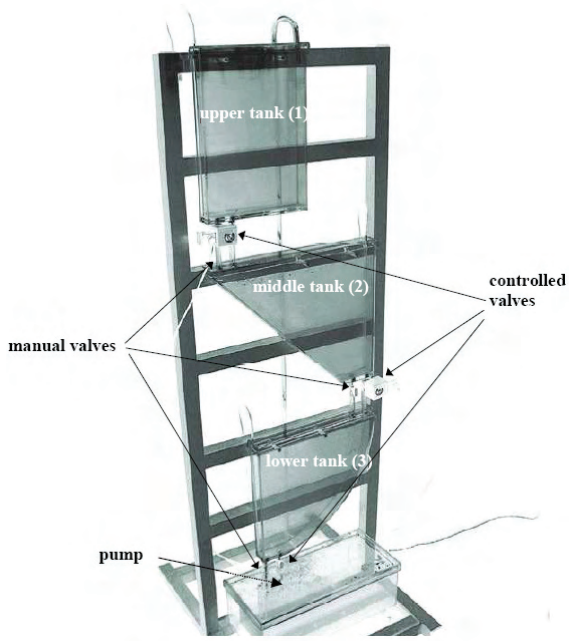


Fig. 2. Plant structure

III. GENETIC ALGORITHM

The principles of the genetic algorithms were first published by Holland in 1962 [5]. The mathematical framework was developed in the late 1960's, and is presented in Holland's book, *Adaptation in Natural and Artificial Systems* published in 1975 [6].

Genetic algorithms are optimization techniques based on simulating the phenomena that takes place in the evolution of species and adapting it to an optimization problem. These techniques imply applying the laws of natural selection onto the population to achieve individuals that are better adjusted to their environment. The population is nothing more than a set of points in the search space. Each individual of the population represents a point in that space by means of his chromosome. The individual's degree of adaptation is given by the objective function. Applying genetic operators to an initial population simulates the evolution mechanism of individuals. 'Survival of the fittest' philosophy is used to speed up the evaluation process.

The genetic algorithms have been used in many diverse areas such as function optimization, image processing, system identification, system modeling..., and have demonstrated very good performances as global optimizers in many types of applications [7-10].

The genetic algorithm used in our experiments was with the following parameters: initial population of 400, number of generations 200, stochastic uniform selection, reproduction with 10 elite individuals, Gaussian mutation with shrinking and scattered crossover. The goal of the experiment was to make a error as small as possible for a chosen input, i.e., to obtain the best parameters of the hydraulic system. So, we used error as the fitness function for the genetic algorithm.

IV. EXPERIMENTAL RESULTS

The tank system consists of a number of tanks placed above each other (Fig. 2). Some of the tanks have a constant cross section, while others are spherical or prismatic, so having variable cross section. Liquid is pumped into the upper tank from the supply tank by the pump driven by a DC motor. The liquid outflows the tanks only due to gravity. The output orifices act as flow resistors, but can also be controlled from the computer.

The levels in the tanks are measured with pressure transducers. The signals from the sensors are connected to the analog inputs of the RT-DAC4 multipurpose PC I/O board. There are four control signals send out from the board to the multitank system: three valve controls and one pump control signal. The appropriate PWM control signals are transmitted from digital outputs of the I/O board to the power interface, and next to the valves and to the DC motor. The speed of the pump motor is controlled by a sequence of PWM pulses configured and generated by the logic of XILINX chip of the RT-DAC4 board.

The plant is closed system, where the liquid that enters the reservoir from the tanks returns to the tanks via the pumps. However, these pumps will be switched off automatically when the liquid level of upper tank or lower tank exceeds a given upper limit.

During the process of parameter identification in our experiments, as a criteria function, mean squared error can be used:

$$J = \frac{1}{T}\int_0^T \left(H_k - H_k^m\right)^2 dt \qquad (4)$$

where H_k represent the measured liquid level for the *k*-th tank and H_k^m represent of the simulation results for the same tank. Parameters of the hydraulic system are obtained by adjusting the model in order to minimize the criteria function (4) by the genetic algorithm.

The complete block diagram, which illustrates the process of identification, is given in Fig. 3.

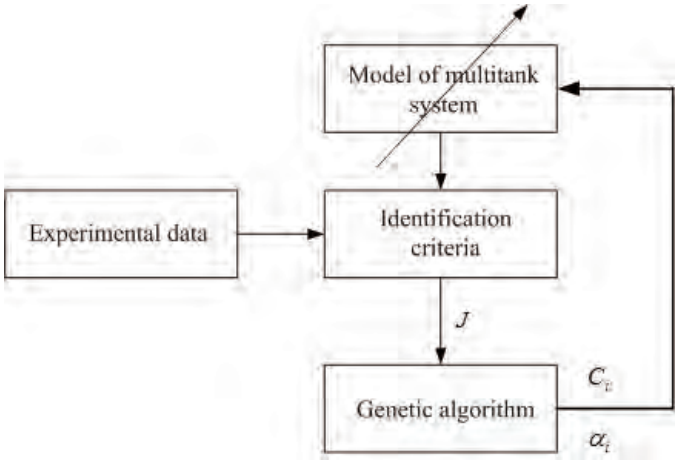


Fig. 3. Priniciple of tank parameters identification

For fixed valves settings the parameters C_i and α_i (i=1,2,3) of the mathematical models of the tank system have to be identified experimentally by using the genetic algorithm. Experimental time was 60 seconds. For each tank the outflow experiment has been performed, the data are collected and the characteristic curves has been fitted to the data.

After the experiments, following parameters for the multitank system were obtained: for upper tank $C_1 = 6.9501 \cdot 10^{-5}$, $\alpha_1 = 0.33692$, for middle tank $C_2 = 6.3214 \cdot 10^{-5}$, $\alpha_2 = 0.30125$, and for bottom tank $C_3 = 6.3954 \cdot 10^{-5}$, $\alpha_3 = 0.32478$.

These results are used in digital simulation in order to validate the proposed solution. Figures 4, 5 and 6 present both the measured and simulation liquid levels of three tanks, respectively. It is obvious that the proposed GA gives good accuracy in estimation of the system parameters. Some larger error in the end of the experiment can be explained as the consequence of additional nonlinearities arising on the low levels of liquid in the tanks.

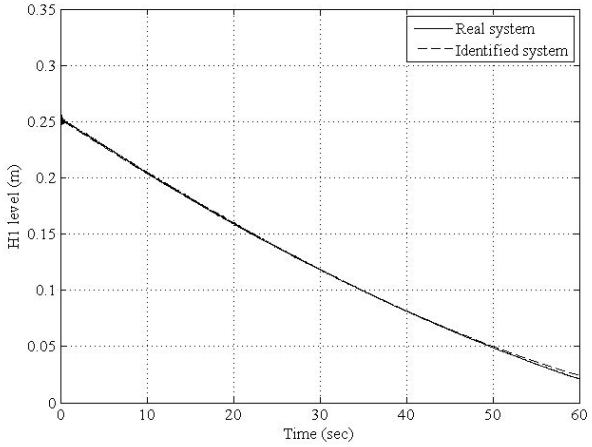


Fig. 4. The upper tank

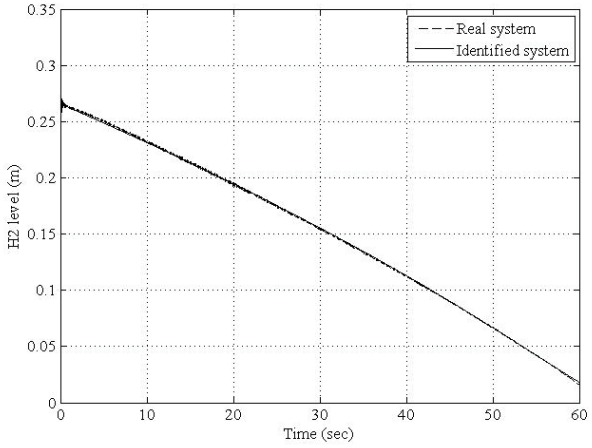


Fig. 5. The middle tank

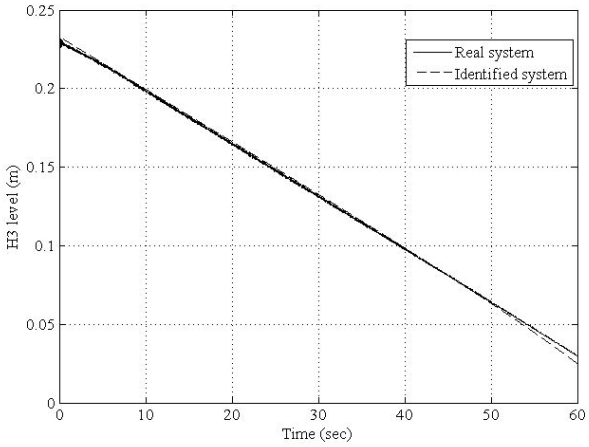


Fig. 6. The bottom tank

V. CONCLUSION

The paper treats the problem of parameter identification of hydraulic system consisting of three tanks by using genetic algorithm optimisation techniques. The system model is highly nonlinear, so the traditional identification techniques can fail in estimation process. The proposed genetic algorithm requires nothing more than lyquid levels measure and do not pose any restrictions to the identification problem. The obtained parameter values are used in digital simulation and the model verification is done by comparing the experimental data with simulated ones. The proposed GA gives good accuracy in estimation of the system parameters.

REFERENCES

[1] P.J. Fleming and R.C. Purshouse, "Evolutionary Algorithms in Control Systems Engineering: A Survey", Control Engineering Practice, vol. 10, pp. 1223–1241, 2002.

[2] Q. Wang, P. Spronck and R. Tracht, "An Overview of Genetic Algorithms Applied to Control Engineering Problems", Proceedings of the Second International Conference on Machine Learning and Cybernetics, Xisan, pp. 1651-1656, 2003

[3] G.J. Gray, D.J. Murray-Smith, Y.Lin, K.C. Sharman and T. Weinbrenner: "Nonlinear Model Structure Identification Using Genetic Programming", Control Engineering Practice, vol. 6, pp. 1341-1352, 1998.

[4] Inteco, *Modular Servo System-User's Manual*, (2008) Available at www.inteco.com.pl

[5] J.H. Holland, "Outline for a Logical Theory of Adaptive Systems", J. ACM, vol. 3, pp. 297-314, 1962.; also in A.W. Burks, Ed., *Essays on Cellular Automata*, University Illinois Press, pp. 297-319, 1970.

[6] J.H. Holland, *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, 1975.

[7] D. Antić and M. Milojković, "Nonlinear System Control by Using Genetic Algorithms and Fuzzy Sliding Mode", TEHNIKA, Elektrotehnika, vol. 56, pp. 9-16, 2007.

[8] B. Danković, D. Antić, Z. Jovanović and M. Milojković, "Genetic Algorithms Applied in Parameter Optimization of Casacade Connected Systems", ICEST 2007, Conference Proceedings, Ohrid, Macedonia, pp. 557-560, 2007.

[9] D. Antić, M. Milojković and D. Mitić, "An Improvement of Fuzzy Sliding Mode Control of Nonminimum Phase Plants by Using Genetic Algorithms", SAUM 2007, Conference Proceedings, Niš, Serbia, pp. 129-132, 2007.

[10] D. Antić, B. Danković, Z. Jovanović, S. Nikolić and M. Milojković, "Systems Modeling Based on Rational Functions", Scien. Bull. UPT, Trans. Autom. Cont. Comp. Scien., vol. 54(68), no. 4, pp. 149-154, 2009.

Identification of One Class of Distributed Parameter Systems Based on Orthogonal Functions

Mariana G. Todorova[1]

***Abstract* – In this paper the identification issue of one class of distributed parameter systems (DPS) by means of two-dimensional orthogonal polynomials and functions is studied. Suitable m-functions are proposed and numerical examples are conducted to demonstrate the validity and accuracy of the method.**

***Keywords* – Identification, Distributed parameter system, Orthogonal functions**

I. INTRODUCTION

Knowledge of the DPS mathematical model is a crucial point in its investigation of dynamic behavior and design of control systems. Two approaches to find models are usually applied: analytical and experimental.

Mathematical modeling in the first approach, which is associated with the design phase of the facility, shall be made on certain assumptions and assumptions that largely determine the accuracy of the model.

The second approach uses identification methods, including questions about the structure of the system model and its relevant parameters. Typically, the type of equations describing the model structure is chosen in advance based on a priori information about the ongoing physical processes and the task of the identification is reduced to parameter estimation. Identification of DPS is far more complex than the case of lumped parameter systems, because the mathematical model of DPS is often partial differential equations (PDE) or integral equations, and the solutions of DPS are concerned with not only initial conditions but also boundary conditions, which increases the complexity.

Two methods of identification of DPS are usually used. The first class of methods uses some optimization approaches, such as Kalman filter algorithm, gradient method, genetic algorithm optimization [10], etc. The second class of methods translates the mathematical model of DPS into algebraic, transcendental or differential equations in advance and then identificates the translated system. Some of the approximation methods are: differential quadrature method (Bellman and Roth, 1979), Walsh functions (Sinha et al. 1980), finite element method (Brahmanandam and Chatterji, 1982), Laguerre polynomials [1, 2, 5], block – pulse functions [4, 6, 7, 8], Haar wavelets [7, 9], etc.

In this paper the attention is focus on the second class of methods. The identification of one class of DPS based on two-dimensional orthogonal polynomials and functions is considered.

The rest of the paper is organized as follows. In section II some mathematical preparations about often used orthogonal functions and polynomials are given. The properties of two-dimensional orthogonal functions are considered. The algorithm of parameter identification for one class of DPS based on two-dimensional orthogonal polynomials and functions is presented in section III. In section IV, numerical simulations to demonstrate the validity and efficiency of the methods are given. Some conclusion in section V is made.

[1]Mariana G. Todorova is with the Faculty of Computing and Automation, Technical University, "Studentska" str. № 1, 9010 Varna, Bulgaria, E-mail: mgtodorova@yahoo.com

II. MATHEMATICAL PREPARATIONS

Orthogonal polynomials have very useful properties in the solution of mathematical and physical problems.

The Laguerre polynomials $\lambda(z)$ are defined over a range [0; ∞). They can be defined recursively, defining the first two polynomials as $\lambda_0(z)=1; \lambda_1(z)=1-z$ and then using the following recurrence relation

$$(n+1).\lambda_{n+1}(z)-(2.n+1-z).\lambda_n(z)+n.\lambda_{n-1}(z)=0\,, \qquad (1)$$

for any $n \geq 1$.

The first class of Chebyshev polynomials *T(z)* is defined over a range [-1, 1]. In this paper we briefly call it Chebyshev polynomials. They are orthogonal polynomials with respect to the weighting function $\rho(z)=\frac{1}{\sqrt{1-z^2}}$ and can be defined recursively by using the following recurrence relation

$$T_{n+1}(z)-2.z.T_n(z)+T_{n-1}(z)=0\,, \qquad (2)$$

where: $T_0(z)=1; T_1(z)=z$.

A block – pulse function (BPF) *B(t)* is defined over a time interval $t \in [0,T]$ as

$$\{B_i(t)\}, \quad i=1,2,...,m\,, \qquad (3)$$

where:

$$B_i(t)=1, \quad for\ t\in[0;T/m]$$

$$B_i(t)=\begin{cases}1, & for\ t\in[(i-1)T/m;\ iT/m]\\ 0, & elsewhere\end{cases}$$

$$for\ i=2,3,...,m$$

The orthogonal set of Haar functions *H(t)* is a group of square waves with magnitude of $\pm 2^{m/2}$ in some intervals and zeros elsewhere. The Haar functions are defined as

$$H_m(t)=H_1(t).(2^j.t-\frac{k}{2^j})\,, \qquad (4)$$

where: $j \geq 0; \quad m=2^j+k; \quad 0<k\leq 2^j$;

$H_1(t)$ - scaling function, pleased during the whole observed interval [0, T].

The two-dimensional orthogonal functions $F_{ij}(x,t)$ are defined as the set of orthogonal functions over the intervals $t \in [0,T]$, $x \in [0,X]$ as

$$\{F_{ij}(x,t)\} = F_i(t).F_j(x) \tag{5}$$

A function y(x,t), absolutely integrable in the region $t \in [0,T]$, $x \in [0,X]$, may be approximated to

$$y(x,t) \cong \sum_{i=1}^{m}\sum_{j=1}^{n} y_{ij}F_{ij}(x,t) = F_M^T(t).Y.F_N(x)\,, \tag{6}$$

where: Y - two-dimensional orthogonal function coefficient matrix of the function $y(x,t)$ $Y = [y_{ij}]_{m\times n}$;

$$y_{ij} = \frac{T.X}{m.n}\int_{(j-1).X/n}^{j.X/n}\int_{(i-1).T/m}^{i.T/m} y(x,t).\,dt.dx\,;$$

$$F_M(t) = [F_1(t)\;\; F_2(t)\;\; \ldots\;\; F_m(t)]^T\,;$$

$$F_N(x) = [F_1(x)\;\; F_2(x)\;\; \ldots\;\; F_n(x)]^T$$

The orthogonal function $F(t)$ has the property

$$\underbrace{\int_0^t \ldots \int_0^t}_{r-times} F(t).dt^r = P_M^r.F(t) \tag{7}$$

If $y(x,t) \cong F_M^T(t).Y.F_N(x)$, applying the property (7) to $y(x,t)$ yields

$$\underbrace{\int_0^x \ldots \int_0^x}_{b-times}\;\underbrace{\int_0^t \ldots \int_0^t}_{a-times} y(x,t).dt^a.dx^b \cong F_M^T(t).(P_M^T)^a.Y.P_N^b.F_N(x)\,, \tag{8}$$

where P_M and P_N are correspondingly $(m\times m)$ and $(n\times n)$ integral operational matrices.

III. Algorithm of identification

A DPS described by the following first order PDE is considered

$$a_1\frac{\partial y(x,t)}{\partial x} + a_2\frac{\partial y(x,t)}{\partial t} + a_3 y(x,t) - u(x,t) = 0\,, \tag{9}$$

where:

u(x,t) and y(x,t) - are the input and the output of the DPS respectively;

x – the variable of location ($x \in [0;L]$);

t – time ($t \in [0;T]$);

a_1, a_2 and a_3 – the unknown system parameters.

The initial condition is $y(x,0) = f(x)$ and the boundary conditions are $y(0,t) = g(x)$.

Orthogonal polynomials and functions implementation reduces the problem of parameter identification of DPS to a computationally convenient form. The identification process includes the following fundamental steps.

- Expansion of the input u(x, t) and output y(x,t) functions of the PDE into two-dimensional orthogonal polynomials or functions;
- Rewriting of the PDE in the matrix form using the orthogonal functions properties and after some well known manipulations, i.e.

$$A.Q = H\,, \tag{10}$$

where: Q – vector of the estimation values of the parameters $\hat{a}_i$, $i = \overline{1,3}$.

- Solving of the obtained matrix equation for the vector of unknown parameters using least – squares technique

$$Q = (A^T.A)^{-1}.A^T.H \tag{11}$$

IV. Numerical simulation

Take a DPS, described by the PDE (9), where the real values of parameters are: $a_1 = 2, a_2 = 4, a_3 = 1$, the input is *u(x,t)=4x+2t+xt, y(x,0)=0, y(0,t)=0,* $x,t \in [0;10]$ *).*

The Laguerre and first class of Chebyshev polynomials, BPF and Haar orthogonal functions are used and m – files are created in Matlab based on the algorithm above. The estimation values of the parameters for a few number *m* of two-dimensional orthogonal polynomials and functions are calculated. The obtained parameters and relative parameter errors E are given in Table 1 and Table 2, where:

$$E = \sqrt{\frac{\sum_{i=1}^{3}(a_i - \hat{a}_i)^2}{\sum_{i=1}^{3} a_i^2}}.100\% \tag{12}$$

Table 1

Estimation values of the parameters $\hat{a}_i$ and relative parameter errors E by using *m=4*

orthogonal polynomials and functions	$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_3$	E [%]
Laguerre	2,0043	3,9965	0,9998	0,12
Chebyshev	1,9979	3,9903	1,0079	0,28
BPF	2,0000	4,0000	1,0000	0
Haar	2,1200	4,2500	0,7000	8,9

Table 2

Estimation values of the parameters $\hat{a}_i$ and relative parameter errors E by using *m=8*

orthogonal polynomials and functions	$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_3$	E [%]
Laguerre	2,0004	3,9864	1,0087	0,35
Chebyshev	2,0001	3,9954	0,9994	0,10
BPF	2,0000	4,0000	1,0000	0
Haar	2,0918	4,1836	0,8469	5,59

Obviously, the estimation precision is bigger when block – pulse function is implemented. By using a few number of two-dimensional Haar orthogonal functions the estimation precision is unsatisfactory.

The noise immunity of the identification algorithms is investigated. The output signal y(x,t) is simulated and independent zero – mean white Gaussian noise is applied for this purpose. The ratio nose-to-signal is q = 10% and q=14%. The relative parameter errors E, obtained by applying considered orthogonal polynomials and functions are shown in Figs.1-2.

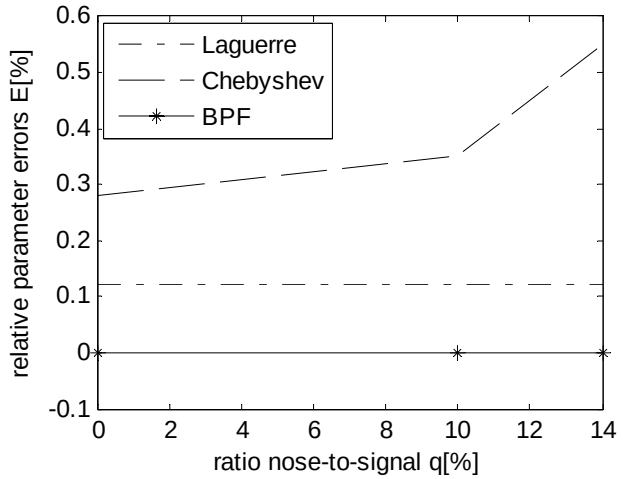


Fig.1 The relative parameter errors, obtained by applying *m=4*

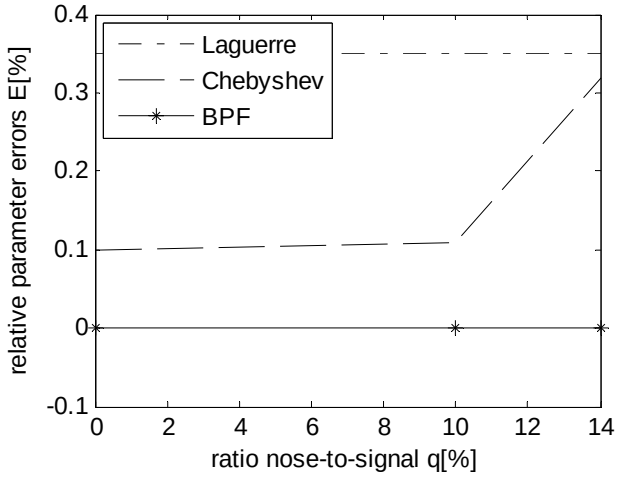


Fig.2The relative parameter errors, obtained by applying *m=8*

As shown in Figs.1-2 the estimates of the parameters via BPF-approach are very accurate.

V. CONCLUSION

The Laguerre and first class of Chebyshev orthogonal polynomials, BPF and Haar orthogonal functions are used for identification of one class of distributed parameter systems. Suitable m – files are created in Matlab based on the considered algorithms. Numerical examples are conducted to demonstrate the validity and accuracy of the method. The estimation values of the parameters for different number of two - dimensional orthogonal polynomials and functions are calculated. The noise immunity of algorithms for different ratio nose-to-signal is investigated.

The following conclusions can be made.

- The algorithms are comparatively simple in form and have low computer memory requirement.
- The obtained identification precision is bigger when BPF are implemented. Results from numerical simulation show that just a few number of BPF are need for high accuracy.
- The parameter estimations can be obtained comparatively accurate when bigger number of Haar orthogonal functions is applied.

REFERENCES

[1] Chang R. Y., S. Y. Yang, M. L. Wang, A new approach for parameter identification of time-varying systems via generalized orthogonal polynomials, Int. J. Control, vol.44, pp.1747-1755, 1986.

[2] Genov D., M. Todorova, T. Kolarov, Aplication of Laguerre polynomials for identification of ship models, Shipbuilding'98, Varna, pp. 139 – 142, 1998.

[3] Karova M., Smarkov V., Comparison Methods of Genetic Schemes in Genetic Algorithms, Proceedings of Third International Bulgarian-Turkish Conference Computer Science, Istanbul, Turkey, pp.375- 380, 2006.

[4] Nath A. K., T. T. Lee, On the multidimensional extension of block – pulse functions and their completeness, Int. J. Systems Sci., vol. 14, no. 2, pp. 201 – 208, 1983.

[5] Ranganathan V., A. Jha, V. Rajamani, Identification of linear distributed systems via Laguerre polynomials, Int. J. Systems Sci., 15, №10, 1984.

[6] Todorova M., T. Pencheva, Using of two-dimensional block pulse functions for identification of wastewater treatment process, AMSE Periodicals, France, Modelling C, pp. 58 – 73, 2007.

[7] Todorova M., Research of the possibilities and application of two dimensional orthogonal functions for dynamic distributed parameters systems identification, PhD Thesis, Varna, 2003.

[8] Wang Shienyu, Jiang Weisun, Identification of nonlinear distributed parameter systems using block pulse operator, IFAC Identification and System Parameter Estimation, York, UK, 1985.

[9] Todorova M., D. Genov, *Identification of non – linear distributed parameter systems via Haar wavelets*, International Conference on Automation and Informatics, ISBN 954-9641-30-9 (vol. 1), Sofia, 2002, pp. 233 – 237.

[10] Todorova M., B. Boyanov, *Genetic Algorithm Based Identification of Linear Distributed Parameter Systems by Finite Difference Technique*, International Conference on Computer Systems and Technologies - CompSysTech'2003, ISBN 954-9641-33-3, Sofia, 2003, pp. III.18 – III.19.

This page intentionally left blank.

Fuzzy Multiple Criteria Decision Making - MCDM

Emilija Kamceva[1] , Oliver Iliev[2] and Jasmin Ramadani[3]

Abstract –**Fuzzy MCDM has been developed by means of the following methods: (i) the outranking, (ii) the value and utility theory based, (iii) the multiple objective programming, (iv) group decision and negotiation theory based methods. Fuzzy-set theory enabled a great numnber of novelties. The most important methods have been considered below as well as the new MCDM method- interdependancy.**

Keywords –**fuzzy,MCDM -multiple criteria decision-making**

I. INTRODUCTION

Multiple Criteria Decision Making was introduced as a promising and important field of study in the early 1970'es. Since then the number of contributions to theories and models, which could be used as a basis for more systematic and rational decision making with multiple criteria, has continued to grow at a steady rate. A number of surveys, cf e.g. Bana e Costa [1], show the vitality of the field and the multitude of methods which have been developed.

Bana e Costa and Vincke [2] argue that with MCDM the first contributions to a truly scientific approach to decision making were made, but find fault with the objectives to carry this all the way as we have to deal with human decision makers who can never reach the degree of consistency needed.

They introduce multiple criteria decision aid MCDA as a remedy; this approach can be given the aim "to enhance the degree of conformity and coherence" in the decision processes carried out among (predominantly groups of) decision makers - this is done with a cross-adaptation of the value systems and the objectives of those involved in the process. Even if there are some distinctions between MCDM and MCDA the overall objective is the same: to help decision makers solve complex decision problems in a systematic, consistent and more productive way.

There are four major families of methods in MCDM:

(i) the outranking , (ii) the value and utility theory based (iii) the largest group is the interactive multiple objective programming approach with pioneering work (iv) group decision and negotiation theory introduced new ways to work explicitly with group dynamics and with diferences in knowledge, value systems and objectives among group members.

[1] Emilija Kamceva- Faculty Information - Communication Technologies- FON University, ul. Vojvodina b.b. 1000 ,Skopje,Macedonia ,e-maill:emilija.kamceva@fon.edu.mk

[2] Oliver Iliev- Faculty Information - Communication Technologies- FON University, ul. Vojvodina b.b. 1000, Skopje,Macedonia ,e-maill:oliver.iliev@fon.edu.mk

[3] Jasmin Ramadani - Faculty Information - Communication Technologies- FON University, ul. Vojvodina b.b. 1000 ,Skopje,Macedonia ,e-maill:jasmin.ramadani@fon.edu.mk

The first category contains a number of ways to find a ranking: degree of optimality, Hamming distance, comparison function, fuzzy mean and spread, proportion, left and right scores, area measurement and linguistic ranking methods.

The second category is built around methods which utilize various ways to assess the relative importance of multiple attributes: fuzzy simple additive weighting methods, analytic hierarchy process, fuzzy conjunctive / disjunctive methods, fuzzy outranking methods and maximin methods.

The category with the most frequent contributions is fuzzy mathematical programming. Inuiguchi et al give a useful survey of recent developments in fuzzy programming in which they work with the following families of applications: flexible prgramming, possibilistic programming, possibilistic linear programming, using fuzzy max, robust programming, possibilistic programming with fuzzy preference relations, possibilistic linear programming with fuzzy goals.

In order to introduce some of the key issues in fuzzy multiple criteria decison making we will work through a number of examples with a novel approach we have recently introduced, a method in which we allow the criteria to be interdependent. Then we will a give a brief overview of the contributions to this issue and close with a fairly comprehensive list of recent publications on fuzzy MCDM problems

II. DECISION-MAKING WITH INTERDEPENDENT CRITERIA

P.L. Yu explains that we have habitual ways of thinking, acting, judging and responding, which when taken together form our habitual domain (HD)[3] . This domain is very nicely illustrated with the following example ([3] page 560):

Example1.

A retiring chairman wanted to select a successor from two finalists (A and B). The chairman invited A and B to his farm, and gave each finalist an equally good horse. He pointed out the course of the race and the rules saying, "From this point whoever's horse is slower reaching the final point will be the new chairman". This rule of horse racing was outside the habitual ways of thinking of A and B. Both of them were puzzled and did not know what to do. After a few minutes, A all of a sudden got a great idea. he jumped out of the constraint of his HD. He quickly mounted B's horse and rode as fast as possible, leaving his own horse behind. When B realized what was going on, it was too late. A became the new chairman.

Part of the HD of multiple criteria decision-making is the intuitive assumption that all criteria are independent; this was initially introduced as a safeguard to get a feasible solution to a multiple criteria problem, as there were no means available

to deal with interdependence. Then, gradually, conflicts were introduced as we came to realize that multiple goals or objectives almost by necessity represent conflicting interests[5,6]. Here we will "jump out of the constraints" of the HD of MCDM and leave out the assumption of independent criteria. Decision-making with interdependent multiple criteria is a surprisingly dificult task. If we have clearly conflicting objectives there normally is no optimal solution which would simultaneously satisfy all the criteria. On the other hand, if we have pairwisely supportive objectives, such that the attainment of one objective helps us to attain another objective, then we should exploit this property in order to find effective optimal solutions.
In classical text Theory of Games and Economic Behavior John von Neumann and Oskar Morgenstern (1947) described the problem with interdependence; in their outline of a social exchange economy they discussed the case of two or more persons exchanging goods with each others :
then the result for each one will depend in general not merely upon his own actions but on those of the others as well. Thus each participant attempts to maximize a function .....of which he does not control all variables. This is certainly no maximum problem, but a peculiar and disconcerting mixture of several conflicting maximum problems. Every participant is guided by another principle and neither determines all variables which affect his interest.

This kind of problem is nowhere dealt with in classical mathematics. We emphasize at the risk of being pedantic that this is no conditional maximum problem, no problem of the calculus of variations, of functional analysis, etc. It arises in full clarity, even in the most "elementary" situations, e.g., when all variables can assume only a finite number of values.

This interdependence is part of the economic theory and all market economies, but in most modelling approaches in multiple criteria decision making there seems to be an implicit assumption that objectives should be independent. This appears to be the case, if not earlier then at least at the moment when we have to select some optimal compromise among a set of nondominated decision alternatives. Milan Zeleny and many others - recognizes one part of the interdependence ,
Multiple and conflixting objectives, for example, "minimize cost" and "maximize the quality of service" are the real stuff of the decision maker's or manager's daily concerns. Such problems are more complicated than the convenient assumptions of economics indicate. Improving achievement with respect to one objective can be accomplished only at the expense of another.
but not the other part: objectives could support each others.

In spite of the significant developements which have taken place in both the theory and the methodology MCDM is still not an explicit part of managerial decision-making [5]. By not allowing interdependence multiple criteria problems are simplified beyondrecognition and the solutions reached by the traditional algorithms have only marginal interest. Zeleny also points to other circumstances [5] which have reduced
the visibility and usefulness of MCDM:
time pressure reduces the number of criteria to be considered; (ii) the more complete and precise the problem definition, the less criteria are needed; (iii) autonomous decision makers are bound to use more criteria than those being controlled by a strict hierarchical decision system; (iv) isolation from the perturbations of changing environment reduces the need for multiple criteria; (v) the more complete, comprehensive and integrated knowledge of the problem the more criteria will be used - but partial, limited and non-integrated knowledge will significantly reduce the number of criteria; and (vi) cultures and organisations focused on central planning and collective decision-making rely on aggregation and the reduction of criteria in order to reach consensus.

Felix [7] presented a novel theory for multiple attribute decision making based on fuzzy relations between objectives, in which the interactive structure of objectives is inferred and represented explicitely.
With the following example in [45] he explains the need for a detailed automated reasoning about relationships between goals when we have to deal with nontrivial decision problems.

Example 2.
Let us suppose that there is a decision maker who wants to earn money (goal 1) and to have fun (goal 2) simultaneously, and the only way to earn money is to work. Then at least two situations are possible:
Situation 1: The decision maker does not like to work. Therefore, while working he will not have fun. The alternative working supports goal 1 but hinders goal 2.

Situation 2: The decision maker likes to work. Therefore, while working he will have fun. The alternative working supports both goal 1 and goal 2.

Relationships between two goals are defined using fuzzy inclusion and non-inclusion between the support and hindering sets of the corresponding goals. Felix [45] also illustrates, with an example, that the decision-making model based on relationships between goals can be used as a powerful MADM-method for solving vector maximum problems.
In multiple objective linear programming (MOLP), application functions are established to measure the degree of fulfillment of the decision maker's requirements (achievement of goals, nearness to an ideal point, satisfaction, etc.) on the objective functions (see e.g. [10, 11]) and are extensively used in the process of finding "good compromise" solutions.
In [9] we demonstrated that the use of interdependences among objectives of a MOLP in the definition of the application functions provides for more correct solutions and faster convergence. Generalizing the principle of application functions to fuzzy multiple objective programs (FMOP) with interdependent objectives, in [9], we have defined a large family of application functions for FMOP and illustrated our

ideas by a simple three-objective program. Let us now discuss our approach to interdependent MCDM.

III. Conclusion

In spite of the significant developements which have taken place in both the theory and the methodology MCDM is still not an explicit part of managerial decision-making. By not allowing interdependence multiple criteria problems are simplified beyond recognition and the solutions reached by the traditional algorithms have only marginal interest.

References

[1] C.A. Bana e Costa and P. Vincke, Multiple Criteria Decision Aid: An Overview, in: C.A. Bana e Costa ed., Readings in Multiple Criteria Decision Aid, Springer Verlag, Berlin-Heidelberg 1990, pp 3-14

[2] C.A. Bana e Costa ed.,Readings in Multiple Criteria Decision Aid, Springer Verlag, Berlin-Heidelberg 1990

[3] Po-Long Yu, To be a Great Operations Researcher from a MCDM Scholar, Computers & Operations Research, (19)1992 559-561.

[4] M.Zeleny, Multiple Criteria decision-making, McGraw-Hill, New-York, 1982.

[5] M. Zeleny, An Essay into a Philosophy of MCDM: AWay of Thinking or Another Algorithm Computers & Operations Research, (19)1992 563-566.

[6] R.R. Yager, On ordered weighted averaging aggregation operators in multicriteria decision-making, IEEE Trans. On Systems, Man, and Cybernetics, SMC- 18(1988) 183-190.

[7] R.Felix, Multiple attribute decision-making based on fuzzy relationships between objectives, in: Proceedings of the 2nd International Conference on Fuzzy Logic and Neural Networks, Iizuka Japan, July 17-22, 1992 805-808.

[8] R.Felix, Relationships between goals in multiple attribute decision-making, Fuzzy Sets and Systems, 67(1994) 47-52.

[9] C.Carlsson and R.Fullfer, Interdependence in fuzzy multiple objective programming, Fuzzy Sets and Systems 65(1994) 19-29.

[10] M.Delgado,J.L.Verdegay and M.A.Vila, A possibilistic approach for multiobjective programming problems. E±ciency of solutions, in: R.SÃlowinski and J.Teghem eds., Stochastic versus Fuzzy Approaches to Multiobjective Mathematical Programming under Uncertainty, Kluwer Academic Publisher, Dordrecht, 1990 229-248.

[11] H.-J.Zimmermann, Fuzzy programming and linear programming with several objective functions, Fuzzy Sets and Systems,1(1978) 45-55.

[12].P.L. Alreck and R.B. Settle, *The Survey Research Handbook* (Irwin, Illinois, 1985).

[13].S. Baas and H. Kwakernaak, Rating and raking of multiple-aspect alternatives using fuzzy sets, *Automatica* 13

[14]. J.F. Baldwin, Evidential support logic, FRIL and case based reasoning, to appear in *lnternat. J. IntelligentSystems.*

[15]. J.F. Baldwin, *FRIL Fuzzy and Evidential Reasoning in AI*(Research Studies Press, 1994).

[16]. J.F. Baldwin and N.C.F. Guild, Comparison of fuzzy setson the same decision space, *Fuzzy Sets and Systems*2

[17]. J,F. Baldwin, T. Martin and B. Pilsworth, *FRIL Manual*(FRIL Systems Ltd., Bristol Business Centre, Bristol BS81QX, UK, 1988).

[18]. J.F. Baldwin and R.A. Ribeiro, Fuzzy reasoning by case for decision support systems, *Uncertainty Fuzziness Knowledge-Based Systems* 2.

[19]. R.E. Bellman and L.A. Zadeh, Decision-making in a fuzzy environment, *Management Sci.* 17(4)(1970).

.

This page intentionally left blank.

RAMP Metering Analysis

Jasmin Ramadani[1], Zoran Gacovski[2], Emilija Kamceva[3]

***Abstract* – The great number of automobiles that are entering the freeways can cause intensification of the traffic which leads to congestions, traffic accidents, higher fuel consumption and CO2 emission in the atmosphere. This brings the need of managing the traffic flow through ramp metering of the vehicles which are trying to join the freeways in order to maintain the traffic flow.**

***Keywords* – *T*raffic, metering, algorithm, freeway, ramp.**

I. INTRODUCTION

The freeways that are connecting the urban areas are faced with congestions in the traffic flow operating at the limit of their capacity when the number of vehicles that are joining is rapidly growing in different periods of the day.

The main reasons for the congestions and traffic disturbances most frequently are accidents, queues of vehicles that are entering the freeway, bottlenecks, small capacity of the roads, growth of the vehicles number over the years and construction works on the freeways.

These occurrences are having negative impact on the traffic flow which brings the freeways to operate on their upper limit of capacity. As a result it leads to traffic congestions and greater time of travel accompanied with more travel costs.

II. THE PROCESS OF RAMP METERING

The metering of the traffic represents a technique of using traffic signs on the access lanes of the freeways in order to manage the number of vehicles entering on the main road. The purpose of this technique is to maintain balance between the demand and supply on the freeways and to enable the normal traffic flow and to avoid interruptions.

The traffic disorder can be classified as predictable and nonpredictable .congestions. The predictable are appearing in particular time periods such as the spillover of the traffic on the local roads, bottlenecks, the difference of the number of vehicles that are entering and leaving the freeway and the occurrence of platoons of entering vehicles. The unpredictable congestions are occurring as a result of temporal traffic flow reductions and traffic accidents. With the use of ramp metering systems the number of vehicles is controlled avoiding the situations of building platoons of vehicles over their segmenting and forcing the drivers for waiting on the ramp for the right moment to enter the freeway.

The purpose of this mechanism for regulation of the traffic flow is to provide normal traffic flow and to avoid the reaching of the peak of the road capacity which can bring to total collapse of the traffic [1]. The solution is found in the use of ramps (Fig. 1) on the access lanes on the freeways where the vehicle will have to wait some amount of time before they can enter the freeway.

The control of the flow on the ramp is regulated with fixed time periods or with flexible time periods which are determined from the condition of the freeway in the moment and the needs of the traffic.

The ramp metering control is considered for one of the most effective ways to manage the traffic flow. This mechanism is representing a strategy for maintaining the effectiveness of the traffic and to maintain the number of vehicles that are using the freeway under maximum capacity [2]. The ramp metering decreases the chances that traffic accidents will appear. The light signals are placed on the entrance of the lanes which are passing through the vehicles in accordance with the used cycle and number of vehicles. This system discourages the drivers to use the freeway for short trips leaving more space for the drivers that are already in the flow of the road and that are traveling on long distances.

The ramp metering management uses different types of algorithms which take real time measured variables detected from various detectors in order to get the results and reach the efficiency that is needed.

The technique of ramp metering is used in various cities in the world. For the first time it has been used in the United States in the 50ies and 60ies of the 20th Century.

III. SYSTEM COMPONENTS

The system of ramp metering consists of several components which are part of a larger architecture:

-*Signalization and controllers*- The signalization is usually placed on the left side of the vehicles or on both sides. The controller is programmed under certain algorithm which controls the access and the metering of the vehicles. The components of the system shown on Fig. 1 are the following:

-*Detector of demand*- this detector is located before the regulating system. It reports the controller that a vehicle is approaching on the ramp and sets the green interval for passing through.

-*Pass through detector*- it is located after the signal for regulating and metering. This detector informs the controller that an vehicles has passed by the ramp which should turn on the red interval which does not pass through any vehicles on the freeway.

[1]Jasmin Ramadani is with the Faculty of Information and Communication Technologies, Bul. Vojvodina bb, 1000 Skopje, Macedonia, E-mail: jasmin.ramadani@fon.edu.mk

[2]Prof. Dr. Zoran Gacovski is with the Faculty of Information and Communication Technologies, Bul. Vojvodina bb, 1000 Skopje, Macedonia, E-mail: zoran.gacovski@fon.edu.mk

[3]Emilija Kamceva is with the Faculty of Information and Communication Technologies, Bul. Vojvodina bb, 1000 Skopje, Macedonia, E-mail: emilija.kamceva@fon.edu.mk

-*Merging detector*- the function of this detector is to sense the presence of any vehicles on the traffic merging point. Before the vehicles gets the merging point the signal is red and after the vehicle has entered the point, the interval is set to green.

-*Queue detector*- the detector is located on the ramp of the system. This detector disables the spillover of the traffic congestion on the regional and access roads. If there is a possibility for this then the degree of metering is reduced.

-*Main flow detectors*- The detectors o the main traffic flow are located before and after the area of merging the highway traffic [3] and the access lane on which the metering is performed. These detectors are signaling the condition the traffic flow and the need traffic regulating. This detector uses the technique of occupancy measurement of the road and the information about the vehicles speed.

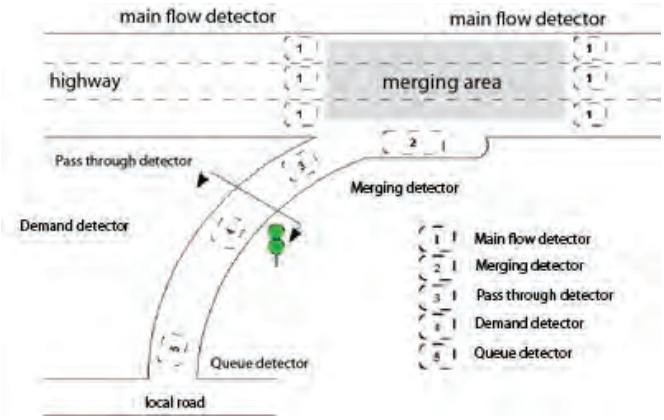


Fig. 1 Components of the Ramp Metering

IV. TYPES OF RAMP METERING SYSTEMS

There are many strategies and types of ramp metering systems which are used in depending the needs and traffic conditions. They can be based on fixed time of action which is coordinated by previous gathered data about the traffic flow or by some simulated predictions.
The strategies can be implemented locally or in one global system. There are mainly three types of ramp metering systems:

-*Systems with fixed operating time*- these systems are the most simple for construction and installation of all types. Their function is to scatter the platoons of vehicles which are entering the freeway. The used technology is very simple and the whole system is not very expensive because there is no central managing system. The system cannot withstand the congestions caused by time factor and traffic accidents. This strategy is implemented on easily predictable situations in cases of road construction sites.

-*System with local operation based on feedback information*- this type of system for ramp metering is operating based on controllers and algorithms which are adjusting the degree of metering over analysis of the road occupancy and the traffic flow. The data is gathered by the detectors on the main highway flow [4]. This kind of systems is more expensive but they are capable to perform very well on unpredictable traffic changes. Sometimes this kind of system is set as backup when the main ramp metering system has failed. The costs are higher because there is need for additional placement of detectors which from other side are justified by the adaptability of the system and the better results compared to the system for fixed ramp metering.

-*Systems with global operation based on feedback information* -is type of ramp metering systems which aims to improve the traffic management over several ramps for metering in order to avoid congestions and bottlenecks. The structure of the system introduces metering doses on a ramp that are influenced by the conditions on the other ramps on the highway or particular part or corridor. Similar to the systems with local operation, the data is collected from various detectors on the ramp and from number of local detectors along the highway controlled by the system. The regulating os performed over central computer which is controlling the ramps. This kind of systems enables full control of different part of the traffic network from one center.. the systems can perform very strong metering at the particular segment of the highway where the traffic accident has occurred. This kind of configuration needs larger technical capacity which can operate and interconnect the detectors and the control center.

V. OPERATIONAL FACTORS

-*The distance between he place of metering and the place of traffic merging*
The vehicles are accessing the highway with different speed of movement. It is found that the vehicles needs a speed of about 60 km/h in order to perform smooth merging of the traffic in the ramp with the traffic on the highway. The distance to the merging point increases with the raising of the speed of merging and the degree of ramp metering.

-*Safe distance to stop and align*
The distance is limited by the construction of the system, the location of the signal and the detectors. The researches have recommended detector of the queue should be placed on a distance of 250 [5] feet after the merging in order to improve save braking behind the queue that is forming on the ramp. Basing on this information there is a generalized model for determining the queries for aligning of the vehicles on the ramp:

$$L = 0{,}820V - 0.002435V^2 \quad (1)$$

Where L is the total distance needed from the queue detector in meters, V is the expected value during the peak in vehicles per hour (vph).The distance for a ramp serving 900 vehicles (vph) and the required distance is about 541 fees.

-*The space on the ramp depending on the time of service*

The available space on the ramp should be long enough to obtain reasonable time of service during peak time and to provide high degree of metering.
The time that one vehicle is spending on the ramp S, depends on the length of the queue and the time of services T_s which is the time of one cycle. Therefore:

$$S = E(n) \times T_s = \frac{L(ft)}{25(ft/veh)} \times \frac{60(\min/hr)}{M(vphpl)} = 2.4 \frac{L(ft)}{M(vphpl)} \quad (2)$$

where:

$S =$ time of service on the ramp over one lane in minutes.

$E(n) =$ expected number of vehicles in the queue when the length is L in vehicles.

$L =$ length of the queue in feet

$M =$ degree of metering of vehicles in hour by lane (vphpl)

VI. ALINEA ALGORITHM

Asservissement Linéaire d'Entrée Auotroutière (ALINEA), is an algorithm which is used by the strategies for ramp metering on local base. It is represented in the late 80ies of the 20th century by Papageorgiu. This algorithm is implemented in several European cities like Paris, Munich, and Amsterdam. As a process that is considered is the traffic flow and the degree of metering is acting as control variable. This algorithm is based on the theory of control of feedback information where the algorithm is aims to maintain ramp metering on the freeway which will not exceed the capacity of the highway. For every time interval the algorithm is implementing the following formula [6] for the degree of ramp metering :

$$\left[r(k) = r(k-1) \pm K_R \{o_c - o_{out}(k)\}\right] \quad (3)$$

where:

$r(k) =$ metering value in the interval k

$K_R =$ variable parameter (weight factor) >0

$o_{out} =$ local occupancy of the ramp (measured by the detector on the main flow)

$o_c =$ predefined value of occupancy on the ramp capacity (the desired maximum capacity)

The algorithm is using the difference between the values of the occupancy or the desired occupancy compared to the measured occupancy. One of the characteristics of this algorithm is the possibility to insert the values of ramp metering from previous time periods.

VII. SIGNALISATION

The traffic signalization is placed on the ramps usually in two levels. The distance from the point of merging with the main traffic where the signalization is placed should be large enough to enable vehicle acceleration to a speed that is enough to perform smooth merge to the traffic flow. In some cases the light signals that are used are identical with the traffic lights placed on the intersections and in some other cases there is a special signalization for this purposes with ability to show more complex messages to the drivers.

VIII. DETECTORS

The devices that supply the system with data are the detectors which can be divided in to two groups. There are detectors on the main stream of the traffic flow and detectors which are placed on the ramp for metering. Particular highways are using detectors placed on the main stream on 500m distance between them and in other cases we have additional equipment that is making evaluation for the position of the detectors according to the traffic situation. It is important to point that the position of the detectors is set mainly by the algorithm that is used in the system of ramp metering.

IX. EVALUATION OF RAMP METERING SYSTEMS

The evaluation of ramp metering systems is done based on data gathered from realworld experiments or simulations [7]. In the particular case the data are obtained from the Paris highway network which is facing large congestions in the morning peak hours. In the first case there are no metering ramps, then we have ramps with fixed metering which were used in the morning hours. Later they are changed in a experiment which introduced ramp metering system based on ALINEA algorithm where each ramp acts locally basing its service on the situation of the traffic flow.
The idea is to collect and evaluate the data in order to note the impact of different ramp metering systems and to check their effectiveness.
The variables that have been used are taken in 6 minutes intervals [8] and they contain the following data: The volume of the traffic measured in vehicles x km, the time spent on a highway section, the average speed, time of travel all measured in different time periods.

X. . APPLICATION FOR GRAPHICAL DATA ANALYSIS

In order to compare the data gathered from the ramp metering system we will use an application called "*RAMP METER DATA COMPARISON*" which is an.NET application that allows us to enter all the gathered data form an experiment or simulation over particular time period of the day. There are three main cases: without existing ramp metering system, with fixed ramp metering and with ramp

metering that uses the ALINEA algorithm. The graphic (Fig 2) represents one of the important measurement and meaningful data in this case that is the total travel time in all three cases.

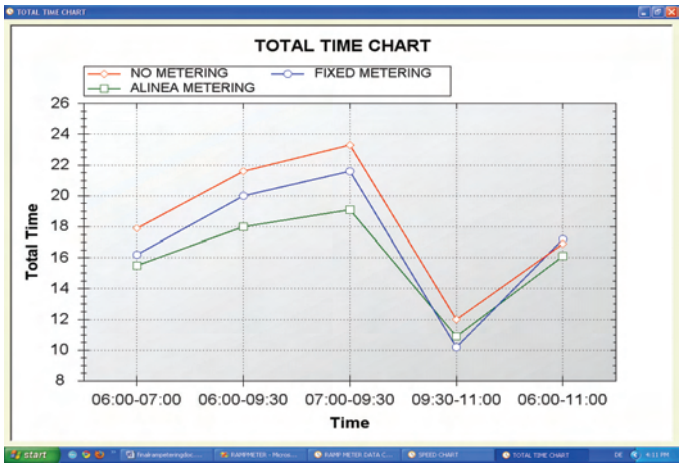


Fig. 2 time of traveling

XI. CONCLUSION

With the intensifying of the traffic and the number of vehicles it becomes necessary to manage the flow of the traffic. The large number of vehicles that are entering the fhighways is causing congestions and traffic accidents as a result of the irregular and dangerous traffic flow.

The use of ramp metering systems should reduce of the load on the highways which will lower the time of traveling and the costs of the trip. In many places in the world the ramp metering systems has been proved as the best solution for improving the traffic condition and avoiding congestions and queues [9]. This kind of systems could reduce the pollution of the air through avoiding frequent stop and go movements when congestions and queues occur.

With the process of regulating the highway access roads the number of platoon so vehicles that are accessing the main road especially in the morning peak hours will be reduced.

The data from the traffic condition are gathered, analyzed and inserted into different experiments and simulations. The simulation will show if the right metering system is chosen or another one will be more suited for implementation. Before the ramp metering systems can be enforced it is important to inspect the capacity and the needs of the traffic infrastructure. The investment in the traffic networks is very expensive so the right strategy should find the optimal solution for the traffic problems.

This is important to achieve the needed benefits and results form the ramp metering system with conduction of extensive analysis and experiments followed by the needed technology and capital which has to be well presented to the society and the authorities for supplying full support in the project implementation.

REFERENCES

[1] Peter Stewart, *Ramp metering Study*, 2003

[2] Lianyu Chu, Will Recker, Guizhen Yu, "*Integrated Ramp Metering Design and Evaluation Platform with Paramics*", California PATH Research Report

[3] Steven I-Jy Chien, Jiangtao Luo, "*evaluation of the potential for using ramp metering in the atms of the i-80 showcase corridor"*, 2007

[4] Markos Papageorgiou, Ioannis Papamichail , "*EURAMP European Ramp Metering Project, Handbook of Ramp Metering"*, 2007

[5] Nadeem A. Chaudhary, Zongzhong Tian, Carroll J. Messer, and Chi-Leung Chu, "*Ramp metering algorithms and approaches for Texas*", 2004

[6] Markos Papageorgiou, Elias Kosmatopoulos, Ioannis Papamichail, Yibing Wang, "*ALINEA maximizes motorway throughput– an answer to flawed criticism*",

[7] Jingcheng Wu, *"traffic diversion resulting from ramp metering"*, 2001

[8] Frederik Kunkel, "*ramp metering: applications in France", 2000*

[9] Levinson, D. and Lei Zhang, Ramp Meters on Trial: *Evidence from the Twin Cities Metering Holiday,* Department of Civil Engineering, University of Minnesota

Model for Calculation of Wasted Time at Pay-Toll Area

Ruben Nuredini[1], Jasmin Ramadani[2] and Emilija Kamcheva[3]

***Abstract* – In this paper, we try to model the natural behavior of traffic around pay-toll plazas. After that, we try to simplify traffic streams around pay-toll plazas by making few reasonable assumptions. We break-up the process in two stages: toll payment and lanes merging. In each stage we use Queuing Theory to model the queuing system. We derive a formula to calculate the average wasted time per driver. The average wasted time is dependent of the parameters of the system. By employing this formula, we calculate the average wasted time for a number of traffic lanes, traffic levels and pay-toll booths number.**

***Keywords* – Toll plaza, Queuing theory, Traffic model.**

I. INTRODUCTION

Toll financing has been used for years for various purposes. Some of them include providing finances for modern freeways and limiting vehicle flow in urban cores which encourages transit usage. Despite its advantages, toll financing has many drawbacks which depend of the way the toll is collected. In the traditional stop-pay-go pay-toll systems used in Macedonia the biggest issue is time spending as the vehicle waits in line to be served, energy spending as the vehicle decreases its speed to 0 and increases it back to permitted speed and pollution caused by unnecessary fuel combustion. The first logical solution for evading long lines hence decreasing waiting time is providing as many tollbooths as possible. This move may cause another disadvantage: the time spending during heavy traffic after toll payment, as the road decreases its width and the vehicles try to get back in line. Choosing the right number of tollbooths for getting an optimal balance between these two factors is crucial for best performance.

II. DEFINITION OF THE SYSTEM

Our system is based on the following assumptions:

- Constant traffic flow over a short time period
- The time between servicing two cars is of exponential distribution
- The traffic streams come into tollbooths evenly. We assume that exiting the tollbooth area does not contribute to delay. We ignore the delay caused by the behavior of irrational drivers.
- The drivers are delayed by waiting in lines for servicing. The driver has to wait for the other drivers that are in front of the line to be serviced first.
- The delay caused by waiting in lines is distributed exponentially.
- The drivers are delayed by the merging process after being serviced.
- The merging process is made with side merging (Fig.1).

Fig. 1: Side merging process

III. THE MODEL

According to our definition the delay in the system can be broken up in two parts: delay at the tollbooths, where drivers are waiting to be served, and delay at merging points, where drivers stop to wait to get into the merged lane. Here we use Queuing Theory and consider each part as a queuing system. Based on the assumptions, the traffic stream coming from the entrance of the toll plaza is evenly divided into each tollbooth and each tollbooth will receive a stream whose inter-arrival time is of exponential distribution. The service time also has exponential distribution. Therefore, each tollbooth can be considered as an independent M/M/1 queue. Burke's Theorem states that the outcoming stream also has exponential distribution with the same rate as the arrival stream [1]. We will also use this property in the queue model of merging points.

The delay caused by the merging process is more complicated to analyze. We first considered a simple 2-to-1 merging process. When a driver on one lane arrives at the merging point, the delay time depends on whether there is a car on the other lane. For simplification, we treated the two incoming lanes as one queue. We suppose that every driver must stop and wait whenever there is another car in the queue. We define the service time of a car as the time it spends through the merging area. According to this, the service rate is equal to when there is more than one car in the system, and

[1]Ruben Nuredini, Faculty of information-communication technologies, FON University, Vojvodina blvd. Skopje, Macedonia E-mail: ruben.nuredini@fon.edu.mk

[2]Jasmin Ramadani, Faculty of information-communication technologies, FON University, Vojvodina blvd. Skopje, Macedonia E-mail: jasmin.ramadani@fon.edu.mk

[3]Emilija Kamcheva, Faculty of information-communication technologies, FON University, Vojvodina blvd. Skopje, Macedonia E-mail: emilija.kamcheva@fon.edu.mk

when the system has only one car. Here, , and are constants that represent whether the driver does or does not have to give advantage to another car when merging. The service pattern of this queuing system is a general function. If the arrival pattern is exponential, the configuration of this queuing system would be M/G/1.

IV. Merging Process

The total merging process at the toll plaza is considered as a multiple 2-to-1 merging points. If there are tollbooths that are merged back into lanes, the number of merging points will be $T -$. However, this depends on the arrival rate at the merging point which equals the traffic flow it receives. If a merging point receives a traffic stream coming from tollbooths, with total flow of , its arrival rate would be:

$$\lambda = \frac{k}{T} \cdot \Phi \qquad (1)$$

The values of for merging points depend on the merging layout. For example, in a toll plaza that uses side merging layout and tollbooths, the first merging point takes the stream from 2 tollbooths, the second from 3 tollbooths etc.

The overall average wasted time is the weighted sum of all averaged wasted time at each merging point, where the corresponding weight is the probability for a driver to reach that point, which is k_i. The arrival rate and the corresponding probability at each merging point for a toll plaza with T tollbooths, N lanes at the exit that receives a total traffic flow is displayed in Table I.

Table I
Arrival Rate and Probability at Each Merging Point

Merging point	1st	2nd	3rd	...	(T-N)th
Arrival rate	$2\Phi/T$	$3\Phi/T$	$4\Phi/T$	...	$(T-N+1)\Phi/T$
Probability	2/T	3/T	4/T	...	(T-N+1)/T

V. Calculation

Here, we will calculate the formula for wasted time depending on number of tollbooths. The calculation is based on some constants which must be defined first:

- Number of incoming lanes ( )

The common number of lanes on a highway (in one direction) in Macedonia is 2. In other countries, this number may be from 1 to 6.

- Total traffic flow ( )

The maximum traffic flow per lane is 2000/hr [2]. With ranging from 1 to 6, various values for traffic flow will be considered, both for light and heavy traffic conditions.

- Service rate at the tollbooth ( )

The service rate at a traditional tollbooth system where person disburses change, issues receipts, etc. is about 350 vehicles per hour [3].

- Service rate at merging point: when merging does not happens ( )

This value represents the time for a vehicle to drive through the merging area at average highway speed. The average highway speed is 90km/h. The length of the merging area is the average car length plus the safety distance, which is [4] :

$$4{,}5\ [m] + 6 \cdot 4{,}5\ [m] = 31{,}5\ [m]$$

According to this, the average service time is:

$$31{,}5\ [m] / 90[km/h] = 1{,}26\ [1/s]$$

And the service rate:

$$\mu_0 = 3600/1{,}26 = 2857\ [1/\mathrm{h}].$$

- Service rate at merging point: when merging happens ( )

To estimate we use the same approach we used to estimate . We considered that a vehicle passes the same distance with initial speed of zero. On this speed, the safe distance is one car length, and the average acceleration of a vehicle is 1.98 [m/s2]. Therefore the average service time is:

$$\sqrt{(2 \cdot (4{,}5 + 4{,}5)/1{,}98)} = 3{,}015$$

And the service rate:

$$\mu_B = 3600 / 3{,}015 = 1194\ [1/$$

VI. Wasted Time

We divided the total wasted time in two parts: wasted time at the tollbooth and wasted time at the merging point.

The average wasted time of each tollbooth can be derived using the performance measure formula [1], given the arrival rate (Φ) and service rate ( ) is:

$$w_A = \frac{1}{\mu_A - \Phi/T} \qquad (2)$$

The merging point is modeled as a Markov system (Fig.2). Here, each state represents the number of vehicles in the system. The single represents the possibility of no conflict at the merging point. The arrival rate is represented with . We first need to find the average waiting time in the system t_{sys} (in order to calculate the average wasted time t_{diff} (.

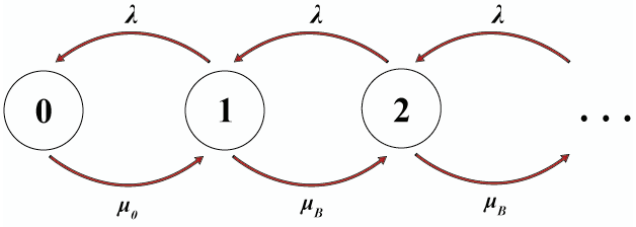


Fig. 2: State Diagram Representing the Queue at Merging Point

Let be the probability that there are drivers in the system. When the system reaches equilibrium, the net probability of transition is zero for each state. The sum of all must be one.

$$\begin{cases} \lambda P_0 = \mu_0 P_1 \\ \lambda P_1 + \mu_0 P_1 = \lambda P_0 + \mu_B P_2 \\ \lambda P_n + \mu_B P_n = \lambda P_{n-1} + \mu_B P_{n+1}, n \geq 2 \\ \sum_0^\infty P_i = 1 \end{cases} \quad (3)$$

The expected numbers of drivers in the system is:

$$L(\lambda) = \sum_0^\infty i P_i = \frac{\lambda}{\mu_B - \lambda} + \frac{\lambda(\mu_B - \mu_0)}{\lambda(\mu_B - \mu_0) + \mu_0 \mu_B} \quad (4)$$

By employing Little's law we obtained:

$$t_{sys}(\lambda) = \frac{L(\lambda)}{\lambda} = \frac{1}{\mu_B - \lambda} + \frac{\mu_B - \mu_0}{\lambda(\mu_B - \mu_0) + \mu_0 \mu_B} \quad (5)$$

The average wasted time of a driver at a merging point is the difference between t_i and the time the driver would spend on a normal lane. The time a driver spends when no merging occurs is .

$$t_{diff}(\lambda) = t_{sys}(\lambda) - \frac{1}{\mu_0} = \frac{1}{\mu_B - \lambda} + \frac{\mu_B - \mu_0}{\lambda(\mu_B - \mu_0) + \mu_0 \mu_B} - \frac{1}{\mu_0} \quad (6)$$

According to Table I, we calculated the corresponding probabilities for reaching the merging points that are 2/T, 3/T, 4/T or: $\frac{T-N}{T}$. Finally, the wasted time during the complete merging process is the weighted sum:

$$w_B = \sum_{i=1}^{T-N} \frac{i \mid 1}{T} \cdot t_{diff}\left(\frac{i \mid 1}{T} \cdot \Phi\right) \quad (7)$$

From Eqs. (2) and (7), we obtain the final equation for total wasted time in toll plaza:

$$w = w_A + w_B = \frac{1}{\mu_A - \Phi/T} + \sum_{i=1}^{T-N} \frac{i+1}{T} \cdot t_{diff}\left(\frac{i+1}{T} \cdot \Phi\right) \quad (8)$$

The total wasted time takes part of the total travel time through the toll plaza, so by minimizing it, the travelling time is also decreasing.

VII. Application

The Eq. (8) depends by the number of lanes , number of tollbooths and the traffic flow . By applying this equation for the specific case on the 2 lane highway between Gostivar and Tetovo (Republic of Macedonia), with 4 maximum available tollbooths we obtain the results in Fig. (3) and (4) which are very close to the results obtained by measurements conducted on the field.

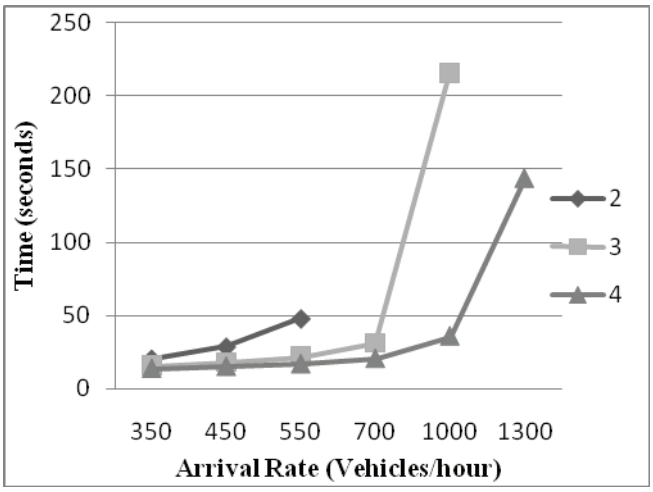


Fig. 3: Waiting time at a tollbooth for toll-plaza with 2, 3 and 4 tolbooths

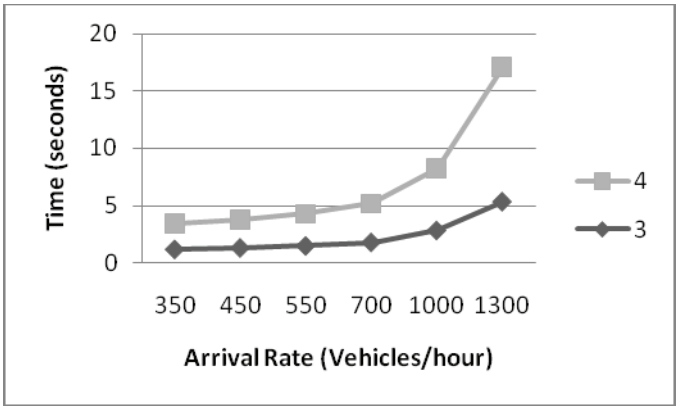


Fig. 4: Waiting time at merging point for a toll plaza with 3 and 4 tollbooths

It is obvious that, when the traffic intensity is greater, the waiting time at tollbooth is decreased by activating more tollbooths at the same time. On the other side, with this move, the waiting time at merging point is increased. The results for highways with more lanes and more tollbooths are more drastic.

References

[1] Hock, Ng Chee, "Queuing Modeling Fundamentals", Wiley, New York, 1996

[2] Homburger, Kell and Perkins, "Fundamentals of Traffic Engineering, 13th Edition", Institute of Transportation Studies, University of California (Berkeley), 1992

[3] Prevedouros, "Automated Tolls for Greece: Systems Review and Performance with AVC", Department of Civil Engineering, University of Hawaii at Manoa

[4] Wand, Liu and Montgomery, "A Simulation Model for Motorway Merging Behaviour", Institute for Transport Studies, University of Leeds

This page intentionally left blank.

Model for Control of Train Traffic on Junctions by Petri Net Simulation and Fuzzy Logic

Slavko Vesković[1], Sanjin Milinković[1], Ilija Tanackov[2], Norbert Pavlović[1], Slaviša Aćimović[1]

Abstract **- Railroad Junction is a place on the railway line where another line diverges. It is a complex system that is very difficult to describe and analyze by analytical or graphical methods. High Level Petri Nets - HLPN (timed, colored, stochastic and hierarchical) as a simulation tool can be used for graphical and mathematical modeling of various systems. HLPN model of junction comprises of interlocking and operating rules and input data for timetable, infrastructure and train characteristics. Model computes parameters of junction system and experiment by varying input data. HLPN have a good graphical presentation of model and can present results by animation or by tables and graphically by train time – distance graph. Trains primary delays are calculated by fuzzy logic system. Model is tested on a part of Belgrade railway node.**

Keywords **– Petri Nets, Railway Simulation, Fuzzy logic, Railway junction**

I. INTRODUCTION

Railroad Junction is a place on the railway line where another line diverges. It is a very complex system that is very difficult to describe with analytical models. Simulation model can compute parameters of junction in different operating conditions. Simulation tool must be able to make a model that will incorporate all interlocking and operating rules and data. Petri Nets are tool for graphical and mathematical modeling of various systems. High Level Petri Nets - HLPN are (timed, colored, stochastic and hierarchical) are tool that can model complex system and have a good graphical presentation of model.

Three different approaches are possible [1]: analytic methods, microsimulation methods and statistical analyses based on empirical data. Simulation models are detailed representation of a railway system and the only reasonable way to model in detail railway processes where different trains interact with each other and with the infrastructure. They require very detailed data about the infrastructure, the performance of the trains and, most importantly, about the timetable. If one of these data is unknown it is necessary to make assumptions and then results depend on the quality of input data.

Fuzzy logic is proved to be a good mathematical tool for modeling traffic processes that are distinguished by subjectivity, uncertainty, ambiguity and imprecision [2]. Many authors use advantage of predictive modeling systems with fuzzy logic. Fay [3] used fuzzy system as a dispatching support system for use in railway operation control systems. The model is defined as a fuzzy Petri net model that combines expert knowledge of fuzzy systems and graphical power of Petri Nets, making the model easy to design, test, improve and maintenance. Cheng and Yang [4] proposed fuzzy Petri Net model that will use professional knowledge of a dispatchers to create database rules to be applied for testing the system in case of disorder.

Fuzzy logic model for calculation of trains primary delays on junction use the experience and knowledge of railway personnel who directly participate in regulating the traffic in system.

[1]University of Belgrade - Faculty of Transport and Traffic Engineering, Vojvode Stepe 305, Belgrade, Serbia, E-mail: veskos@sf.bg.ac.rs

[2]Faculty of Technical Science, University of Novi Sad, Trg Dositeja Obradovića 6, Novi Sad, Serbia, E-mail: ilijat@uns.ac.rs

II. PETRI NETS AS A TOOL FOR MODELING OF A JUNCTION SYSTEM

Petri nets are a mathematical tool for modeling used for analysis and simulation of concurrent systems [5]. The theory of Petri nets is based on a mathematical theory of bipartite graphs. A bipartite graph (or bigraph) is a graph which nodes can be divided into two disjoint sets V_1 and V_2 such that every edge connects a node in V_1 to one in V_2; that is, there are no two identical nodes in the same set. Petri net is one of several mathematical descriptions of discrete distributed systems. The system is modeled as a bipartite directed graph with two sets of nodes: the set of places which represent state or system objects and the set of events or transitions that determine the dynamics of the system.

Petri Net model of junction

In Petri net model *places* represent sections, *transitions* represent conditions for train movement, and *tokens* represent trains. Hierarchy of the model enables defining insulated section as a subsystem or module. Insulated section can be block section, switch section and station track section. A module is defined for each distinctive section. Model is created by positioning and connecting modules according to the railway line section plan. Although this approach requires more time for initial programming, it allows using defined modules for modeling systems with similar processes. Module for generating trains imports timetable data from an external database for generating tokens (trains). Also, each token is filled with information about train it represent (train number,

category, time of entering into the system, train route etc.). Defined tokens leave the module when the simulation clock and time of train departure match. Module of block section (Fig. 1.) represents a block section on an open line. The module contains *places*, *transitions*, storages for storing parameters and objects for connecting with other modules. Transitions in module enable or forbid entering and leaving the section based on the storage data. Storages contain information about state of the connected sections and signals and simulation clock. When the transition *trainin* enables firing, token are placed in *sectionbusy*. At the same moment, information about the occupancy of the section is sent to the previous two sections and to the signal. Token remains in place while conditions defined in the transition *trainout* are met:

- time needed for train to cross section - section occupancy time (train traveling time on section),
- next section is occupied,
- and signal does not allow further movement.

When the conditions are met, transition is allowed; the signal is set to allow train movement to next section; token leaves the section module; additional token is firing to the place *sectionfree;* and information about leaving the section are sent to connected modules. The purpose of storages in module is keeping data about section state. Data is used in transition processor for imposing logical conditions and for section journey time calculation. There are several types of *storages*: *time*, *info* and *ceka*. The other type of storages (*sectionstate*) serves for gathering data sent by transitions. These storages have information about the state of signals and sections. During the simulation, data from storages is sent to an external database.

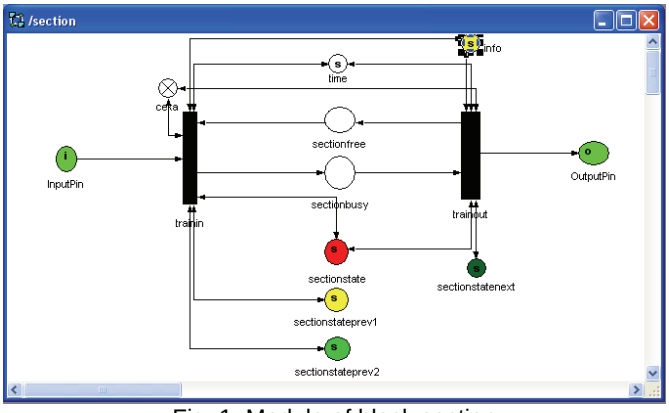


Fig. 1. Module of block section

Modules for other specific sections are similar to those already described, with difference in implementation of the train traffic rules dependent on the location of the section in the system.

After defining the modules, creating the junction simulation model requires connecting modules according to the layout of sections in railway line plan. Then, modules should be connected with storages (Fig. 2.).

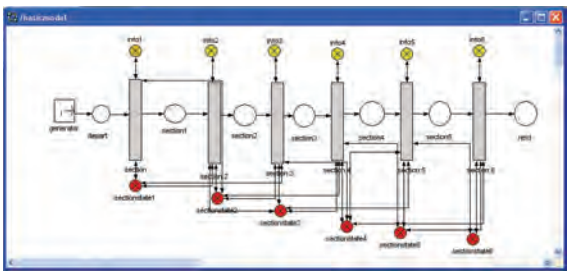


Fig. 2. Creation of the model by connecting section modules

Model is connected to external databases for storing input and output data. Output data is classified and filtered for statistical analysis, as well as for graphical representation of the simulation results. The simulation program gives data about movement of each train through the model, as well as data about section state (total and physical occupancy of each section). The database is customized for creating quick reports based on queries and for filtering data by train, section, signals or train delay time in the model. Data can be presented in tables or graphs, and can be easily validated.

Animation of the simulation run is done in the simulation program itself, animating section states in the model. During the simulation, parameters (from storages) of each section change when a train enters the section. These numerical data are used for animation of sections. Numerical value from storage gives visual identity of the section occupancy symbol in animation window.

III. MODEL FOR GENERATING TRAIN DELAYS BY FUZZY LOGIC

Train delays have great influence on timetable and on technological processes related to train traffic, but also on railways planning procedures. There are primary and secondary delays [6]. Primary delays are timetable disorders caused by disturbed train operation. It is not always possible to avoid primary delays, because of many factors that generate them. Secondary delays are result of “spreading“ the primary delays through the timetable, e.g. transferring the delay factors from one train to the others. Secondary (or knock-on) delays and their propagation are subject to primary delays, railway network structure and train timetable.

Fuzzy set theory is a suitable mathematical approach to modeling processes characterized by subjectivity, uncertainty, ambiguity and imprecision, which makes it a very good tool for modeling timetable disorders (primary delays). Fuzzy logic enables making decisions even when they are based on imprecise information. Models based on fuzzy logic are defined by set of IF-THEN rules. Input variables are linguistic variables which describes current operating conditions for each train. Defuzzified output result of fuzzy inference system is train delay. Values of fuzzy model parameters are defined in collaboration with traffic dispatchers, operators and experts familiar with functioning of the system. Their knowledge and experience, as well as train delay statistics, are used for defining input variables, rules base and output variables.

Fuzzy model is defined with four input variables: train category, timetable influence, train mileage and infrastructure

influence (Fig. 3.). The train category and probability of train delay are highly dependent. This is defined with a description from 0 to 10 where numerical scores can be matched by words. The lowest score 0 is given to a freight train (μ_F), score 5 is regional train (μ_R), and score 10 is high category passenger train (μ_P). The membership function for train category is:

$$\mu_F(x)=\begin{cases}1 & x\le 0\\ 1-\frac{x}{4} & 0\le x\le 4\\ 0 & x\ge 4\end{cases} \qquad \mu_R(x)=\begin{cases}0 & x\le 1\\ \frac{x}{4}-\frac{1}{4} & 1\le x\le 5\\ -\frac{x}{4}+\frac{9}{4} & 5\le x\le 9\\ 0 & x\ge 9\end{cases}$$

$$\mu_P(x)=\begin{cases}0 & x\le 6\\ \frac{x}{4}-\frac{6}{4} & 6\le x\le 10\\ 1 & x\ge 10\end{cases}$$

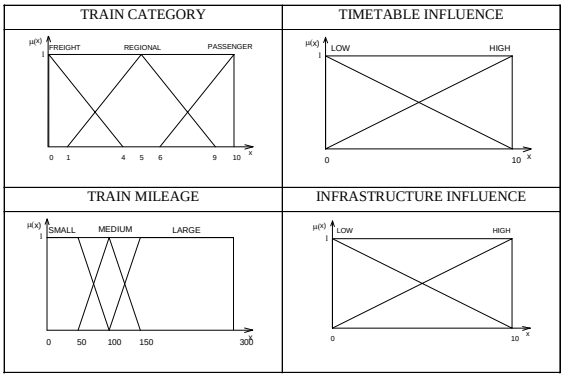


Fig. 3. Membership functions of input parameters

The membership function for output variable train delay (Figure 6) is defined with 5 fuzzy sets: very small (μ_{VS}), small (μ_S), medium (μ_M), high (μ_H) and very high delay (μ_{VH}) (Fig. 4.).

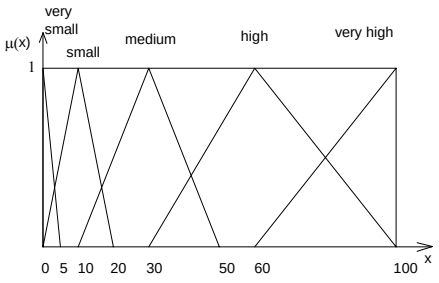


Fig. 4. The membership function of fuzzy set of output variable (train delay)

Fuzzy logic system comprises of 36 rules. These fuzzy rules translate 4 input fuzzy variables into 1 output fuzzy variable. Logical AND operator is employed (the rule of minimum for AND relationships, the so-called Mamdani rule of minimum). In creating the consequent fuzzy set, MAX – MIN inference is used. Defuzzification of the output fuzzy variable is by center of gravity method (COG). Fig. 5. shows relationship between train delay and selected input variables: train category and train distance.

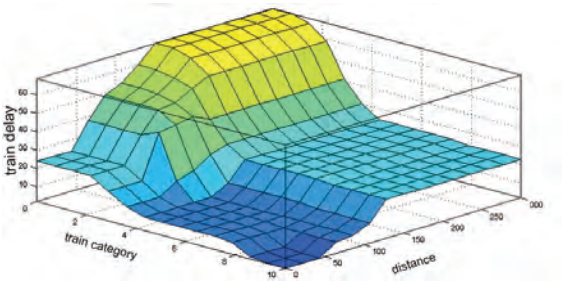


Fig. 5. Relationship between input and output fuzzy variables

Fig. 6. presents an example of fuzzy inference system of model. Test case is for following input parameters:

- train category – 5 (regional train);
- timetable influence – score 4;
- train distance traveled – 40 km;
- infrastructure influence – score 2.

Four rules are related to this set of input data: rule 13, 14, 19 and 20. Defuzzified output variable gives calculated train delay: 7.15 minutes.

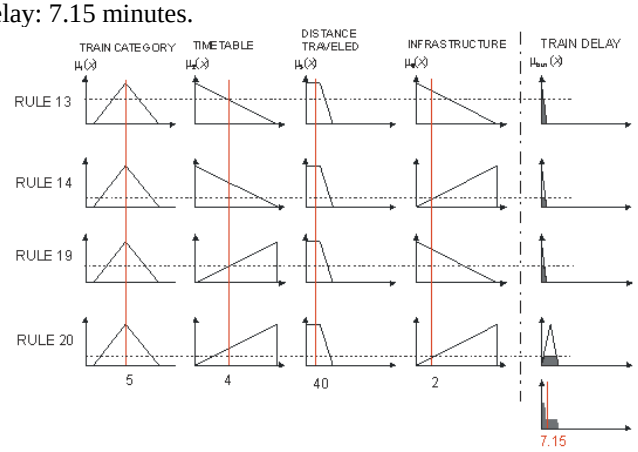


Fig. 6. Example of fuzzy inference system

IV. PETRI NET MODEL EXAMPLE

Junction G in Belgrade railway node is chosen as an example of the Petri Net simulation model. It is a complex double-track diamond crossing junction. The boundaries of the model are stations Belgrade Center, Topcider, Rakovica and Karadjordjev Park. Model is defined with data from section and railway signals layout plan, as well as from timetable for year 2008.

Petri net Model of the junction G is shown in Fig. 7.

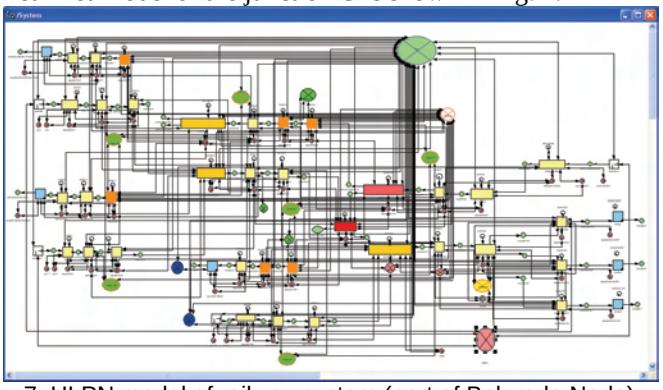


Fig. 7. HLPN model of railway system (part of Belgrade Node)

The simulation can be observed by viewing:

- token movement through the Petri Net Graph,
- storage state indicators and
- animation window which shows the state of each section in the model.

The model allows experimenting by inserting new tokens in *places* during simulation (marking of a Petri Net model). After the simulation, the results can be viewed in the external database and subsequently edited in the form of reports, graphs, sketches, and tables.

Verification and validation of Petri Net junction model

Validation and verification of Petri Net simulation model can be done by:

- monitoring if tokens movement is according to operating rules;
- inspecting storage indicators;
- animation window which shows the state of each section in the model (Fig. 8.);
- data analysis. Simulation results are stored in database where reports can be generated.
- Train time–distance diagram generated from simulation results (Fig. 9.). Any irregularities and conflictsituations in model can be easily identified on a train diagram.

Fig. 8. Animation window indicating state of sections

Animation window is modeled to visually resemble to dispatcher interlocking control panel. Sections of model are animated by using section state data located in storage.

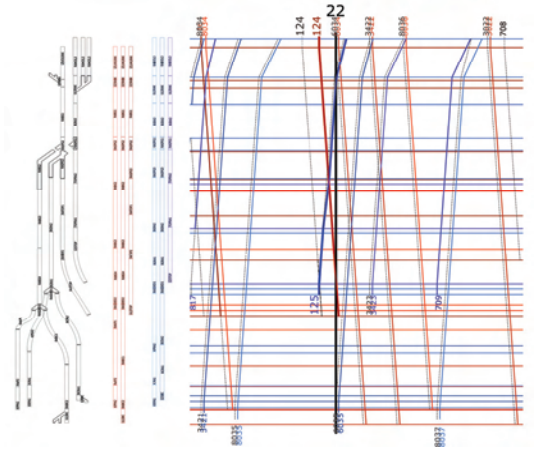


Fig. 9. Time - distance graph of train simulation

V. Results of a Petri Net Model

Results can be classified and filtered for analysis in external base that collects simulation data. Multiple queries can be used to generate reports. Reports can assist at timetable analysis, analysis of train delay by train category, by section etc. Database can generate reports for comparing results of experimenting with model by changing timetables, section layout, junction design or introducing new train operating rules policy.

Average section occupancy data can indicate possible bottlenecks of system (Fig.10.)

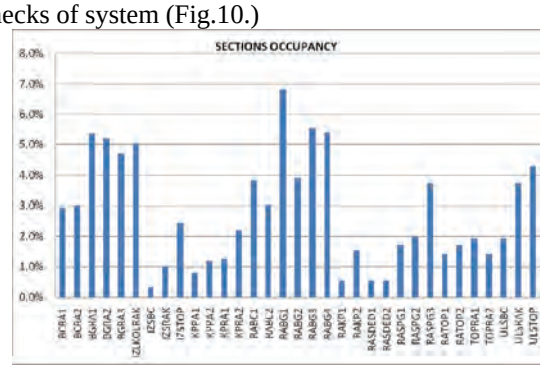


Fig. 10. Percentage occupancy of sections in model

Sections with highest percentage of physical occupancy are sections between junction and stations, and their occupancy is highest because of their length. Fig. 9. shows a part of train time – distance diagram for sections of a model. Dashed line present train routes by timetable and continuous line (colors represent different train categories) represent train routes with supplemented primary delay calculated in fuzzy logic system

Simulation results could be utilized for:

- decision making in investment projects,
- testing infrastructure or technological design of junction, station, sidings, etc.,
- analyzing effects from experimenting with train operation rules,
- analyzing timetables,
- testing railway line capacity,
- analyzing train delays.

REFERENCES

[1] Mattsson, L.-G. (2007), Railway capacity and train delay relationships, 129-150, in Murray, A. and Grubesic, T.H. (eds), Critical Infrastructure: Reliability and Vulnerability, Springer-Verlag

[2] Teodorovic, D. (1999). Fuzzy logic systems for transportation engineering: the state of the art. Transportation Research Part A, 33, 337-364

[3] Fay, A. (2000)., Fuzzy knowledge-based system for railway traffic control, Engineering Applications of Artificial Intelligence, 13 (6), 719-729.

[4] Cheng, Y.-H. , & Yang, L.-A. (2008), A Fuzzy Petri Nets approach for railway traffic control in case of abnormality, Expert Systems with Applications doi:10.1016/j.eswa.2008.10.070

[5] Murata T. Petri nets: Properties, analysis and applications, Proceedings of the IEEE, Vol.77, No.4, 541-580

[6] Landex, A., Kaas, A. H. and Hansen, S.,(2006). Railway Operation, Technical report at Centre for Traffic and Transport, Technical University of Denmark,

[7] Hansen, I. A.; Pachl, J. (2008). Railway Timetable & Traffic. Analysis - Modeling – Simulation. Eurailpress, Hamburg,

Management of Road and Rail Freight Transport with the View of Energy Efficiency in Serbia

Siniša R. Sremac[1], Svetlana D. Bašić[2], Jovan Đ. Tepić[3] and Bojan J. Matić[4]

***Abstract* – This paper defines the characteristics of road and rail freight transport in Serbia in terms of energy efficiency. Energy consumption per unit of work is very unfavourable, as a result of a number of factors that indicate the irrational use of energy.**

Significant possibility to boost energy efficiency of the transport sector in Serbia is in management of road and rail transport using modern technology of combined transport.

***Keywords* – energy efficiency, combined transport.**

I. INTRODUCTION

In The systematic study of energy efficiency of road and rail freight transport subsystem is done in order to get realistic assessment of the level of energy efficiency of the transportation system in Serbia. Transport participates with 26-30% in total energy consumption [5]. Analyzed the energy consumption by means of transport it is note that road transport contributes with 86%, air transport with 9%, rail transport with and 4%, and river transport with 1%.

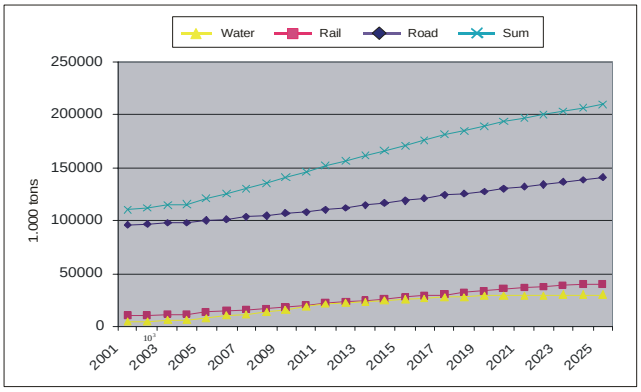


Fig. 1. Prognosis the volume transport of goods by means of transport

Over the past ten years, contribution of road transport significantly increases on the transport market in Serbia (Fig. 1). Bearing in mind the characteristics of transport in Europe, as well as information and knowledge about participation of some forms of transport in the territory of Serbia, it is necessary in further research to focus attention on reroute of commodity flows from road to rail and water transport, more exactly applied technology of combined transport.

The aim of this research is management of road and rail freight traffic using the combined transport in the function of energy efficiency.

II. ENERGY EFFICIENCY OF ROAD FREIGHT TRANSPORT IN SERBIA

A. Road transport

All the indicators of road freight transport are related to the public transport, as individual transport is not recorded by the competent institutions. In the Fig. 2 it is obvious that the transport of goods by road transport has increased nearly three times in net tone kilometers.

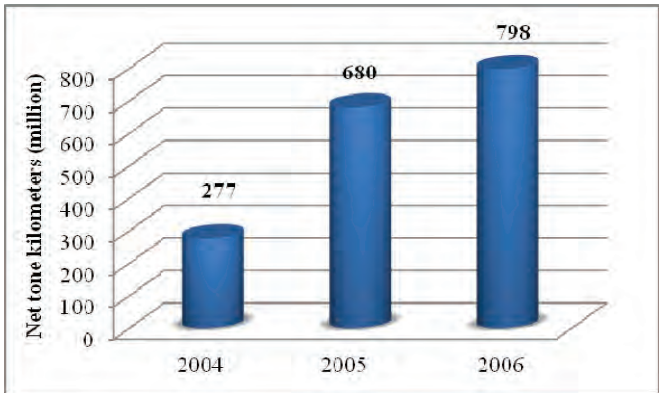


Fig. 2. The work in the road freight transport

Total energy consumption in the year 2006 increased by 48,71% if compared to the year 2004, and reduced by 10,31% if compared to the year 2005, and this can be seen in Fig. 3. [3]

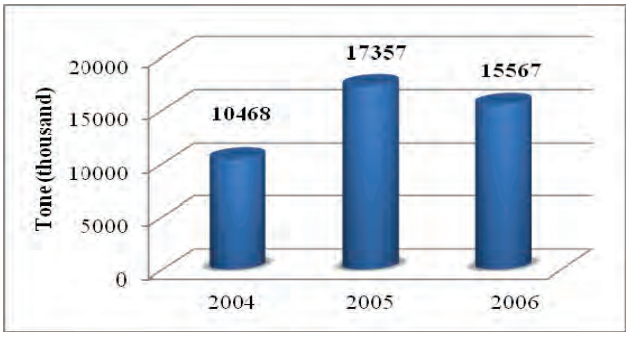


Fig. 3. Energy consumption in the road freight transport

[1]Siniša R. Sremac is with the Faculty of Technical Sciences, Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia, E-mail: sremacs@uns.ac.rs

[2]Svetlana D. Bašić is with the Faculty of Technical Sciences, Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia, E-mail: basic@uns.ac.rs

[3]Jovan Đ.Tepić is with the Faculty of Technical Sciences, Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia, E-mail: jot@nspoint.rs

[4] Bojan J. Matić is with the Faculty of Technical Sciences, Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia, E-mail: bojanm@uns.ac.rs

Energy efficiency of the road freight transport in Serbia is shown in Fig. 4. Specific energy consumption per unit of work has a downward trend, which means that energy efficiency increased in the observed period. The average specific consumption in freight road transport is 0,027 kg/ntk [5].

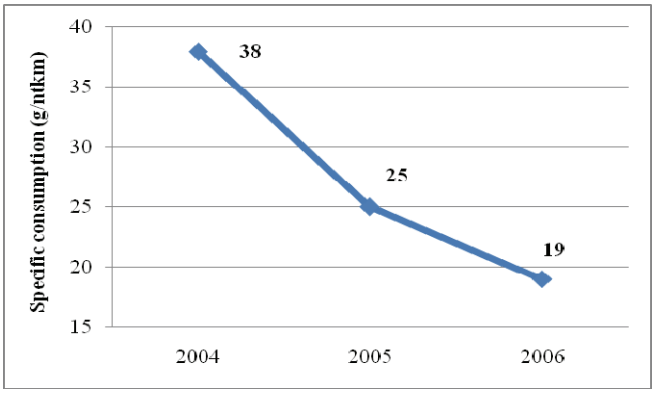


Fig. 4. Energy efficiency of the road freight transport

B. Rail freight transport

Volume of the transport of goods in the Serbian Railways increased in the period since 2004 to 2006 by 33,8% in net tone kilometers (Fig. 5).

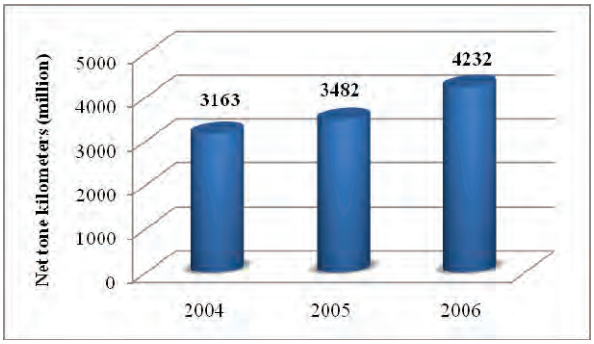


Fig. 5. Energy efficiency of the road freight transport

Energy consumption in freight rail transport consists of the consumption of diesel fuel for diesel traction, and electricity consumption for electric traction.

The specific consumption of diesel traction ranges from 8,84 to 9,42 g/ntk (Table I).

TABLE I: THE SPECIFIC CONSUMPTION OF THE DIESEL TRACTION

Year	Consumption liquid fuel (10^3 g)	Work achieved (ntkm)	Specific consumption (g/ntkm)
2004	6 404 660	706 402 912	**9,07**
2005	7 083 143	801 274 629	**8,84**
2006	8 676 078	921 078 411	**9,42**

In the Table II we can see that the specific consumption of electric traction is in the range of 3,77 to 4,25 g/ntk. The specific fuel consumption in 2006 decreased by 11,29% compared to 2004.

TABLE II: THE SPECIFIC CONSUMPTION OF THE ELECTRIC TRACTION

Year	Energy consumption (kWh)	Work achieved (ntkm)	Specific consumption (kWh/ntkm)	Specific consumption (g/ntkm)
2004	121 979 559	2 456 724 005	0,0496	**4,25**
2005	131 826 121	2 679 392 721	0,0492	**4,21**
2006	146 141 860	3 309 875 670	0,0441	**3,77**

The specific energy consumption per unit of work in the railway transport has a trend of reduction in the electric traction, while the diesel traction increases. The increase of energy efficiency of the electric traction has influenced the greater degree of utilization of freight cars for cargo transport [5].

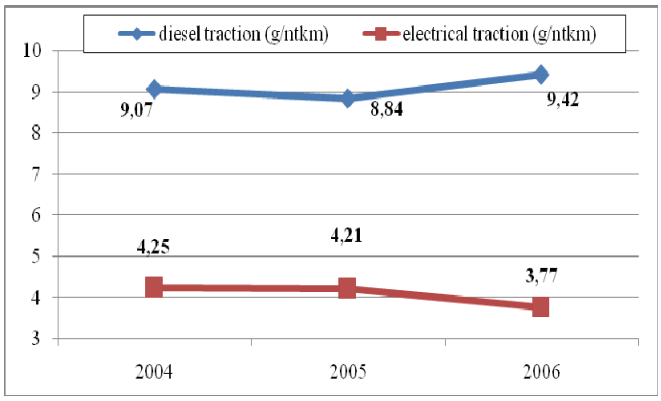


Image 5: Energy efficiency of the railway freight transport

Image 5 shows that the electrical traction energy is more efficient than the diesel traction, and that the specific consumption per unit of transport work is more than twice smaller. One of the basic problems of energy efficiency of the railway transport is that only 32% of the total railway network in Serbia (4.347 km) is electrified. Therefore, one of the priority measures would be to increase the number of the electrified railways.

III. IMPROVING THE ENERGY EFFICIENCY OF TRANSPORT USING THE COMBINED TRANSPORT IN SERBIA

Energy consumption in Serbia is much higher than in developed countries so there are significant opportunities to reduce it.

To increase energy efficiency in transport, it is necessary to perform the integration of some forms of transport, diversion of commodity flows to the energy efficient forms of transport such as rail or water. Application of combined transport requires development and production of new means of transport, use of alternative fuels, introduction of information technology in managing and monitoring transport vehicles and etc.

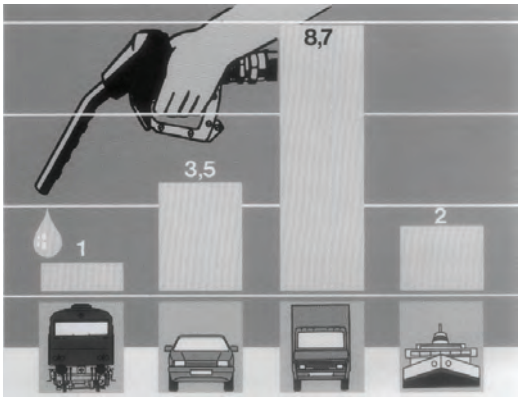


Image 1. Energy consumption some mood of transport [1]

Conducted research and the best European experience have shown that the specific energy consumption in the current transportation system with introduction of technology designed for optimal combined transport can be reduced by 5.8% [2]. This is even more important if we bear in mind that the reduction in specific energy consumption, as a rule, means the reduction of air pollution.

Combined transport in Serbia until 2005. participated in the overall transport with about 0.5%, and at the European Union it participated with 6 to 9% [2]. Given the size of the countries in the region and the limited potential for this type of transport, it is very important to cooperate with neighbouring countries, as strategic, based on regional and bilateral initiatives for the combined transport, as well as operational by operator for the combined transport. With the recovery of the economy in the region and international transport processes the importance and role of combined transport has increased.

The study found that the use of modern technology of combined transport in Serbia, on the whole, is insignificant and it is primarily related to intercontinental transport of container. There are only three terminals, which have a certain capacity for handling transport units of the combined transport: ŽIT -Belgrade railway terminal, Container terminal in the Port of Belgrade and the Container terminal at the port of Pančevo. Presence of other technologies for combined transport (Hucke-pack, Ro-Ro and etc.) at the total commodity flow is negligibly small, primarily because of the lack of appropriate infrastructure elements of the transportation system (the terminals for the handling and transportation unit for combined transport) [3].

Terminal capacity should be developed gradually, according to market demands. It is necessary to ensure that the terminals, according to the needs, except for transfer Container are also capable for handling other units of combined transport (for example exchange transport vessels and semi-trailers), as well as complete road vehicles.

Special attention should be pay on the environmental aspects of combined transport. In fact, this technology of transport should contribute significantly reduction of the harmful effects of transport on the environment.

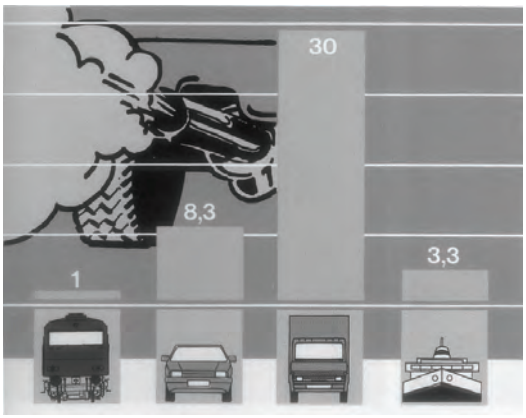


Image 2. Impact some mood of transport in emissions of exhaust gases [1]

Rail traffic is able to use electricity as a fuel and thereby make a huge transport work. Electrical energy is the cleanest fuel and there is no necessary to stored it, also it can be obtained from renewable natural resources. It least from all other sources of energy threatens the natural environment, which provides to railway undoubted advantages over all other forms of transport, especially road transport.

IV. CONCLUSION

Energy efficiency of road and rail transport in Serbia is very unfavourable, as a result of a number of factors that point to the irrational use of energy. The main causes of low energy efficiency of road transport are outdated road network, lack of bypass, obsolete technology traffic management, traffic congestion, unfavourable age structure of vehicles, inadequate vehicle capacity utilization, etc. The causes of lower energy efficiency of rail transport are outdated infrastructure, low capacity of railway lines, old vehicles and low capacity utilization.

Combined transport, as a transport of the wider public interest, is environmentally acceptable, economically justified and safe. Bearing in mind the current state of the energy consumption of the traffic system in Serbia, it is necessary for the government to conduct "aggressive" policy of application of combined transport in order to increase the energy efficiency of transportation. The intensity of activity in the combined transport in particular it will be increase with formation of institutions in this area, removing bottlenecks in the corridors, construction of terminals and logistics centres, improving the organization of all participants in the combined transport chain and the implementation of stimulating measures by the government.

REFERENCES

[1] Planco Cusulting Studie ABB HENSHCEK ag, Manheim, Germany, 2001.

[2] The development and application of optimal controlling system for monitoring energy efficiency in transport, Republican program program of Energy efficiency, The project number 293005, Ministry of Science and Technology Republic of Serbia, 2006-2008.

[3] Development and application of optimal technology of combined transport in the function of the energy efficiency in transport as well as logistical support to increase the competitiveness of our economy, Republican program of Energy efficiency, The project number 293002, Ministry of Science and Technology Republic of Serbia, 2006-2009.

[4] Statistical Yearbook of Republic of Serbia for the 2004, 2005. and 2006. year, the Statistics Institute.

[5] Strategy of Energy Development Republic of Serbia until 2015., Belgrade, 2004.

[6] S. Sremac, S. Bašić, I. Tanackov, "Energy efficiency of road and rail freight transport in Serbia", Junior Scientist Conference, pp. 367-368, Vienna University of Technology, Austria, 2010.

AUTHOR INDEX