

**XLVII INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION,  
COMMUNICATION AND ENERGY SYSTEMS AND TECHNOLOGIES**



**iCEST**

**PROCEEDINGS OF PAPERS**

**VOLUME 1**

**SOFIA, 2012**

**ICEST 2012 Proceedings of the XLVII International Scientific Conference on  
Information, Communication and Energy Systems and Technologies**  
organized by the Faculty of Telecommunications, Technical University  
of Sofia, June 28-30, 2012, Veliko Tarnovo, Bulgaria

*Proceedings of Papers:*      *Volume 1 of 2 volumes*  
*Editor:*                              *Prof. Rumen Arnaudov, PhD*  
*Published by:*                      *Faculty of Telecommunications*  
*Printed by:*                              *Publishing Company, TU-Sofia*

*All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic, or mechanical, including photocopying or any information storage and the retrieval system now known or to be invented, without written permission from the Publisher.*

**ISBN: 978-619-167-002-4**

## WELCOME TO ICEST 2012



Dear Colleagues,

First of all I would like to thank you for attending our conference, in this way you contribute to the improvement of the scientific value of this conference and make it one of the most popular conferences in Balkan region. At the same time this conference turns to be a tool for spreading our scientific achievements with the help of colleagues from Japan, USA, United Kingdom, France and many more. For second year in addition to the founders of this conference: Faculty of Telecommunications, Technical University of Sofia, Technical Faculty, University "St. Kl. Ohridski", Bitola and Faculty of Electronics, University of Nish, we enjoy the participation and hospitability of Faculty of Mathematics and Informatics, "St. Cyril and St. Methodius" University of Veliko Tarnovo and on the behalf of all of us I would like to express our deep gratitude for their support.

This year there are 169 contributions with 352 authors from Bulgaria, Macedonia, Serbia, Canada, Portugal, Romania, Hungary, Finland and Germany to be presented at our conference. The submissions are evaluated by 74 reviewers from Bosnia and Herzegovina, Bulgaria, Macedonia, Romania, Serbia and USA.

Dear colleagues and guests, I wish you successful participation in the conference. I believe that this conference will go forward in raising its standard and influence and it will be helpful in the scientific development and maturity of our young colleagues. I believe that here we will witness the birth of new ideas for joint projects and future scientific achievements, because we all work in the same direction – the development of technical and technological foundations of future economy.

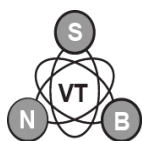
Good luck to all participants!

A handwritten signature in black ink, appearing to read 'R. Arnaudov'. The signature is stylized and cursive.

*Prof. Dr. Rumen Arnaudov*  
ICEST 2012 Conference General Chairman



# XLVII INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION, COMMUNICATION AND ENERGY SYSTEMS AND TECHNOLOGIES



# iCEST 2012

*Organized by:*



**Technical University of Sofia**  
**Faculty of Telecommunications**  
**Bulgaria**



**St. Cyril and St. Methodius University of Veliko Tarnovo**  
**Faculty of Mathematics and Informatics**  
**Bulgaria**



**University "St. Kliment Ohridski" of Bitola**  
**Faculty of Technical Sciences**  
**Macedonia**



**University of Niš**  
**Faculty of Electronic Engineering**  
**Serbia**

## TECHNICAL PROGRAM COMMITTEE

### *General Chairman:*

**Arnaudov R.** Technical University of Sofia, Bulgaria

### *Vice Chairmen:*

**Milovanović B.** University of Niš, Serbia

**Mitrovski C.** University "St. Kliment Ohridski" Bitola, Macedonia

### *Members:*

**Acevski N.** University "St. Kliment Ohridski" Bitola, Macedonia

**Atanasov I.** Technical University of Sofia, Bulgaria

**Bekiariski Al.** Technical University of Sofia, Bulgaria

**Bock J.** University of Ottawa, Canada

**Boumbarov O.** Technical University of Sofia, Bulgaria

**Ceselkoska V.** University "St. Kliment Ohridski" Bitola, Macedonia

**Dimitrov K.** Technical University of Sofia, Bulgaria

**Dochev I.** Technical University of Sofia, Bulgaria

**Dončov N.** University of Niš, Serbia

**Iliev G.** Technical University of Sofia, Bulgaria

**Iliev I.** Technical University of Sofia, Bulgaria

**Janković D.** University of Niš, Serbia

**Janković N.** University of Niš, Serbia

**Jeftić M.** University of Niš, Serbia

**Jolevski I.** University "St. Kliment Ohridski" Bitola, Macedonia

**Jordanova L.** Technical University of Sofia, Bulgaria

**Kostov M.** University "St. Kliment Ohridski" Bitola, Macedonia

**Makal J.** Technical University of Byalistok, Poland

**Marković V.** University of Niš, Serbia

**Markovski A.** University "St. Kliment Ohridski" Bitola, Macedonia

**Mitrevski P.** University "St. Kliment Ohridski" Bitola, Macedonia

**Nakamatsu K.** University of Hyogo, Japan

**Nedelkovski I.** University "St. Kliment Ohridski" Bitola, Macedonia

**Nikolov T.** Technical University of Sofia, Bulgaria

**Nikolova B.** Technical University of Sofia, Bulgaria

**Nikolova Zl.** Technical University of Sofia, Bulgaria

**Pencheva E.** Technical University of Sofia, Bulgaria

**Perić Z.** University of Niš, Serbia

**Pleshkova Sn.** Technical University of Sofia, Bulgaria

**Popova A.** Technical University of Sofia, Bulgaria

**Poulkov Vl.** Technical University of Sofia, Bulgaria

**Radevska P.** University "St. Kliment Ohridski" Bitola, Macedonia

**Stanković R.** University of Niš, Serbia

**Stanković Z.** University of Niš, Serbia

**Stefanović M.** University of Niš, Serbia

**Stefanovski M.** University "St. Kliment Ohridski" Bitola, Macedonia

**Stojčev M.** University of Niš, Serbia

**Stojmenov L.** University of Niš, Serbia

<b>Tasić D.</b>	University of Niš, Serbia
<b>Todorov G.</b>	St. Cyril and St. Methodius University of Veliko Tarnovo, Bulgaria
<b>Todorova M.</b>	St. Cyril and St. Methodius University of Veliko Tarnovo, Bulgaria
<b>Trpezanovski Lj.</b>	University "St. Kliment Ohridski" Bitola, Macedonia
<b>Tsenov A.</b>	Technical University of Sofia, Bulgaria
<b>Valtchev S.</b>	NOVA-University, Lisbon, Portugal
<b>Zieleznik L.</b>	Brookes University Oxford, UK

## ORGANIZING COMMITTEE

### *Chairman:*

<b>Arnaudov R.</b>	Technical University of Sofia, Bulgaria
--------------------	---

### *International coordinators:*

<b>Todorov G.</b>	St. Cyril and St. Methodius University of Veliko Tarnovo
<b>Milovanović B.</b>	University of Niš, Serbia
<b>Mitrovski C.</b>	University "St. Kl. Ohridski", Bitola, Macedonia

### *Local coordinators:*

<b>Todorova M.</b>	St. Cyril and St. Methodius University of Veliko Tarnovo
<b>Poulkov V.</b>	Technical University of Sofia, Bulgaria

### *Members of the organizing committee:*

<b>Iliev G.</b>	Technical University of Sofia, Bulgaria
<b>Dochev I.</b>	Technical University of Sofia, Bulgaria
<b>Goleva R.</b>	Technical University of Sofia, Bulgaria
<b>Dimitrov K.</b>	Technical University of Sofia, Bulgaria
<b>Koleva P.</b>	Technical University of Sofia, Bulgaria
<b>Tsankova J.</b>	Technical University of Sofia, Bulgaria
<b>Stanković Z.</b>	University of Niš, Serbia
<b>Milijić M.</b>	University of Niš, Serbia
<b>Dimitrijević T.</b>	University of Niš, Serbia
<b>Pargovski J.</b>	University "St. Kl. Ohridski", Bitola, Macedonia
<b>Petkovski M.</b>	University "St. Kl. Ohridski", Bitola, Macedonia

## CONFERENCE SECRETARIAT

<b>Tsankova J.</b>	Conference Coordinator
<b>Dochev I.</b>	Technical Secretariat

### *Members:*

<b>Nikolova M.</b>	Technical University of Sofia, Bulgaria
<b>Stoyanova K.</b>	Technical University of Sofia, Bulgaria
<b>Kirilova A.</b>	Technical University of Sofia, Bulgaria
<b>Ivanova M.</b>	Technical University of Sofia, Bulgaria
<b>Popova K.</b>	Technical University of Sofia, Bulgaria
<b>Markova G.</b>	St. Cyril and St. Methodius University of Veliko Tarnovo
<b>Kalushkov T.</b>	St. Cyril and St. Methodius University of Veliko Tarnovo
<b>Markov A.</b>	St. Cyril and St. Methodius University of Veliko Tarnovo

*Address:*

Technical University of Sofia  
Faculty of Telecommunications  
Kl. Ohridski Blvd. 8, 1000, Sofia, Bulgaria  
Phone/Fax: (+359 2) 965 30 95  
E-mail: [fktt-dekan@tu-sofia.bg](mailto:fktt-dekan@tu-sofia.bg)  
[icest@tu-sofia.bg](mailto:icest@tu-sofia.bg)

## CONFERENCE INTERNET SITE

For further information, please visit the Conference Internet Site: <http://www.icestconf.org>



## BRIEF ICEST HISTORY

The ICEST Conference is successor of a series of conferences started in 1963 at Technical University of Sofia under the name "Day of the Radio". In 1977 the name of the Conference was changed into "Communication, Electronic and Computer Systems". Since 2000 has become an international conference organized jointly by The Faculty of Telecommunications, Sofia and the Faculty of Technical Sciences, Bitola.

In 2001 the conference was renamed EIST (Energy and Information Systems and Technologies). In 2002 the Faculty of Electronic Engineering, Niš, Serbia became a co-organizer of the conference and the conference changed its name into ICEST (International Scientific Conference on Information, Communication and Energy Systems and Technologies).

This year the conference is organized by the Faculty of Telecommunications at the Technical University of Sofia and Faculty of Mathematics and Informatics of St. Cyril and St. Methodius University of Veliko Tarnovo.

## CONFERENCE TOPICS

- [1] Radio Communications, Microwaves, Antennas
- [2] Telecommunication Systems and Technology
- [3] Signal Processing
- [4] Digital Image Processing
- [5] Computer Systems and Internet Technologies
- [6] Informatics and Computer Science
- [7] Electronics
- [8] Energy Systems and Efficiency
- [9] Control Systems
- [10] Measurement Science and Technology
- [11] Remote Ecological Monitoring
- [12] Engineering Education



# TABLE OF CONTENTS

## VOLUME 1

### RADIO COMMUNICATIONS, MICROWAVES, ANTENNAS – PART 1

<b>Multiuser IR-UWB System Performance</b> .....	1
Razvan Craciunescu, Simona Halunga, Octavian Fratu <i>University Politehnica of Bucharest, Romania</i>	
<b>Doppler Fading Effects on OFDM Transmissions</b> .....	5
Ioana Bucsa, Razvan Craciunescu, Simona Halunga, Octavian Fratu <i>University Politehnica of Bucharest, Romania</i>	
<b>Investigation of the modulation type's influence on the DVB-T signals quality</b> .....	9
Oleg Panagiev <i>Technical University of Sofia, Bulgaria</i>	
<b>Improving the reception of class DVB-T receivers</b> .....	13
Oleg Panagiev <i>Technical University of Sofia, Bulgaria</i>	
<b>Monolithic Integrated Antennas with High Radiation Efficiency</b> .....	17
Hristomir Yordanov <i>Technical University of Sofia, Bulgaria</i>	
<b>SCP-RPSC – The New Technology for Microwave Broadband Mobile Communications</b> .....	21
Veselin Demirev <i>Technical University of Sofia, Bulgaria</i>	
<b>Study on Hybrid FSO/RF Systems Availability Depending on the Meteorological Conditions</b> .....	25
Tsvetan Mitsev, Maxim Shupak, Boncho Bonev <i>Technical University of Sofia, Bulgaria</i>	
<b>Cylindrical Mesh TLM Model of Probe-Coupled Cavity Loaded with Planparallel Dielectric Layers</b> .....	29
Tijana Dimitrijevic, Jugoslav Jokovic, Bratislav Milovanović <i>University of Niš, Serbia</i>	

### RADIO COMMUNICATIONS, MICROWAVES, ANTENNAS – PART 2

<b>Implementation of pseudo random noise generator in FPGA for Free Space Optics BER testing</b> .....	33
Nikolay Kolev, Tsvetan Mitsev <i>Technical University of Sofia, Bulgaria</i>	
<b>Experimental Estimation and Correction of the Methods for Radio Waves Attenuation Prediction in Rain</b> .....	37
Boncho Bonev, Kliment Angelov, Emil Altimirski <i>Technical University of Sofia, Bulgaria</i>	
<b>Multiresolution Analysis of Multiple Reflections in Transmission Lines</b> .....	39
András Fehér, Ádám Békefi, Szilvia Nagy <i>Széchenyi István University, Hungary</i>	

<b>Ad-Hoc Supported, Connection Fault-Tolerant Model for Mobile Distributed Transaction Processing .....</b>	<b>43</b>
Tome Dimovski, Pece Mitrevski <i>University "St. Kliment Ohridski" of Bitola, Macedonia</i>	
<b>Fast Synthesis of High Order Microwave Filters by Coupling Matrix Optimization.....</b>	<b>47</b>
Marin Nedelchev, Ilija Iliev <i>Technical University of Sofia, Bulgaria</i>	
<b>Random High Voltage Impulses Modeling for EMC Testing.....</b>	<b>51</b>
Kliment Angelov, Miroslav Gechev <i>Technical University of Sofia, Bulgaria</i>	
<b>Optimum Divergence of the Transmitter Optical Radiation in FSO Systems.....</b>	<b>55</b>
Tsvetan Mitsev, Nikolay Kolev, Hristo Ivanov, Kalin Dimitrov <i>Technical University of Sofia, Bulgaria</i>	
<b>Body Bias Influence on Ring Oscillator Performance for IR-UWB Pulse Generator in 0.18<math>\mu</math>m CMOS technology .....</b>	<b>59</b>
Jelena Radic, Alena Djugova, Laszlo Nagy, Mirjana Videnovic-Misic <i>University of Novi Sad, Serbia</i>	

## TELECOMMUNICATION SYSTEMS AND TECHNOLOGY – PART 1

<b>Comparative Performance Studies of Laboratory WPA IEEE 802.11b,g Point-to-Point Links.....</b>	<b>63</b>
José Pacheco de Carvalho, Cláudia Pacheco, Hugo Veiga, António Reis <i>University of Beira Interior, Portugal</i>	
<b>Customer Satisfaction based Demand Analysis of Mobile Services .....</b>	<b>67</b>
Aleksandar Tsenov <i>Technical University of Sofia, Bulgaria</i>	
<b>Investigate common work of software phone systems in virtual environments and real switching systems .....</b>	<b>71</b>
Borislav Necov, Krasen Bankov, Mario Georgiev <i>Technical University of Varna, Bulgaria</i>	
<b>Analysis of current methods and technologies for encoding, distribution and consumption of IPTV services.....</b>	<b>74</b>
Jordan Kanev, Stanimir Sadinov <i>Technical University of Gabrovo, Bulgaria</i>	
<b>Average SIR Comparison for SC Systems Using Different Decision Algorithms in the Presence of Interference .....</b>	<b>77</b>
Aleksandra Panajotović, Dragan Drača, Nikola Sekulović* <i>University of Niš, Serbia</i> *School of Higher Technical Professional Education, Serbia	
<b>Optimization of Traffic Distribution Coefficients in IP Radio-Relay Network with Path Diversity .....</b>	<b>81</b>
Dragana Perić, Miroslav Perić*, Branislav Todorović**, Milan Šunjevarić**, Miroslav Popović** <i>IMTEL Komunikacije a.d., Serbia</i> *VLATACOM d.o.o., Serbia **Institute for Computer Based Systems, Serbia	
<b>Optical Receiver Sensitivity Evaluation in Presence of Noise in Digital Communication System .....</b>	<b>85</b>
Krasen Angelov, Stanimir Sadinov, Nataliya Varbanova <i>Technical University of Gabrovo, Bulgaria</i>	

## TELECOMMUNICATION SYSTEMS AND TECHNOLOGY – PART 2

<b>New Teletraffic Loss System – Polya/G/n/0 .....</b>	<b>89</b>
Seferin Mirtchev, Rossitza Goleva, Georgi Balabanov, Velko Alexiev <i>Technical University of Sofia, Bulgaria</i>	
<b>An Evaluation of an UMTS/WLAN Interworking Architecture using IEEE 802.21 .....</b>	<b>93</b>
Alexandru Vulpe, Octavian Fratu <i>University Politehnica of Bucharest, Romania</i>	
<b>Simulation of Rare Events in Teletraffic Systems with Single Queue .....</b>	<b>97</b>
Elena Ivanova, Rostislav Raev, Dimitar Radev <i>University of Ruse "Angel Kanchev", Bulgaria</i>	
<b>VoIP over a Cognitive Network with Limited Availability.....</b>	<b>101</b>
Yakim Mihov, Boris Tsankov <i>Technical University of Sofia, Bulgaria</i>	
<b>BEP Performance of DE-QPSK and DE-OQPSK over composite fading channels in the presence of imperfect signal extraction.....</b>	<b>105</b>
Milica Petković, Bojana Nikolić, Bata Vasić, Goran Đorđević <i>University of Niš, Serbia</i>	
<b>Quality of Service (QoS) – main principles and managing tools .....</b>	<b>109</b>
Miroslav Slavov, Pencho Penchev <i>Technical University of Gabrovo, Bulgaria</i>	
<b>Pitch perception in complex sound .....</b>	<b>113</b>
Marko Janković, Dejan Ćirić <i>University of Niš, Serbia</i>	

## SIGNAL PROCESSING

<b>Digital Bandpass IIR Filers with High Selectivity .....</b>	<b>117</b>
Peter Apostolov <i>College of Telecommunications and Posts, Bulgaria</i>	
<b>Features of time-frequency analysis visualization of large dynamic range signals.....</b>	<b>121</b>
Tihomir Trifonov, Ivan Simeonov*, Rosen Dzhakov** <i>St. Cyril and St. Methodius University of Veliko Tarnovo, Bulgaria</i> *Vasil Levski National Military University, Bulgaria **Vasil Levski National Military University - Shumen, Bulgaria	
<b>Accuracy Improvement of Allpass-based Digital Hilbert Transformers .....</b>	<b>125</b>
Kamelia Nikolova, Georgi Stoyanov <i>Technical University of Sofia, Bulgaria</i>	
<b>Acoustic Standing Waves in Closed Cylindrical Enclosures .....</b>	<b>129</b>
Ekaterinoslav Sirakov, Hristo Zhivomirov <i>Technical University of Varna, Bulgaria</i>	
<b>Control of Radiation Directivity Applying Independent Element Dodecahedral Loudspeaker.....</b>	<b>131</b>
Marko Jelenković, Dejan Ćirić, Jelena Zdravković, Stefan Tomić <i>University of Niš, Serbia</i>	
<b>Modulated bandpass Farrow Decimators and Interpolators .....</b>	<b>135</b>
Djordje Babic, Vesa Lehtinen* <i>Union University, Serbia</i> *Tampere University of Technology, Finland	

<b>Simulation of Codec for Adaptive Linear Prediction .....</b>	<b>139</b>
Rumen Mironov <i>Technical University of Sofia, Bulgaria</i>	

## DIGITAL IMAGE PROCESSING

<b>Image Compression with Inverse Pyramid Decomposition over Wavelet Spectrum.....</b>	<b>143</b>
Teodora Sechkova, Ivo Draganov <i>Technical University of Sofia, Bulgaria</i>	
<b>Efficient Adaptive Local Binarization Algorithm for Text Extraction from Image with Complex Background .....</b>	<b>147</b>
Antoaneta Popova <i>Technical University of Sofia, Bulgaria</i>	
<b>Text Skew Detection using Log-polar Transformation .....</b>	<b>151</b>
Darko Brodić, Zoran Milivojević*, Dragan Milivojević** <i>University of Belgrade - Bor, Serbia</i> <i>*Technical College Niš, Serbia</i> <i>**Institute for Mining and Metallurgy, Serbia</i>	
<b>Directional Transforms Applicability in Image Coding .....</b>	<b>155</b>
Ivo Draganov <i>Technical University of Sofia, Bulgaria</i>	

## COMPUTER SYSTEMS AND INTERNET TECHNOLOGIES

<b>An Algorithm and a program module for calculating the border height of the mass centre of a vessel.....</b>	<b>159</b>
Emiliya Koleva, Mariya Nikolova, Mariya Eremieva, Viktoriya Sabeva <i>Naval Academy "Nikola Vaptsarov", Bulgaria</i>	
<b>Power consumption analysis of fault tolerant real-time systems .....</b>	<b>163</b>
Sandra Djosic, Milun Jevtic, Milunka Damnjanovic <i>University of Niš, Serbia</i>	
<b>Challenges of Personalization and Collaboration Learning Process by Using Blogs .....</b>	<b>167</b>
Teodora Bakardjieva, Boyka Gradinarova* <i>Varna Free University, Bulgaria</i> <i>*Technical University of Varna, Bulgaria</i>	
<b>Implementation of Web 2.0 in the Bitola Museum - Successful Marketing Tool .....</b>	<b>171</b>
Pargovski Jove, Irena Ruzin*, Aleksandra Lozanovska** <i>Cultural Heritage Protection Office, Macedonia</i> <i>*NI Institute and Museum Bitola, Macedonia</i> <i>**Gauss Institute, Macedonia</i>	
<b>Attacking the cloud .....</b>	<b>175</b>
Vlad-Andrei Poenaru, George Suciu, Cristian-George Cernat, Gyorgy Todoran, Traian-Lucian Militaru <i>University Politehnica of Bucharest, Romania</i>	
<b>An Implementation of SMS Communication with Patients in a Medical Information System .....</b>	<b>178</b>
Ivica Marković, Aleksandar Milenković, Dragan Janković <i>University of Niš, Serbia</i>	

<b>A Comparative Analysis of Mobile AR Software with the Application to the Archeological Site Medijana</b> .....	182
Dušan Tatić, Časlav Stefanović, Dragan Stanković*	
<i>University of Niš, Serbia</i>	
*University of Pristina Kosovska Mitrovica, Serbia	
<b>A comparative analysis of dynamic programming languages for application in multi-agent systems</b> .....	186
Ana Stankovic, Dragan Stanković*, Dušan Tatić**	
<i>Metropolitan University, Serbia</i>	
*University of Pristina Kosovska Mitrovica, Serbia	
**University of Niš, Serbia	

## INFORMATICS AND COMPUTER SCIENCE – PART 1

<b>GPU Accelerated Construction of Characters of Finite Abelian Groups</b> .....	190
Dušan Gajić, Radomir Stanković	
<i>University of Niš, Serbia</i>	
<b>Modern Processor Architectures Overview</b> .....	194
Danijela Jakimovska, Aristotel Tentov, Goran Jakimovski, Sashka Gjorgjievska, Maja Malenko	
<i>University Ss Cyril &amp; Methodius in Skopje, Macedonia</i>	
<b>The mechanism for flexible symbology in mobile GIS</b> .....	198
Miloš Roganović, Bratislav Predić, Dragan Stojanović, Marko Kovačević	
<i>University of Niš, Serbia</i>	
<b>GinisED tools for spatial analysis of electric power supply network</b> .....	202
Aleksandar Stanimirović, Leonid Stoimenov, Danilo Vulović	
<i>University of Niš, Serbia</i>	
<b>Computer Methods and New Values for Cut Set Catalan Numbers</b> .....	206
Iuliana Dochkova-Todorova	
<i>St. Cyril and St. Methodius University of Veliko Tarnovo, Bulgaria</i>	
<b>Accelerating Strategies in Evolutionary Algorithms</b> .....	208
Vassil Guliashki, Leoneed Kirilov	
<i>Bulgarian Academy of Science, Bulgaria</i>	

## INFORMATICS AND COMPUTER SCIENCE – PART 2

<b>2D Weather product visualization using Marching Squares algorithm</b> .....	212
Igor Antolović, Dejan Rančić, Vladan Mihajlović, Dragan Mihić*, Marija Đorđević*	
<i>University of Niš, Serbia</i>	
*Republic Hydrometeorological Service of Serbia, Serbia	
<b>Efficient Implementation of BDD Packages on the GPU Platform</b> .....	216
Miloš Radmanović, Dušan Gajić	
<i>University of Niš, Serbia</i>	
<b>Architecture of Distributed Multiplatform GIS for Meteorological Data Analysis and Visualization</b> .....	220
Marko Kovačević, Aleksandar Milosavljević, Vladan Mihajlović, Dejan Rančić	
<i>University of Niš, Serbia</i>	

## ELECTRONICS

<b>Tracking Analogue to Digital Converter Modelling using VHDL-AMS</b> .....	224
--	-----

Marieta Kovacheva, Ivailo Pandiev  
*Technical University of Sofia, Bulgaria*

**Development of Parameterized Verilog-AMS Model of Photovoltaic Cells .....228**

Elissaveta Gadjeva, Georgi Valkov  
*Technical University of Sofia, Bulgaria*

**Optical Control through Stencils Cutting in Surface Mount Technology .....232**

Valentin Videkov, Aleksei Stratev, Georgi Furkov  
*Technical University of Sofia, Bulgaria*

## ENERGY SYSTEMS AND EFFICIENCY

**Power Quality According to EN 50160 .....235**

Nikolce Acevski, Kire Mijoski, Tomce Mijoski  
*University "St. Kliment Ohridski" of Bitola, Macedonia*

**Using  $H_{\infty}$  synthesis for finding settings of single channel power system stabilizers of synchronous generators .....239**

Konstantin Gerasimov, Petko Petkov\*, Krum Gerasimov  
*Technical University of Varna, Bulgaria*  
*\*Technical University of Sofia, Bulgaria*

**Technical Conditions for PV Plants Connection on the MV Distribution Grids in the Republic of Macedonia .....243**

Ljupco Trpezanovski, Metodija Atanasovski, Dimitar Dimitrov\*  
*University "St. Kliment Ohridski" of Bitola, Macedonia*  
*\*University Ss Cyril & Methodius in Skopje, Macedonia*

## CONTROL SYSTEMS

**Control cards. Control cards and control points as part of the manufacturing process .....247**

Violina Georgieva, Alexander Hadjidimitrov\*  
*Technical University of Sofia, Bulgaria*  
*\*Team VISION Bulgaria Ltd., Bulgaria*

**Computer Simulation and Analysis of Two-Coordinate Position Electric Drive Systems .....251**

Mikho Mikhov, Marin Zhilevski  
*Technical University of Sofia, Bulgaria*

## MEASUREMENT SCIENCE AND TECHNOLOGY

**Automated multichannel broadband spectrum analysis of fiber-optic grating sensors .....255**

Plamen Balzhiev, Wojtek Bock\*, Tinko Eftimov\*\*, Rumen Arnaudov  
*Technical University of Sofia, Bulgaria*  
*\*Université du Québec en Outaouais, Canada*  
*\*\*University of Plovdiv, Bulgaria*

**Measurement of the Position by Using Hybrid Pseudorandom Encoder .....259**

Dragan Denić, Goran Miljković, Jelena Lukić, Miodrag Arsić, Milan Simić  
*University of Niš, Serbia*

**Method for calculating the stability at moderate and big heeling angles of a vessel .....263**

Mariya Eremieva, Viktoriya Sabeva, Mariya Nikolova, Emiliya Koleva  
*Naval Academy "Nikola Vaptsarov", Bulgaria*

**Design of a high – sensitive capacitive sensor for wireless monitoring of bulk material's level.....265**

Teodora Trifonova, Valentina Markova, Valentin Todorov, Ventseslav Draganov  
*Technical University of Varna, Bulgaria*

**Different Implementations of Serial Pseudorandom/Natural Code Converters .....269**

Dragan Denić, Goran Miljković, Jelena Lukić, Miodrag Arsić, Dragan Živanović  
*University of Niš, Serbia*

## REMOTE ECOLOGICAL MONITORING

**Cloud systems for environmental telemetry – A case study for ecological monitoring in agriculture .....273**

George Suciu, Octavian Fratu, Cristian Cernat, Traian Militaru, Gyorgy Todoran, Vlad Poenaru  
*University Politehnica of Bucharest, Romania*

## ENGINEERING EDUCATION

**High-quality Primary School Education in the Field of Electrotechnics and Informatics - Beginning of the Development of Successful Engineers .....277**

Sonja Cvetkovic, Zoran Stankovic\*  
*Primary school "Cele Kula", Serbia*  
*\*University of Niš, Serbia*

# VOLUME 2

## RADIO COMMUNICATIONS, MICROWAVES, ANTENNAS (POSTER)

**Sensor Network Topology as Low-interference Factor .....283**

Vasil Dimitrov, Rozalina Dimova, Paskal Novakov  
*Technical University of Varna, Bulgaria*

**A Proposal for Harmonic Rejection Mixer Avoiding Irrational Weighting Ratios .....287**

Ludwig Lubich  
*Technical University of Sofia, Bulgaria*

**Efficient Estimation of the Antenna Noise Level Using Neural Networks .....291**

Ivan Milovanovic, Zoran Stankovic\*, Marija Milijic\*  
*Singidunum University, Serbia*  
*\*University of Niš, Serbia*

**Software for Automated Measuring Pattern Diagrams of Wide Frequency Bands Antennas with Integrated Receivers .....295**

Dragan Obradović, Igor Stančić, Aleksandar Kopta, Zoran Mičić, Predrag Manojlović  
*IMTEL Komunikacije a.d., Serbia*

**Numerical Model of Enclosure with Receiving Dipole Antenna for Shielding Effectiveness Calculation .....299**

Tatjana Cvetković, Vesna Milutinović, Nebojša Dončov\*, Bratislav Milovanović\*  
*Republic Agency for Electronic Communications, Serbia*  
*\*University of Niš, Serbia*

<b>Reliability of Radio-Relay Systems .....</b>	<b>303</b>
Nataša Bogdanović, Dejan Blagojević, Dragiša Milovanović*	
<i>School of Higher Technical Professional Education, Serbia</i>	
<i>*University of Niš, Serbia</i>	
<b>New Architectural Solutions to Improve the CATV System Performances .....</b>	<b>307</b>
Lidia Jordanova, Dobri Dobrev, Kalin Dimitrov	
<i>Technical University of Sofia, Bulgaria</i>	
<b>Measurement Site and Procedures for Experimental 2D DOA Estimation .....</b>	<b>311</b>
Marija Agatonovic, Zoran Stankovic, Bratislav Milovanović, Nebojša Dončov	
<i>University of Niš, Serbia</i>	
<b>Simulation influence of the thermal noise of PIN photodetector on performance DWDM optical network .....</b>	<b>315</b>
Petar Spalević, Dejan Milić*, Branimir Jakšić, Mile Petrović, Ilija Temelkovski*	
<i>University of Pristina Kosovska Mitrovica, Serbia</i>	
<i>*University of Niš, Serbia</i>	
<b>Study of ICI in PRS-OFDM systems .....</b>	<b>319</b>
Stanio Kolev, Ilija Iliev, Stoicho Manev	
<i>Technical University of Sofia, Bulgaria</i>	
<b>BER simulation analysis of PRS-OFDM systems with MLSD.....</b>	<b>321</b>
Ilija Iliev, Stanio Kolev, Stoicho Manev	
<i>Technical University of Sofia, Bulgaria</i>	

## TELECOMMUNICATION SYSTEMS AND TECHNOLOGY (POSTER)

<b>A New Modified Algorithm for Multi Constraint Routing .....</b>	<b>323</b>
Yavor Tomov, Georgi Iliev	
<i>Technical University of Sofia, Bulgaria</i>	
<b>A Schema Based Approach to Access Network Discovery and Selection in EPS .....</b>	<b>327</b>
Ivaylo Atanasov	
<i>Technical University of Sofia, Bulgaria</i>	
<b>System for thermal comfort monitoring in working and living environment .....</b>	<b>331</b>
Uros Pesovic, Dusan Markovic, Zeljko Jovanovic, Sinisa Randjic	
<i>University of Kragujevac, Serbia</i>	
<b>Some Integral Characteristics of MRC Receiver in Nakagami-m fading Environment .....</b>	<b>335</b>
Hana Stefanovic, Dejan Milić*, Dimitrije Stefanovic*, Srdjan Milosavljevic**	
<i>College of Electrical Engineering and CSAS, Serbia</i>	
<i>*University of Niš, Serbia</i>	
<i>**University of Pristina Kosovska Mitrovica, Serbia</i>	
<b>Adaptive Filtering Algorithms Suitable for Real-Time Systems.....</b>	<b>339</b>
Maria Nenova	
<i>Technical University of Sofia, Bulgaria</i>	
<b>Presentation of a model to study facsimile coded signals from a fourth group.....</b>	<b>343</b>
Todorka Georgieva	
<i>Technical University of Varna, Bulgaria</i>	
<b>Design and implementation of a device for a cloudiness measurement.....</b>	<b>347</b>
Cvetan Kitov, Bulgaria	
<b>Classification and comparative analysis of localization approaches for Wireless Sensor Networks.....</b>	<b>351</b>
Ivanka Tsvetkova, Plamen Zahariev, Georgi Hristov, Mihail Iliev	



*University of Ruse "Angel Kanchev", Bulgaria*

**Comparative analysis of routing approaches for wireless sensor networks.....355**

Plamen Zahariev, Georgi Hristov, Ivanka Tsvetkova, Mihail Iliev

*University of Ruse "Angel Kanchev", Bulgaria*

**Upstream Power Control for Digital Subscriber Lines Based on Role Game**

**Approach .....359**

Pavlina Koleva, Oleg Asenov\*, Vladimir Poulkov

*Technical University of Sofia, Bulgaria*

*\*St. Cyril and St. Methodius University of Veliko Tarnovo, Bulgaria*

## SIGNAL PROCESSING (POSTER)

**A possibility for Edge detection using LabVIEW graphical programming environment.....363**

Liljana Docheva

*Technical University of Sofia, Bulgaria*

**Implementation of Analog Neural Networks with Labview .....367**

Liljana Docheva, Alexander Bekjarski

*Technical University of Sofia, Bulgaria*

**Noise-Resistance Performance Estimation of a Chaos Shift Keying Signals .....371**

Galina Cherneva, Georgi Pavlov, Elena Dimkina

*Todor Kableshkov University of Transport, Bulgaria*

**On a combination of amplitude and frequency modulation used for processing speech signals in cochlear implants.....375**

Svetlin Antonov, Snejana Pleshkova-Bekjarska

*Technical University of Sofia, Bulgaria*

**Safe Operating Area Limitations in Class B Amplifiers .....379**

Hristo Zhivomirov

*Technical University of Varna, Bulgaria*

**Complex Criterion for Linearity Segment Detection in the Subtraction Procedure for Removing Power-line Interference from ECG .....383**

Georgy Mihov

*Technical University of Sofia, Bulgaria*

## DIGITAL IMAGE PROCESSING (POSTER)

**School promoting tool Multimedia project – Documentary film .....387**

Ilche Acevski, Valentina Acevska, Mimoza Jankulovska, Igor Nedelkovski

*University "St. Kliment Ohridski" of Bitola, Macedonia*

**Animation of shadow -- Advantages and disadvantages when rendering 3D project.....391**

Valentina Acevska, Ilche Acevski, Igor Nedelkovski

*University "St. Kliment Ohridski" of Bitola, Macedonia*

**Investigation of Mixture of Gaussians Method for Background Subtraction in Traffic Surveillance .....395**

Boris Nikolov, Nikolay Kostov, Slava Yordanova

*Technical University of Varna, Bulgaria*

**Applied Aspects In Static Images Processing .....399**

Gergana Markova

*St. Cyril and St. Methodius University of Veliko Tarnovo, Bulgaria*

## COMPUTER SYSTEMS AND INTERNET TECHNOLOGIES (POSTER)

- Management of Software Project using Genetic Algorithm .....403**  
Milena Karova, Nevena Avramova, Ivaylo Penev, Yulka Petkova  
*Technical University of Varna, Bulgaria*
- Railway Infrastructure Maintenance Efficiency Improvement by Using Tablet PCs.....407**  
Slobodan Mitrović, Svetlana Čičević, Slađana Janković, Norbert Pavlović, Slaviša Aćimović, Snežana Mladenović, Sanjin Milinković  
*University of Belgrade, Serbia*
- Intelligent learning system for High education .....411**  
Aleksandar Kotevski, Gjorgi Mikarovski  
*University "St. Kliment Ohridski" of Bitola, Macedonia*
- Using Cloud Computing in e-learning.....415**  
Gjorgi Mikarovski, Aleksandar Kotevski  
*University "St. Kliment Ohridski" of Bitola, Macedonia*
- An Approach to Define Interfaces for Mobile Telemetry.....418**  
Ivaylo Atanasov, Ventsislav Trifonov, Evelina Pencheva  
*Technical University of Sofia, Bulgaria*
- Architecture of Automated System Software for Testing Petrol Engines .....422**  
Georgi Krastev  
*University of Ruse "Angel Kanchev", Bulgaria*
- Mazes - Classification, Algorithms for Finding an Exit.....425**  
Maya Todorova, Nedyalko Nikolov  
*Technical University of Varna, Bulgaria*
- Integration of Biometrics to the E-Health .....428**  
Milena Stefanova, Oleg Asenov  
*St. Cyril and St. Methodius University of Veliko Tarnovo, Bulgaria*
- Psychology of the Perpetrators of Computer Criminal Acts and Review of Legal and Economic Consequences for the Community .....431**  
Zaklina Spalevic, Jelena Matijasevic, Dejan Rančić\*  
*University Business Academy, Serbia*  
*\*University of Niš, Serbia*
- P2P Wireless Network Based on Open Source Linux Routers .....435**  
Hristofor Ivanov, Miroslav Galabov  
*St. Cyril and St. Methodius University of Veliko Tarnovo, Bulgaria*
- An Approach to Optimization of the Links' Load in the MPLS Domain .....439**  
Veneta Aleksieva  
*Technical University of Varna, Bulgaria*
- Performance Study of Virtualization Platforms for Virtual Networking Laboratory.....443**  
Hristo Valchanov  
*Technical University of Varna, Bulgaria*
- Integration of Video Conference into eLearning Platform Based on Moodle for the Vocational School .....447**  
Ilche Acevski, Valentina Acevska, Linda Fahlberg-Stojanovska  
*University "St. Kliment Ohridski" of Bitola, Macedonia*
- System for Multi-variant Multi-parametric WEB-based Test Control .....451**  
Vladimir Karailiev, Raicho Ilarionov, Hristo Karailiev  
*Technical University of Gabrovo, Bulgaria*

<b>Methods for Assessing Information Sites .....</b>	<b>455</b>
Tihomir Stefanov	
<i>St. Cyril and St. Methodius University of Veliko Tarnovo, Bulgaria</i>	

## INFORMATICS AND COMPUTER SCIENCE (POSTER)

<b>One Approach for Development of Software Modules Adding New Geometric Primitives in 3D Graphics Applications .....</b>	<b>459</b>
Emiliyan Petkov	
<i>St. Cyril and St. Methodius University of Veliko Tarnovo, Bulgaria</i>	
<b>Analysis of Platform Dependencies in Software Solution for Auction and Trading in Electric Energy Market.....</b>	<b>463</b>
Milos Gajic, Marko Djukovic, Sasa Devic, Branislav Atlagic, Zvonko Gorecan, Dragan Tomic	
<i>Telvent DMS LLC, Serbia</i>	
<b>Rapid development of GUI Editor for Power grid CIM models .....</b>	<b>467</b>
Sasa Devic, Lajos Martinovic, Branislav Atlagic, Zvonko Gorecan, Dragan Tomic	
<i>Telvent DMS LLC, Serbia</i>	
<b>Building an 8085 Microprocessor Module for the HADES Simulation Framework .....</b>	<b>471</b>
Goce Dokoski, Dimitar Bojchev, Aristotel Tentov	
<i>University Ss Cyril &amp; Methodius in Skopje, Macedonia</i>	
<b>Algorithms for scheduling of resource-constrained jobs .....</b>	<b>475</b>
Ivaylo Penev, Milena Karova	
<i>Technical University of Varna, Bulgaria</i>	
<b>Hybrid Automatic Repeat Request (HARQ) Overview .....</b>	<b>479</b>
Ginka Marinova, Slava Yordanova, Nikolay Kostov	
<i>Technical University of Varna, Bulgaria</i>	
<b>Digital information transfer systems an overview .....</b>	<b>482</b>
Ginka Marinova, Slava Yordanova, Nikolay Kostov, Boris Nikolov	
<i>Technical University of Varna, Bulgaria</i>	
<b>Towards applicability of agile software development methodologies .....</b>	<b>485</b>
Aleksandar Dimov, Stavros Stavru, Dessislava Petrova-Antonova, Iva Krasteva	
<i>University of Sofia "St. Kl. Ohridski", Bulgaria</i>	

## ELECTRONICS (POSTER)

<b>FPGA (Field Programmable Gate Arrays) – Based Systems-On-a-Programmable-Chip (SOPC) Development for Educational Purposes.....</b>	<b>489</b>
Valentina Rankovska	
<i>Technical University of Gabrovo, Bulgaria</i>	
<b>Electronic Simulator of Sound (Noise) Effects for Electric Vehicles in Urban Areas .....</b>	<b>493</b>
Georgi Pavlov, Galina Chemeva, Radoslav Katsov, Ivaylo Nenov, Ilko Tarpov	
<i>Todor Kableshkov University of Transport, Bulgaria</i>	
<b>Spray Deposition of PVDF Layers with Application in MEMS Pressure Sensors .....</b>	<b>495</b>
Georgi Kolev, Mariya Aleksandrova, Krassimir Denishev	
<i>Technical University of Sofia, Bulgaria</i>	
<b>Different Technological Methods for Offset Compensation in Si Hall Effect Sensors.....</b>	<b>499</b>
Ivelina Cholakova	
<i>Technical University of Sofia, Bulgaria</i>	

<b>Multipoint Video Control System Applicable in Assistance of Elderly and People with Disabilities.....</b>	<b>502</b>
Ivo Iliev, Serafim Tabakov, Velislava Spasova <i>Technical University of Sofia, Bulgaria</i>	
<b>Computer Modeling of RF MEMS Inductors Using SPICE.....</b>	<b>505</b>
Elissaveta Gadjeva <i>Technical University of Sofia, Bulgaria</i>	
<b>Deposition of Transparent Electrodes for the Future Generation of Flexible Displays .....</b>	<b>509</b>
Mariya Aleksandrova, Georgy Dobrikov, Kostadinka Gesheva*, Georgy Bodurov*, Ivelina Cholakova, Georgy Kolev <i>Technical University of Sofia, Bulgaria</i> <i>*Bulgarian Academy of Science, Bulgaria</i>	
<b>Investigation of Over Voltage Protection Circuit for Low Power Applications .....</b>	<b>513</b>
Tihomir Brusev, Nikola Serafimov, Boyanka Nikolova <i>Technical University of Sofia, Bulgaria</i>	
<b>Realization of Low-frequency Amplitude Modulator and Demodulator with FPAAs .....</b>	<b>517</b>
Ivailo Pandiev <i>Technical University of Sofia, Bulgaria</i>	
<b>Modification method to determining the output parameters in the audio power stage with complex load.....</b>	<b>521</b>
Plamen Angelov <i>Burgas Free University, Bulgaria</i>	
<b>Modified method for design of the low-frequency audio driver .....</b>	<b>525</b>
Anton Petrov, Plamen Angelov <i>Burgas Free University, Bulgaria</i>	
<b>SPICE Modelling of Magneto-resistive Sensors.....</b>	<b>529</b>
Boyanka Nikolova, Georgi Nikolov, Milen Todorov <i>Technical University of Sofia, Bulgaria</i>	
<b>Maximizing Power Transfer to the Remote Terminal of PCM4 System .....</b>	<b>533</b>
Zoran Zivanovic, Vladimir Smiljakovic <i>IMTEL Komunikacije a.d., Serbia</i>	
<b>Design, Analysis and Modifications of a Telecom Converter.....</b>	<b>537</b>
Zoran Zivanovic, Vladimir Smiljakovic <i>IMTEL Komunikacije a.d., Serbia</i>	
<b>Image processing of infrared thermograms for hidden objects .....</b>	<b>541</b>
Anna Andonova <i>Technical University of Sofia, Bulgaria</i>	

## ENERGY SYSTEMS AND EFFICIENCY (POSTER)

<b>Study and analysis of optimization approaches for insulation of an industrial grade furnace with electrical resistance heaters .....</b>	<b>544</b>
Borislav Dimitrov, Hristofor Tahrilov, Georgi Nikolov <i>Technical University of Varna, Bulgaria</i>	
<b>Improving energy efficiency of industrial grade furnaces with electrical resistance heaters and comparative model-experiment analysis.....</b>	<b>548</b>
Borislav Dimitrov, Hristofor Tahrilov, Angel Marinov <i>Technical University of Varna, Bulgaria</i>	

<b>Dynamic Braking in Induction Motor Adjustable Speed Drives .....</b>	<b>552</b>
Nebojša Mitrović, Milutin Petronijević, Vojkan Kostić, Bojan Banković <i>University of Niš, Serbia</i>	
<b>Application of Active Front End Rectifier in Electrical Drives.....</b>	<b>556</b>
Bojan Banković, Nebojša Mitrović, Vojkan Kostić, Milutin Petronijević <i>University of Niš, Serbia</i>	
<b>Cyclic Current Rating of Single-Core XLPE Cables with Respect to Designed Life Time .....</b>	<b>560</b>
Miodrag Stojanović, Dragan Tasić, Aleksa Ristić <i>University of Niš, Serbia</i>	
<b>The Influence of the Geometry of the Inductor on the Depth and Distribution of the Inductively Hardened Layer .....</b>	<b>564</b>
Maik Streblau, Bohos Aprahamian, Vladimir Shtarbakov*, Hristofor Tahrilov <i>Technical University of Varna, Bulgaria</i> <i>*METAL PLC, Bulgaria</i>	
<b>Examination of Frequency Controlled Asynchronous Drives at Variable Load Torque - Laboratory Simulator .....</b>	<b>567</b>
Vasil Dimitrov <i>Todor Kableshkov University of Transport, Bulgaria</i>	

## CONTROL SYSTEMS (POSTER)

<b>Multi Leveled Hierarchical Approach for Monitoring and Management Information Systems Construction .....</b>	<b>571</b>
Emiliya Dimitrova, Galina Chemeva <i>Todor Kableshkov University of Transport, Bulgaria</i>	
<b>Improving Control System in the Sulfuric Acid Plant .....</b>	<b>573</b>
Viša Tasić, Dragan Milivojević, Vladimir Despotović*, Darko Brodić*, Marijana Pavlov <i>Institute for Mining and Metallurgy, Serbia</i> <i>*University of Belgrade - Bor, Serbia</i>	
<b>DLadder – an Integrated Environment for Programming PIC Microcontrollers .....</b>	<b>577</b>
Viša Tasić, Dragan Milivojević, Vladimir Despotović*, Darko Brodić*, Marijana Pavlov, Vladan Miljković <i>Institute for Mining and Metallurgy, Serbia</i> <i>*University of Belgrade - Bor, Serbia</i>	

## MEASUREMENT SCIENCE AND TECHNOLOGY (POSTER)

<b>Frequency Measurement Using Compact DAQ Chassis .....</b>	<b>581</b>
Georgi Nikolov, Boyanka Nikolova <i>Technical University of Sofia, Bulgaria</i>	

## REMOTE ECOLOGICAL MONITORING (POSTER)

<b>Wind Energy and Steps Towards 100 Percent of Renewable Energy Penetration .....</b>	<b>585</b>
Aleksandar Malecic <i>University of Niš, Serbia</i>	

## ENGINEERING EDUCATION (POSTER)

- Multitool Online Assisted Design of Communication Circuits and Systems .....589**  
Galia Marinova  
*Technical University of Sofia, Bulgaria*
- E-learning Systems as a Behavioural Analyst .....593**  
Valentin Videkov, Rossen Radonov  
*Technical University of Sofia, Bulgaria*
- Application of Remote Instrumentation in Learning using LabView .....595**  
Ivo Dochev, Liljana Docheva  
*Technical University of Sofia, Bulgaria*
- Simulation of third-order dispersion in single optical channel.....599**  
Kalin Dimitrov, Tsvetan Mitsev, Lidia Jordanova  
*Technical University of Sofia, Bulgaria*
- Developing of a Video Information System for the Technical University of Sofia .....603**  
Kalin Dimitrov, Rumen Mironov, Alexander Bekjarski  
*Technical University of Sofia, Bulgaria*

# Multiuser IR-UWB System Performance

Razvan Craciunescu<sup>1</sup>, Simona Halunga<sup>2</sup>, Octavian Fratu<sup>3</sup>

*Abstract – This paper analyzes a multiuser system performance under the Standard Gaussian Approximation - SGA. In particular it focuses on the analysis of the multiuser interference contribution. Simulations have been performed for a pulse position modulation and a pulse amplitude modulation for 5, 20 and 50 interfering users. The performances have been evaluated in terms of the error probability over the signal to noise ratio, assuming that the signal is transmitted through a AWGN channel.*

*Keywords – Ultra Wideband communication, Standard Gaussian Approximation, Probability of Error.*

## I. INTRODUCTION

Ultra Wideband Technology (UWB) has been described as one of the most promising technologies during the last decade [1-3]. It offers the possibility of achieving higher rates for indoor systems with a reduced range of action due to the resistance encountered in multi-path environments. Moreover, this technology offers lower implementation cost and reduced power requirements than most of other technologies mentioned in the literature [1-2]. Taking into consideration the FCC regulations, a UWB is defined as being any signal in which the 3 dB bandwidth is at least 25% of the central frequency or any signal with a bandwidth larger than 500 MHz [1].

The UWB radio channels can use frequencies from 3.1 GHz to 10.6 GHz, using a frequency bandwidth larger than 7GHz, with the restrictions imposed by the spectral frequency-power masks given in standards [1]. Every radio channel may occupy a bandwidth of at least 500 MHz, in accordance to its central frequency. Regarding the multiple access techniques used, the original proposal for UWB was to use Time Hopping – TH combined with the Pulse Position Modulation – PPM. Later on, various modulation techniques have been used: PAM (pulse amplitude modulation), OOK (On-off keying) or other multiple access techniques such as DS (direct sequence) [2].

In this paper we will present the performances of two types of pulse modulation, in position and in amplitude (TH-PPM and TH-PAM). The communication channel will be modeled as affected by Additive White Gaussian Noise

<sup>2</sup>Razvan Craciunescu is with the Faculty of Electronics, Telecommunications and Information Technology at University POLITEHNICA of Bucharest, 313 Splaiul Independentei, 060042 Bucharest-6, Romania, E-mail: rcraciunescu@radio.pub.ro .

<sup>3</sup>Simona Halunga is with the Faculty of Electronics, Telecommunications and Information Technology at University POLITEHNICA of Bucharest, 313 Splaiul Independentei, 060042 Bucharest-6, Romania, E-mail: shalunga@elcom.pub.ro .

<sup>4</sup>Octavian Fratu is with the Faculty of Electronics, Telecommunications and Information Technology at University POLITEHNICA of Bucharest, 313 Splaiul Independentei, 060042 Bucharest-6, Romania, E-mail: ofratu@elcom.pub.ro .

(AWGN), while the UWB signal will be an Impulse Radio Ultra Wideband signal (IR-UWB). The IR-UWB transmitting method is the most common way of transmitting UWB signals. This type of signals are radiated pulses very short in time.

Several methods of evaluating the effects of the multi-user interference have been developed over the years [3,4]. This paper focuses on the Standard Gaussian Approximation (SGA) hypothesis, which models the effects of all the interferences as a Gaussian additive noise, with a uniform power spectrum distribution over all frequencies of interest. The hypothesis is very accurate for a large number of interfering users, and has very optimistic results for low values of bit rates or for a small number of interfering users. [

## II. THE PERFORMANCES OF A MULTI-USER IR-UWB SYSTEM IN SGA HYPOTHESIS

In this section we will present the general simulation scenario and the analytical expressions of the probability of error for the two types of modulation which will undergo our analysis, PPM and PAM.

In the following work, the following assumptions are assumed to be valid:

1. All sources produce binary vectors  $\mathbf{b}$ ,  $b_k \in \{0,1\}$ ,  $\forall k$ .
2. All sources use the same pulse code period frequency,  $T_s$ .
3. The spreading codes  $c_k \in \{\pm 1\}$ ,  $(\forall)k$  are independent and equally likely, with the same code period  $T_c$ .
4. For each transmission/reception path, a different code – known at the receiver – is used.
5. It is assumed that the base impulse has a limited duration,  $T_m$  and a symmetrical frequency shape.
6. Propagation is achieved on a channel with multi-paths. For a given user,  $n$ , the channel impulse response is a function of the path gain,  $\alpha(n)$ , and of the path time delay  $\tau(n)$ . Delays are considered to be independent and uniformly distributed within the  $[0, T_s)$  interval. The channel impulse response is thus given by:

$$h^{(n)}(t) = \alpha^{(n)} \delta(t - \tau^{(n)}) \quad (1)$$

7. The channel is affected only by AWGN, with the spectral power density of  $N_0/2$ (W/Hz)
8. Coherent single user correlation reception is implemented at the receiver for all users. The signal will be thus correlated with the user code and integrated over a bit period  $T_b = N_s T_c$ .

### A. TH-PPM

The binary signal, TH-PPM, transmitted by user  $n$  can be written as [3,4]:

$$s_{TX}^{(n)}(t) = \sum_{j=-\infty}^{\infty} \sqrt{E_{TX}^{(n)}} p_0(t - jT_s - c_j^{(n)}T_c - a_j^{(n)}\varepsilon) \quad (2)$$

where  $p_0(t)$  is the normalized base impulse,  $E_{TX}^{(n)}$  is the energy transmitted by each pulse,  $c_j^{(n)}T_c$  is the time shift imposed by the TH code,  $c_j^{(n)}$  is the  $j$ -th TH code sequence used by user  $n$ , and  $T_c$  is the chip duration. Each TH code is a sequence of  $N_p$  identically and independently distributed random variables, each of them with a probability of  $1/N_h$ , and with values within  $[0, N_h-1]$  interval, where  $N_h$  is the cardinality of the TH code. In order to identify the users, each of them will be assigned a specific TH code in such a way to avoid collision at the receiver. The term  $a_j^{(n)}\varepsilon$  the time-shift introduced by the data;  $\varepsilon$  is the specific PPM time delay and  $a_j$  is the binary value assigned to the  $j$ th pulse for user  $n$ . The binary vector  $\mathbf{a}$  represents the output of a  $(N_s, 1)$  repetition coder, that receives as input the binary vector  $\mathbf{b}$ , meaning that  $N_s$  pulses carry the information of one bit. The binary vector  $\mathbf{a}$  length is the length of  $\mathbf{b}$  time  $N_s$ , [1,3,4,5]

Assuming that the channel is modeled by Eq. (1), in the presence of AWGA noise the received signal can be written as

$$r(t) = \sum_{n=1}^N \sum_{j=-\infty}^{\infty} \sqrt{E_{RX}^{(n)}} p_0(t - jT_s - c_j^{(n)}T_c - a_j^{(n)}\varepsilon - \tau) + n(t) \quad (3)$$

where  $N_u$  is the number of users and  $E_{RX}$  is the energy of each transmitted pulse at the receiver. [1,3,4,5].

Referring to user (1) and assuming that the receiver is perfectly synchronized, such that the time delay is accurately known at the receiver and can be assumed 0, for simplicity, the received signal can be written as [4]:

$$r(t) = r_u(t) + r_{mui}(t) + n(t) \quad (4)$$

Next, focusing our analysis on the bit interval  $T_b$  and taking into account the symmetry of the system, the analysis can be performed within  $[0, T_b]$  interval. The  $r_u(t)$  and  $r_{mui}(t)$  contribution can be written for  $t \in [0, T_b]$  as [4]:

$$r(t) = \sum_{j=0}^{N_S-1} \sqrt{E_{RX}^{(1)}} p_0(t - jT_s - c_j^{(1)}T_c - a_j^{(1)}\varepsilon) \quad (5)$$

$$r_{mui}(t) = \sum_{n=1}^{N_U} \sum_{j=-\infty}^{\infty} \sqrt{E_{RX}^{(n)}} p_0(t - jT_s - c_j^{(n)}T_c - a_j^{(n)}\varepsilon - \tau^{(n)}) \quad (6)$$

In the decision process performed at the reception, the correlation output at reception is thus given by [2,4,5]:

$$Z = \int_0^{T_b} r(t)m(t)dt \quad (7)$$

,where  $m(t)$  is the correlation mask upon reception defined by:

$$m(t) = \sum_{j=0}^{N_S-1} v(t - jT_s - c_j^{(1)}T_s) \quad (8)$$

$$v(t) = p_0(t) - p_0(t - \varepsilon) \quad (9)$$

Combining Eqs. (7) and (9) we obtain [4]:

$$Z = Z_u + Z_{mui} + Z_n \quad (10)$$

Under the SGA hypothesis,  $Z_{mui}$  and  $Z_n$  represents random Gaussian processes with a 0 mean and a variance of  $\sigma_{mui}^2$  and  $\sigma_n^2$  respectively. The average bit error rate  $Pr_b$  can be written as [3,4,5]:

$$\begin{aligned} Pr_b &= \frac{1}{2} \operatorname{erfc} \left( \sqrt{\frac{E_b}{2(\sigma_n^2 + \sigma_{mui}^2)}} \right) \\ &= \frac{1}{2} \operatorname{erfc} \left( \sqrt{\frac{((SNR_n)^{-1} + (SIR)^{-1})^{-1}}{2}} \right) \end{aligned} \quad (11)$$

The bit energy of the received signal,  $E_b$ , can be obtained by calculating the energy of the useful components at the output of the receiver for all  $N_s$  pulses that form a bit. Therefore

$$E_b = E_{RX}^{(1)} N_s^2 (1 - R_0(\varepsilon))^2 \quad (12)$$

,where  $R_0(t)$  is the autocorrelation function of the base impulse  $p_0(t)$  pulse. In the presence of the thermal noise, the signal to noise ratio (SNR) can be written as:

$$SNR_n = \frac{N_s E_{RX}^{(1)}}{N_0} (1 - R_0(\varepsilon)) = \frac{E_b^{(1)}}{N_0} (1 - R_0(\varepsilon)) \quad (13)$$

Regarding the signal to interferences ratio (SNI) it can be written as

$$SIR = \frac{E_{RX}^{(1)} N_s^2 (1 - R_0(\varepsilon))^2}{\frac{1}{T_s} N_s \sigma_M^2 \sum_{n=2}^N E_{RX}^{(n)}} = \frac{(1 - R_0(\varepsilon))^2 \gamma_R}{\sigma_M^2} \frac{1}{R_b \sum_{n=2}^N \frac{E_{RX}^{(n)}}{E_{RX}^{(1)}}} \quad (14)$$

By combining Eqs. (13) and (14) and replacing them in Eq. (11) we obtain the average bit error rate [3,4,5]:

$$Pr_b = \frac{1}{2} \operatorname{erfc} \left( \sqrt{\frac{1}{2} \left( \frac{E_b^{(1)}}{N_0} (1 - R_0(\varepsilon)) \right)^{-1} + \frac{(1 - R_0(\varepsilon))^2 \gamma_R}{\sigma_M^2 R_b \sum_{n=2}^N \frac{E_{RX}^{(n)}}{E_{RX}^{(1)}}}} \right)^{-1} \quad (15)$$



C. TH-PAM

The TH-PAM signal can be analyzed by following a similar procedure as the one presented above. The binary signal transmitted by user  $n$  can be written as [6]:

$$s_{TX}^{(n)}(t) = \sum_{j=-\infty}^{\infty} \sqrt{E_{TX}^{(n)}} a_j^{(n)} p_0(t - jT_s - c_j^{(n)}T_c) \quad (16)$$

The signal at the output of the correlator at reception has the same expression as in Eq. (7) with a correlation mask  $m(t)$  defined as [1,2,6]:

$$m(t) = \sum_{j=0}^{N_s - 1} p_0(t - jT_s - c_j^{(1)}T_c) \quad (17)$$

With the same decision criterion as in the above paragraph, the error probability,  $Pr_b$  obtained in Eq. (11) is still valid.

The expressions for the signal to noise ratio (SNR) in the presence of the thermal noise, as well as the signal to interference ratio (SIR) can be written as:

$$SNR_n = \frac{N_s E_{RX}^{(1)}}{\frac{N_0}{2}} = \frac{2E_b^{(1)}}{N_0} \quad (18)$$

$$SIR = \frac{E_{RX}^{(1)} N_s T_s}{\sigma_M^2 \sum_{n=2}^N E_{RX}^{(n)}} = \frac{\gamma_R}{R_b \sum_{n=2}^N \frac{E_{RX}^{(n)}}{E_{RX}^{(1)}}} \quad (19)$$

Therefore, the probability of error,  $Pr_b$ , is given by [6]

$$Pr_b = \frac{1}{2} \operatorname{erfc} \left( \sqrt{\frac{1}{2} \left[ \left( \frac{2E_b^{(1)}}{N_0} \right)^{-1} + \frac{\gamma_R}{R_b \sum_{n=2}^N \frac{E_{RX}^{(n)}}{E_{RX}^{(1)}}} \right]^{-1}} \right) \quad (20)$$

III. NUMERICAL RESULTS AND SIMULATIONS

As far as the numerical results are concerned we will analyse the performances of a IR-UWB system in the presence of the multi-user interferences (MUI). First, we will evaluate the probability of error,  $Pr_b$ , in the case of using the TH multiple access technique both for a PPM binary modulation and for a PAM one. In both circumstances  $Pr_b$  will be estimated in accordance with the theoretical results obtained in Sections II.A. and II.B.

We compared the performances for PPM and PAM in three different scenarios, using 5, 20 and 50 interference signals. The results are presented in Figs 1, 2 and 3. The transmitted signal has a rate of 20 Mbit/s. The used pulse is given and has the shape of the second Gaussian derivative with a shaping factor of 0.25ns. In the case of PPM the timeshift is  $\epsilon=0.5$  ns.

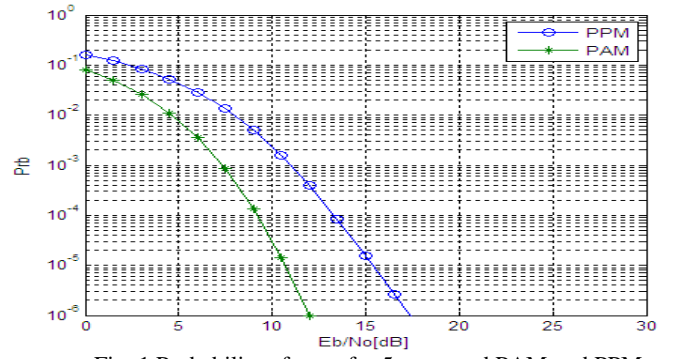


Fig. 1 Probability of error for 5 users and PAM and PPM

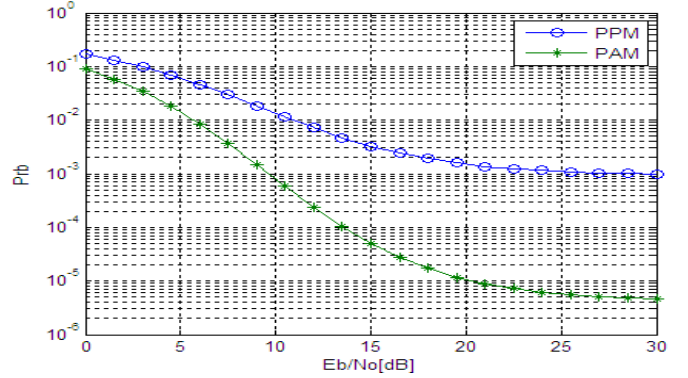


Fig. 2 Probability of error for 20 users and PAM and PPM

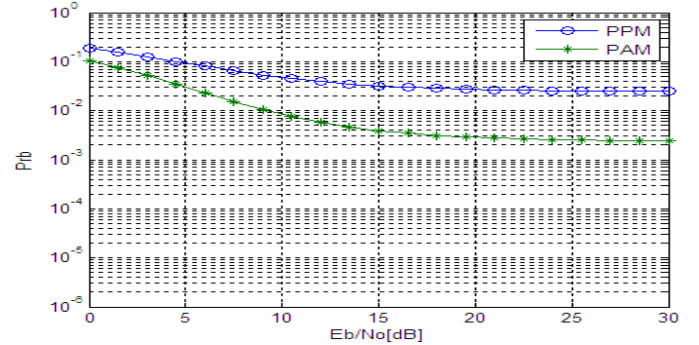


Fig. 3 Probability of error for 50 users and PAM and PPM

From these figures we can extract the following conclusions. First, in Fig. 1 we can notice that the multi-user interference term can be neglected, because the probability of error decreases as the SNR increases. We can say that, if the number of users is small enough, only the thermal noise affects the probability of error. The  $E_b/N_0$ [dB] distance between PAM and PPM is approximately 3dB, for a  $Pr_b=10^{-2}$ , and it increases for the lower values of  $Pr_b$ . From Figs. 2 and 3 we can observe that the probability of error  $Pr_b$  tends to a constant value, as the signal to noise ratio increases, showing the fact that, for large  $E_b/N_0$ , the system performances are dominated by the multi-user interference. We can, moreover, identify two regions: for low values of  $E_b/N_0$ ,  $Pr_b$  is determined mostly by the thermal noise, in which case we can improve the performances of the system by increasing the transmission power; for high  $E_b/N_0$  ratios, the systems performances trends asymptotically to a constant value and does no longer depend on  $E_b/N_0$ . In this case the performances are dominated by the multi-user interference effects.

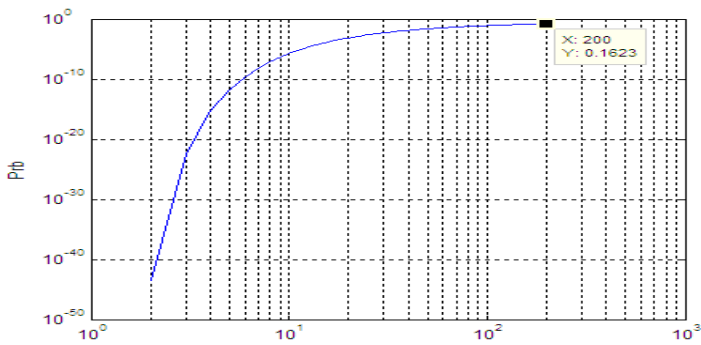


Fig. 3 Probability of error for 1: 200 users

As the number of users increases, the  $Pr_b$  is asymptotically limited to a higher value. If the number of users increases up to 50, the probability of error decrease only to  $2 \cdot 10^{-3}$ , in the PAM case, and to  $2.5 \cdot 10^{-2}$  in the PPM case. All graphs shows that PAM is slightly more robust then PPM with respect to  $Pr_b$  performances.

As it can be seen in Figs. 2 and 3,  $Pr_b$  increases with the augmentation of the number of users. In order to evaluate the performances of the system with the number of users increases. we represented graphically  $Pr_b$  as a function of number of users. Thus, in Fig. 4 we can see that in the presence of more than 200 users, the error probability is very high, with a magnitude order of  $10^{-1}$ ,  $10^{-2}$ , and the system is dominated by M.U.I.

In order to validate the theoretical results, we simulated the UWB receiver in the presence of multi-user interference. We have been particularly interested in how precise the SGA hypothesis used for error probability estimation is fulfilled. We simulated a system with 7 and 10 users. Each user generates stream of data with a bit period  $T_b=18ns$ , leading to a bit rate of 55.55Mbit/s. Every bit period is organized in 3 frames with a duration  $T_s=6ns$ , meaning 3 pulses are transmitted for every bit. Each frame is then divided into 6 slots with a length of  $T_c=1ns$ . All users transmit with the same format of the signal. The results are shown in Fig. 5.

We can notice that the theoretical model used for the evaluation  $Pr_b$  underestimates the effect of MUI. The error probability obtained based on the simulation is larger than the one the SGA hypothesis predicted. In the 7 user case the difference between the error probability from the theoretical model and the simulated one is larger than in the 10 user case:

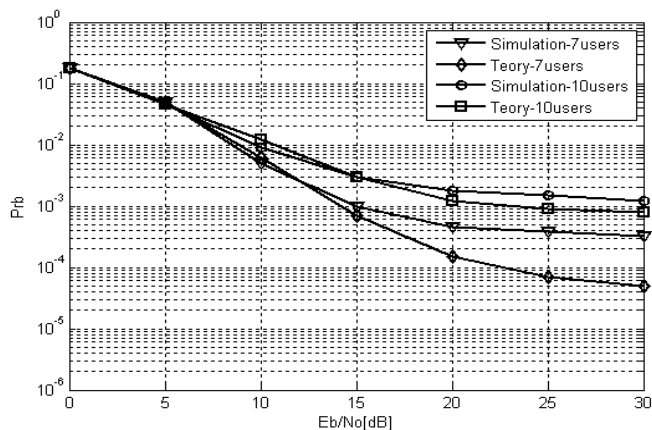


Fig. 4 Comparasion between theoretical ans simulation results for 7 and 10 users

#### IV. CONCLUSION

In this paper we have evaluated the performances of an IR-UWB system in the presence of multi-user interferences and in the Standard Gaussian Approximation hypothesis. We have noticed that the multi-user interferences influence on the performances of such a system is important only if the number of users is large enough. We have also noted that the PAM modulation is far more robust than the PPM one, achieving error probabilities ten times smaller at the same  $E_b/N_0$  ratio. As the number of users increases, the  $Pr_b$  is asymptotically limited to a higher value.

Next, investigating the validity of the SGA hypothesis in comparison with simulation results, we observed that, for a relatively low number of users (up to 10) the results obtained based on simulation are better with respect to the error probability then the ones obtained under the SGA assumptions, showing that the formulas developed represents an upper limit for the  $Pr_b$ .

In a future work we will focus on increasing the number of users in the simulation, developing simulations under different sets of parameters and checking other hypotheses, like the multi-user interference model based on package collision or the chip-synchronous hypothesis.

#### ACKNOWLEDGEMENT

This research activity was supported by Ministry of Communications and Information Society of Romania under the grant no.106/2011 "Evolution, implementation and transition methods of DVB radiobroadcasting using efficiently the radio frequencies spectrum".

#### REFERENCES

- [1] M. Ghavami, L. B. Michael, R. Kohno, "Ultra Wideband Signals and Systems in Communication Engineering", John Wiley & Sons, 2007.
- [2] Kharrat-Kammoun, F. Le Martret, C.J. Ciblat, "Performance analysis of IR-UWB in a multi-user environment", Wireless Communications, IEEE Transactions on, Issue 11, Noiembre 2009.
- [3] G. Durisi, S. Benedetto, "Performance evaluation and comparison of different modulation schemes for UWB multiaccess systems", Communications, 2003. ICC '03. IEEE International Conference on
- [4] Shing TenqChen, Ying-Haw Shu, Ming-Chang Sun, Wu-Shiung Feng and Chao-Hao Lee, "Performance comparison of PPM-TH, PAM-TH, and PAM-DS UWB Rake receivers with channel estimators via correlation mask", ICCOM'05 Proceedings of the 9th WSEAS International Conference on Communications
- [5] M. Z. Win and R. A. Scholtz. "Ultra-wide bandwidth time-hopping spread spectrum impulse radio for wireless-access communication". IEEE Transactions on Communications, vol.48, no. 4 pp.:679-691, 2000.
- [6] Zhiquan Bai, Kyungsup Kwak, "Performance analysis of TH-PAM of UWB system and the coded scheme" Wireless Communications, Networking and Mobile Computing, 2005. Proceedings. 2005 International Conference on

# Doppler Fading Effects on OFDM Transmissions

Ioana Bucsa<sup>1</sup>, Razvan Craciunescu<sup>2</sup>, Simona Halunga<sup>3</sup>, Octavian Fratu<sup>4</sup>

**Abstract** – This paper presents the performances of different Doppler fading models for a OFDM communication. The simulations were carried out over a AWGN channel with a BPSK modulation, and the performances have been evaluated in terms of probability of error for each model.

**Keywords** – Fading, Doppler Effect, OFDM, MATLAB.

## I. INTRODUCTION

Radio propagation for mobile communications, especially in dense urban areas or in indoor environment, is characterized by severe multipath phenomenon that causes fading and distortion effects, leading to inter-symbol interference (ISI) [1].

The fading is one of the major factors affecting the performance of such systems [1,2]. The large-scale fading is given by path loss and shadowing effects of buildings or prominent terrain contours. The small-scale fading is the common reference to the rapid changes in signal amplitude and phase. In this work, the combined effects of large- and small-scale fading are considered.

Multi-carrier communication is a way to increase bandwidth without amplifying the noise in the signal on account of the frequency selectivity fading that affects the channel. This type of communication implies sending several narrower band signals (called subcarriers) instead of a single broadband one. The signals are multiplexed in frequency and they are transmitted together to the same receiver, on the same radio link. By sending  $M$  signals in parallel on the same radio link we increase the rate of transfer by  $M$  times. In the same time, the impact of frequency selectivity fading depends on the bandwidth of each subcarrier [1].

An extended band involves higher transfer rates, especially for the descending connection. As the symbol rate for each subcarrier is much smaller than the initial symbol rate, the effects of delayed scattering, for example ISI (Inter Symbol

Interferences), considerably decrease and reduce the complexity of the equalizer at the receiver [1,2].

## II. TYPES OF DOPPLER FADING THAT AFFECT THE RADIO CHANNEL

### A. Multipath Channel

The performances of any wireless communication system is strongly affected by the multipath phenomenon, very common in dense urban areas, when the transmission path between the transmitter and the receiver is severely obstructed by buildings, trees or other objects.

The variation of the signal in the communication systems can be caused by the short and/or long term fading. Short term fading includes multi-path fading (Rayleigh, Rice fading) and the Doppler fading [2].

The multi-path fading determines whether a channel is flat or selective in frequency, while the Doppler fading divides channels into slow fading channels and fast fading ones. While the multi-path fading causes a scattering of the pulse in time, the Doppler fading causes a scattering in frequency [2,3].

### B. The Doppler Fading

The multi-path fading (Rayleigh, Rice) does not take into consideration the possible movements of the emitter and/or of the receiver. If the emitter/receiver is mounted on a moving vehicle the Doppler fading occurs. If a signal is emitted with the  $f_0$  frequency, then the spectrum of the received signal will broaden and it will contain spectral components from  $f_0 - f_d$  to  $f_0 + f_d$ , where  $f_d$  is the Doppler deviation given by [4]:

$$f_d = \frac{vf_0 \cos(\theta)}{c} \quad (1)$$

where  $v$  is the velocity of the emitter/receiver, and  $\theta$  is the angle between the emitter/receiver forward velocity and the line of sight from the emitter to the receiver.

If the band which is occupied by the useful signal is wider than the Doppler bandwidth, the Doppler scattering will cause no problem either for emission or reception. In this case we are dealing with a channel affected by a slow fading. On the other hand, if the band is smaller than the Doppler bandwidth, the movement produces a fast variation of the channel during the length of the pulse. Thus the fading is considered to be fast.

In the following paragraphs we will analyze in detail some Doppler power spectrum models from the point of view of their applicability and that of the theoretical expressions of the power spectral density (PSD) of fading processes [2,4,5].

The **Jakes Doppler** power spectrum model is applied to a mobile receiver. It is called the classic model and it is built based on the following hypothesis: propagation of radio

<sup>1</sup>Ioana Bucsa is with the Faculty of Electronics, Telecommunications and Information Technology at University POLITEHNICA of Bucharest, 313 Splaiul Independentei, 060042 Bucharest-6, Romania.

<sup>2</sup>Razvan Craciunescu is with the Faculty of Electronics, Telecommunications and Information Technology at University POLITEHNICA of Bucharest, 313 Splaiul Independentei, 060042 Bucharest-6, Romania, E-mail: rcraciunescu@radio.pub.ro .

<sup>3</sup>Simona Halunga is with the Faculty of Electronics, Telecommunications and Information Technology at University POLITEHNICA of Bucharest, 313 Splaiul Independentei, 060042 Bucharest-6, Romania, E-mail: shalunga@elcom.pub.ro .

<sup>4</sup>Octavian Fratu is with the Faculty of Electronics, Telecommunications and Information Technology at University POLITEHNICA of Bucharest, 313 Splaiul Independentei, 060042 Bucharest-6, Romania, E-mail: ofratu@elcom.pub.ro .

waves parallel to the ground; at receiver, the angle of arrival is uniformly spread in  $[-\pi, \pi]$ ; the receiver antenna is omnidirectional.[3,4].

The baseband normed Jake spectrum is:

$$S_j(f) = \frac{1}{\pi f_d \sqrt{1 - (f/f_d)^2}}, |f| \leq f_d \quad (2)$$

where  $f_d$  is the maximum Doppler deviation.

It has been proven that in a 3-D isotropic scattering environment, in which the arrival angles are uniformly distributed in the elevation plan and in that of the azimuth, the total PSD for the  $\theta$  angle of elevation and the  $\alpha$  azimuth angle is:  $p_{\theta, \alpha}(\theta, \alpha) = \frac{\sin \theta}{4\pi}$ ,  $0 \leq \theta < \pi$ ,  $0 \leq \alpha < 2\pi$ , the theoretical spectrum is flat.[2,3,4].

The baseband normed **flat Doppler** spectrum is:

$$S_p(f) = \frac{1}{2\pi f_d}, |f| \leq f_d \quad (3)$$

The next model corresponds to the multi-path components with high rates of delay in the UHF communications. It has also been put forward in the case of the high frequency (HF) channels as well as for the aeronautical channels with a VHF band [3,4].

The baseband normed **Jake Gaussian** spectrum is:

$$S_g(f) = \frac{1}{\sqrt{2\pi\sigma_g^2}} e^{-\frac{f^2}{2\sigma_g^2}} \quad (4)$$

where  $\sigma_g$  is the standard deviation.

The **Jakes bi-Gaussian** model is built from two Gaussian spectrums which are shifted in frequency. This is used for modeling the long echoes which can occur in urban areas and hilled terrains [2,3,4].

The baseband normed Jakes bi-Gaussian spectrum is :

$$S_{bg}(f) = A_{bg} \left[ \frac{C_{g1}}{\sqrt{2\pi\sigma_{g1}^2}} e^{-\frac{(f-f_{g1})^2}{2\sigma_{g1}^2}} + \frac{C_{g2}}{\sqrt{2\pi\sigma_{g2}^2}} e^{-\frac{(f-f_{g2})^2}{2\sigma_{g2}^2}} \right] \quad (5)$$

where  $\sigma_{g1}$  and  $\sigma_{g2}$  are the standard deviations,  $f_{g1}$  and  $f_{g2}$  are the central frequencies,  $C_{g1}$  and  $C_{g2}$  are the power gains, and  $A_{bg} = \frac{1}{C_{g1} + C_{g2}}$  is the norming coefficient. If  $C_{g1} = 0$  or  $C_{g2} = 0$ , we can obtain a Gaussian Doppler spectrum which is shifted in frequency. In case both central frequencies are 0 and the standard deviations are equal, the result is a Gaussian Doppler spectrum.

As we have mentioned beforehand, the Jakes Doppler spectrum is build based on the fact that the angle of arrival at the mobile receiver is uniformly distributed, and the spectrum covers the  $[-f_d, f_d]$  frequencies. If the angles are not uniformly distributed, the spectrum does not cover this interval, a fact which occurs in the case of a directional antenna. This type of spectrum is called restricted. The spectrum will also be considered symmetrical in order to obtain a real impulse response [4].

The baseband normed **restricted Jakes Doppler(rjakes)** spectrum is:

$$S_{rj}(f) = \frac{A_{rj}}{\pi f_d \sqrt{1 - (f/f_d)^2}} \quad (6)$$

where  $0 \leq f_{d,min} \leq |f| \leq f_{d,max} \leq f_d$  and the norming factor defined as:

$$A_{rj} = \frac{\pi/2}{\sin^{-1}\left(\frac{f_{d,max}}{f_d}\right) - \sin^{-1}\left(\frac{f_{d,min}}{f_d}\right)} \quad (7)$$

where  $f_{d,min}$  și  $f_{d,max}$  are the minimum and maximum positive frequencies for which the spectrum is non-zero. These frequencies can be determined from the PSD of the angle of arrival.

The restricted Jakes Doppler spectrum was considered to be symmetrical. The asymmetric spectrums occur in the case of directional antennae, of aeronautical channels and of the satellite mobile radio channels. Taking an asymmetrical spectrum in consideration, the pulse response will be complex.

The baseband **asymmetrical restricted Jakes Doppler(ajakes)** spectrum is expressed analytically [2,4]:

$$S_{aj}(f) = \frac{A_{aj}}{\pi f_d \sqrt{1 - (f/f_d)^2}} \quad (8)$$

where  $-f_d \leq f_{d,min} \leq f \leq f_{d,max} \leq f_d$  and the norming factor defined as:

$$A_{aj} = \frac{\pi}{\sin^{-1}\left(\frac{f_{d,max}}{f_d}\right) - \sin^{-1}\left(\frac{f_{d,min}}{f_d}\right)} \quad (9)$$

where  $f_{d,min}$  și  $f_{d,max}$  are the minimum and maximum positive frequencies for which the spectrum is non-zero.

The round spectral power density is approximated by the measured PSD of a scattering component, while taking into consideration a wireless channel of 2.5 GHz. In this case, the PSD representation is also influenced by the frequency of the central carrier [1,2,3,4].

The baseband **rounded normed Doppler** spectrum is defined as follows:

$$S_r(f) = C_r \left[ a_0 + a_2 \left(\frac{f}{f_d}\right)^2 + a_4 \left(\frac{f}{f_d}\right)^4 \right] \quad (10)$$

where  $|f| \leq f_d$  and the norming factor defined as:

$$C_r = \frac{1}{2f_d \left[ a_0 + \frac{a_2}{3} + \frac{a_4}{5} \right]} \quad (11)$$

We can notice that the rounded Doppler spectrum is a polynomial in frequency function, of the fourth order, in which only even exponents appear. The real numbers  $-a_0, a_2, a_4$  are the coefficients of the polynom. In the IEEE 802.16 standard the following values are used:  $a_0 = 1$ ,  $a_2 = -1.72$ ,  $a_4 = 0.785$ .

### III. ORTHOGONAL FREQUENCY-DIVISION MULTIPLEXING – BASIC PRINCIPLES

The Orthogonal frequency-division multiplexing (OFDM) transmission is a multi-carrier type of transmission.

A complex OFDM signal,  $x(t)$ , during the  $mT_u < t < (m+1)T_u$  interval, can be written as follows [5,6,7]:

$$x(t) = \sum_{k=0}^{N_c-1} x_k(t) = \sum_{k=0}^{N_c-1} a_k^{(m)} e^{j2\pi k\Delta f t} \quad (12)$$

where  $x_k(t)$  is the  $k$ -th subcarrier modulated with the frequency of  $f_k = k\Delta f$  and  $a_k$  is the modulator symbol for the  $k$ -th subcarrier in the  $m$ -th order OFDM time interval and  $T_u$  is the useful symbol duration [1,5,7].

The OFDM transmission is based on blocks, which means that during the length of every OFDM symbol,  $N_c$  modulating symbols are transmitted in parallel. The symbols can be modulated using one of the following modulations: BPSK, QPSK, 16QAM sau 64QAM.

At any rate, for a channel with time dispersion, the orthogonality between the subcarriers will be partially or definitely lost. Consequently, in the case of a channel with time dispersion there will be both inter-symbol interferences in the same subcarrier and interferences between different subcarriers. In order to solve this problem and to make the OFDM signal impervious to the time dispersion that occurs within the channel, the cyclical prefix is introduced. The insertion of the cyclical prefix consists of copying the last part of the OFDM symbol and introducing that at the start of the symbol. Once this cyclical prefix is inserted, the OFDM signal increases from  $T_u$  to  $T_u + T_{cp}$ , where  $T_{cp}$  is the length of the cyclical prefix, but also there occurs a decrease in the symbol rate. If at the receiver end, the correlation is performed during the  $T_u = 1/\Delta f$  interval, then the orthogonality between the subcarriers will be kept even if there is a channel with a time dispersion, with the condition that the duration of the dispersion is lesser than the length of the cyclical prefix [1,2,6].

Due to its features and to the choice of an adequate spacing,  $\Delta f$ , between subcarriers, the OFDM allows for a less complex implementation from the point of view of the calculus efficiency by using the Fast Fourier Transform (FFT) [1,2].

### IV. PERFORMANCE ANALYSIS

Simulations were carried out over an Additive white Gaussian noise (AWGN) multipath channel, with Ricean/Rayleigh fading and different types of Doppler fading presented in II.B. For the OFDM communication, the number of subcarriers is 52, with a BPSK modulation and a frequency spacing of 312,5 kHz. The bandwidth is 20MHz, the useful symbol duration is 3,2  $\mu$ s and a cycling prefix of 0.8  $\mu$ s. The length of the FFT algorithm used is 64.

We are focusing our interest on the variation of the bit probability of error for the different types of Doppler fading presented in II.B. First of all a AWGN channel with Rayleigh fading is analyzed. The Doppler frequency used in the

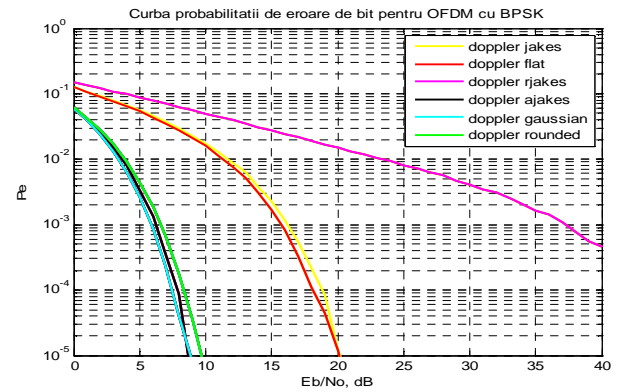


Fig. 1. Probability of error for a AWGN, Rayleigh fading channel, and the Doppler frequency of 10Hz

simulations is 10Hz, 50Hz and 100Hz. Thus, the following variations of the bit probability of error are obtained.

For a Doppler deviation of 10Hz, we can notice in figure 1 that the probability of error ( $P_e$ ) reaches its minimum value of  $10^{-5}$  for  $E_b/N_0 = 8-10$  dB in the case of rounded, Akajes, Gaussian type of Doppler scattering. For a flat type of Doppler and a Jakes one,  $P_e$  is minimum when  $E_b/N_0 = 20$  dB, and then the error increases significantly for a Rjakes type

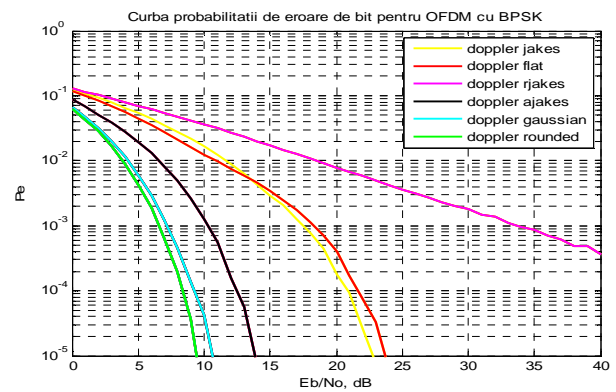


Fig. 2. Probability of error for a AWGN, Rayleigh fading channel, and the Doppler frequency of 50Hz

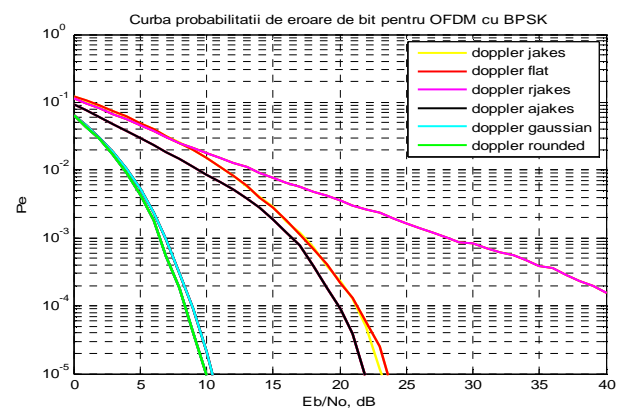


Fig. 3. Probability of error for a AWGN, Rayleigh fading channel, and the Doppler frequency of 100Hz

of Doppler.

In figure 2 the most favorable cases from the point of view of the bit error probabilities are obtained for the rounded and Gaussian type of Doppler, while the most unfavorable situation is that of the channel which is affected by the Rjakes Doppler. For the Ajakes Doppler, the proportion between the bit energy and the power spectral density reaches the value of 13,8 dB, increasing until 22,8 dB and 23,7 dB in the case of the channel which is affected by the Jakes Doppler, flat respectively. Modifying the Doppler deviations to 100Hz, the only major difference that occurs in figure 4.25 is noticeable around the value of the  $E_b/N_0$  proportion that increases by 8 dB, from 13,8 dB to 21,8 dB, in the case of the channel affected by the Ajakes Doppler.

## V. CONCLUSION

In this paper we compared the performances of several Doppler fading models, presented in II.B, for an OFDM communication. Thus, for lower Doppler frequencies we obtained a better probability of error than for higher ones. Furthermore, in the case of Rayleigh fading, for the same probability of error,  $E_b/N_0$  is at minimum in the case of the rounded Doppler model and at a maximum for the Rjakes Doppler model. For Rician fading we observed the minimum for  $E_b/N_0$  is still for the rounded Doppler model but the maximum is, in this case for the Jakes Doppler model.

Therefore, both for Rician and Rayleigh fading the more robust Doppler model is the rounded Doppler model. The worst model, in terms of probability of error is the Rjakes respectively the Jakes Doppler model.

## ACKNOWLEDGEMENT

This research activity was supported by Ministry of Communications and Information Society of Romania under the grant no. 106/2011 "Evolution, implementation and transition methods of DVB radiobroadcasting using efficiently the radio frequencies spectrum".

## REFERENCES

- [1] Andreas F. Molisch, *Wireless Communication Second Editions*, Wiley & Sons.,2011.
- [2] Franz Hlawatsch , Gerald Matz , *Wireless Communications Over Rapidly Time-Varying Channels* , Elsevier 2011.
- [3] R. H. Clarke, "A statistical theory of mobile-radio reception", *Bell Sys.Tech. J.*, vol. 47, no. 6, pp. 957–1000, July-Aug. 1968.
- [4] P. M. Shankar, *Fading and Shadowing in Wireless Systems*, Springer 2011.
- [5] Haiying Zhu, Bouchard, L., Boucher, L. , *Performance of OFDM based wireless LAN system under Doppler over Rayleigh fading*, *Communication Technology Proceedings, International Conference on*. Vol.2,pp 1234 - 1237 , 2003.
- [6] K.A.Hamdi, *Unified Error-Rate Analysis of OFDM over Time-Varying Channels*, *Wireless Communications*, *IEEE Transactions on*, Vol.10, No. 8,pp. 2692 – 2702, 2011.
- [7] L. Noor A. , Anpalagan ;S. Kandeepan. *SNR and BER derivation and analysis of downlink OFDM systems with noisy fading Doppler channels*, *Signal Processing and Its Applications*, 9th International Symposium on, pp 1-4, 2007.

# Investigation of the modulation type's influence on the DVB-T signals quality

Oleg Borisov Panagiev<sup>1</sup>

**Abstract** – In this paper are made and described studies that represent the influence of key critical factors, ranging from RF performance characteristics, such as error vector magnitude (EVM), modulation error ratio (MER), through to roll-off and FEC for M-ary QAM modulation of subcarriers in the OFDM signal.

**Keywords** – DVB-T, EVM, BER, S/N, roll-off.

## I. INTRODUCTION

Advanced digital terrestrial systems behave quite differently when compared to traditional analog TV as the signal is subjected to noise, distortion, and interferences along its path. The secret to maintaining reliable and high-quality services over DVB-T transmission systems is to focus on critical factors that may compromise the terrestrial reception. Once reception is lost, the path to recovery isn't always obvious. The problem could be caused by MPEG table errors, or merely from the RF power dropping below the operational threshold or the cliff point. RF problems can include any of the following: terrestrial RF signal reflections, poor noise performance, or channel interference [1-4]. The results are based on the recent analytical research by used laboratory tests and theoretical analysis.

## II. MATHEMATICAL ANALYSIS

This mathematical analysis is made according to the international standard requirements [1], [2] for terrestrial radio and television broadcasting and the characteristic parameters and features of the broadcasting in Republic of Bulgaria. The researches are for the main figures of merit and factors of OFDM modulation signal with channel frequency and 16-QAM/ 64-QAM modulation of the subcarriers.

### A. Calculation of $E_b/N_0$

$$E_b/N_0 \text{ [dB]} = C/N - k_{RS} - k_{QAM}, \text{ where} \quad (1)$$

$E_b/N_0$  is the energy per useful information bit  $E_b$  referred to the normalized noise power  $N_0$ ;

$C/N$  - carrier-to-noise ratio in-channel [dB];

$k_{RS}$  - the factor for FEC to Reed-Solomon [dB];

$k_{QAM}$  - the factor for the QAM modulation [dB].

When we study the dependence of  $E_b/N_0$  of  $S/N$  can be used  $k_{r-f}$ , which gives the relationship between  $C/N$  and  $S/N$ :

$$S/N \text{ [dB]} = C/N + k_{r-f}, \text{ where} \quad (2)$$

$k_{r-f}$  - the factor for the roll-off Nyquist filtering in the demodulator/receiver.

Then Eq.1 can be written as following:

$$E_b/N_0 \text{ [dB]} = S/N - k_{r-f} - k_{RS} - k_{QAM}. \quad (3)$$

The values and the expressions for their obtaining for every parameter and factor are given in Table 1, where  $\alpha$  is the roll-off factor by Nyquist skew,  $b$  is the number of bits per symbol and  $R_{RS}$  is Reed-Solomon code rate.

TABLE I  
PARAMETERS AND FACTORS FOR QAM

modulation Parameters and Factors	16-QAM		64-QAM	
	$M$	16		64
$\alpha$	0,10	0,30	0,10	0,30
$R_{RS}$	188/204			
$b = \log_2 M$	4		6	
$k_{r-f} = 10 \lg(1-0,25\alpha)$	-0,1169	-0,3345	-0,1169	-0,3345
$k_{RS} = 10 \lg R_{RS}$	- 0,3547			
$k_{QAM} = 10 \lg b$	6,0206		7,7815	

### B. Calculation of BER

For the calculation of the bit-error-rate (BER) for OFDM DVB-T signal with 16-QAM and 64-QAM modulations of the subcarriers we used expressions [1], [2]:

$$P_B = \frac{2^{(b-1)}}{2^b - 1} \cdot P_S \text{ or} \quad (4)$$

$$P_B = \frac{1}{b} \cdot P_S, \quad (5)$$

which give the relationship between Symbol Error Probability  $P_S$  and the Bit Error Probability  $P_b$  of QAM with  $M$  constellation points, arranged in a rectangular set, for  $b$  even.

The Eq.4 makes no assumption about the constellation mapping and is based on the probability that any particular bit in a symbol of  $b$  bits is in error, given that the symbol itself is in error. The Eq.5 assumes that an erroneous symbol contains just one bit in error. This assumption is valid as long as a Gray coded mapping is used and the BER is not too high. These equations give different results for symbols of two or more bits. The second approach is generally adopted because DVB systems employ Gray code mapping.

<sup>1</sup>Oleg B. Panagiev is with the Technical University of Sofia, Bulgaria, E-mail: olcomol@yahoo.com.

As per Eq.5 when the received signal is distraught by Additive White Gaussian Noise (AWGN) there is a probability that any particular symbol will be wrongly decoded into one of the adjacent symbols. The  $P_S$  is given by:

$$P_S(E_b/N_0) = 2 \cdot \left(1 - \frac{1}{\sqrt{M}}\right) \cdot \operatorname{erfc} \left[ \sqrt{\frac{3b \cdot (E_b/N_0)}{2 \cdot (M-1)}} \right] \times \left\{ 1 - \frac{1}{2} \cdot \left(1 - \frac{1}{\sqrt{M}}\right) \cdot \operatorname{erfc} \left[ \sqrt{\frac{3b \cdot (E_b/N_0)}{2 \cdot (M-1)}} \right] \right\}, \quad (6)$$

where  $\operatorname{erfc}(x)$  is the complimentary error function given by:

$$\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt. \quad (6a)$$

For practical purposes Eq.6 can be simplified by omitting the (generally insignificant) joint probability term to give the approximation:

$$P_S(E_b/N_0) = 2 \cdot \left(1 - \frac{1}{\sqrt{M}}\right) \cdot \operatorname{erfc} \left[ \sqrt{\frac{3b \cdot (E_b/N_0)}{2 \cdot (M-1)}} \right]. \quad (7)$$

This approximation introduces an error which increases with degrading  $E_b/N_0$ , but is still less than 0,1 dB for 64-QAM at  $E_b/N_0 = 10$  dB. As already stated, the above equations for Symbol Error Probability are based on certain simplifying assumptions which can be summarized as "the system is perfect except for the presence of additive white Gaussian noise", but within this rather generous constraint the equations for  $P_S$  are exact.

### C. Calculation of EVM and MER

$EVM$  and  $MER$  take into account the combined effects of  $C/N$ ; transmitter, upconverter phase noise; impairments such as second and third order distortions; group delay; in-channel frequency response problems (amplitude tilt or ripple) and microreflections.  $EVM$  and  $MER$  measure essentially the same quantity and easy conversion is possible between the two measures if the constellation is known.

#### 1) Calculation of EVM

For calculation of  $EVM$ , we use the relationship between  $EVM$  and  $S/N$  [5]. After performing the appropriate mathematical transformations we obtain the expression for  $EVM$ :

$$EVM = 10^{-\frac{S/N+k}{20}} \cdot 100\%, \quad \text{where} \quad (8)$$

$k$  is the peak-to-average energy ratio (Table II). The Eq. 8 used to calculate the  $EVM$  is directly from vector considerations of signal plus noise in relationship to the ideal constellation points of any M-QAM signal. The minus sign in Eq. 8 is necessary because  $S/N$  is the ratio of signal to noise whereas  $EVM$  is the ratio of noise to signal.

TABLE II  
EVM VALUES FOR VARIOUS MODULATIONS

modulation	$k$
	dB
16-QAM	2,5527
64-QAM	3,6796

#### 2) Calculation of MER

$MER$  is the preferred measurement for the following reasons [1]:

- The sensitivity of the measurement, the typical magnitude of measured values, and the units of measurement combine to give  $MER$  an immediate familiarity for those who have previous experience of  $C/N$  or  $S/N$  measurement;
- $MER$  can be regarded as a form of  $S/N$  measurement that will give an accurate indication of a receiver's ability to demodulate the signal, because it includes, not just Gaussian noise, but all other uncorrectable impairments of the received constellation as well;
- If the only significant impairment present in the signal is Gaussian noise then  $MER$  and  $S/N$  are equivalent (they are often used interchangeably).

The relationship between  $EVM$  and  $MER$  given by the expression:

$$EVM_V = 1/(MER_V \cdot V), \quad (9)$$

where the peak to mean voltage ratio  $V$ , is calculated over a large number of symbols (10 times the number of points in the constellation is adequate if the modulation is random) and each symbol has the same probability of occurrence then it is a constant for a given transmission system. The value tends to a limit which can be calculated by considering the peak to mean of all the constellation points. Table III lists the peak-to-mean voltage ratios for the DVB constellation sizes. When expressed as simple voltage ratios  $MER_V$  is equal to the reciprocal of the product of  $EVM_V$  and the peak-to-mean voltage ratio for the constellation.

TABLE III  
PEAK-TO-MEAN RATIOS FOR THE DVB CONSTELLATION SIZES

modulation	Peak-to-mean voltage ratio
	$V$
16-QAM	1 341
64-QAM	1 527

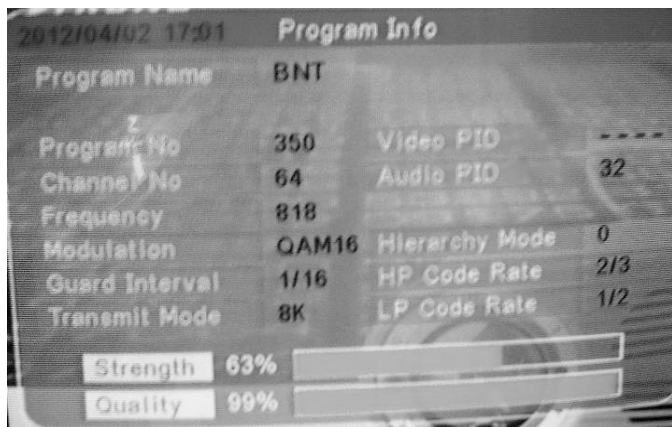
## III. SIMULATION INVESTIGATION

In this section the simulated researches, of main parameters of DVB-T signals with 16-QAM and 64-QAM modulations of the subcarrier, are made for an existing (real-life) channels in the area of the city of Sofia (Figs.1 and 2). The research uses the mathematical dependences from above and its results are given in tables and graphics.

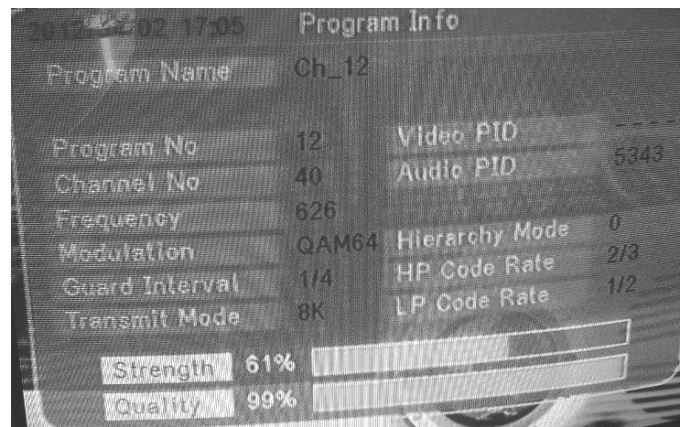
On Fig.1a is shown a picture with the parameters of received television channel by D/K standard at a frequency of 818 MHz (ch.64), and on Fig1b – the signal's spectrum. Table IV contains the numeric results from the simulation research for  $E_b/N_0$ ,  $SER$ ,  $BER$  and  $EVM$  at an amendment of  $C/N$  from 2÷30dB and respectively  $S/N$  from 1,88-29,88dB at  $\alpha=0,1$  [6].

Fig.2a shows a picture with the parameters of received television channel by D/K standard at a frequency of 626 MHz (ch. 40), and Fig2b - the signal's spectrum. Table V gives the numeric results of the simulation research [7] for  $E_b/N_0$ ,  $SER$ ,

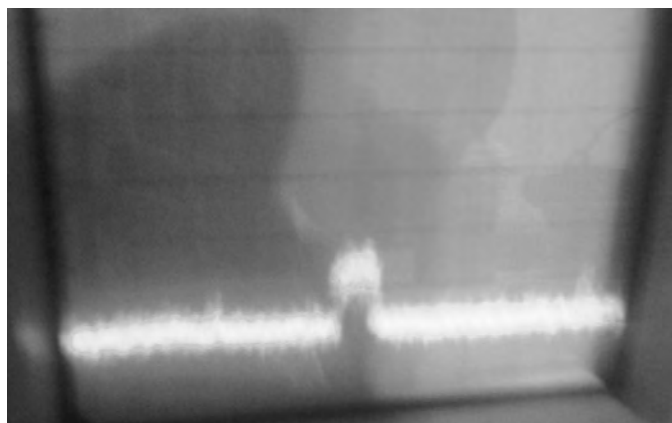




a) Menu for ch.64 (16-QAM)

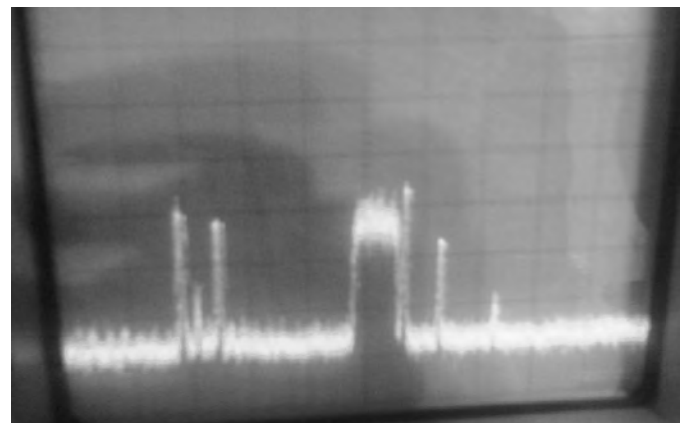


a) Menu for ch.40 (64-QAM)



b) Spectrum ch.64 (DVB-T)

Fig.1. ch.64 (DVB-T)



b) Spectrum ch.40 (DVB-T), ch. 36 and ch.41 (PAL-K)

Fig.2. ch.40 (DVB-T)

TABLE IV  
16-QAM VALUES

C/N dB	S/N dB	$E_b/N_0$ dB	SER	BER	EVM %
2	1,8831	-3,6659	0,3743	0,1995	60,01
4	3,8831	-1,6659	0,3153	0,1681	47,67
6	5,8831	0,3341	0,2469	0,1316	37,86
8	7,8831	2,3341	0,1732	0,0923	30,08
10	9,8831	4,3341	0,1027	0,0547	23,89
12	11,8831	6,3341	0,0472	0,0251	18,98
14	13,8831	8,3341	0,0146	7,79E-03	15,07
16	15,8831	10,3341	2,46E-03	1,31E-03	11,97
18	17,8831	12,3341	1,61E-04	8,60E-05	9,51
20	19,8831	14,3341	2,39E-06	1,27E-06	7,56
22	21,8831	16,3341	3,37E-09	1,80E-09	6,00
24	23,8831	18,3341	1,16E-13	6,17E-14	4,77
26	25,8831	20,3341	0	0	3,79
28	27,8831	22,3341	0	0	3,01
30	29,8831	24,3341	0	0	2,38

TABLE V  
64-QAM VALUES

C/N dB	S/N dB	$E_b/N_0$ dB	SER	BER	EVM %
2	1,8831	-5,4268	0,5629	0,2871	52,71
4	3,8831	-3,4268	0,5298	0,2702	41,87
6	5,8831	-1,4268	0,4878	0,2488	33,26
8	7,8831	0,5732	0,4351	0,2219	26,42
10	9,8831	2,5732	0,3704	0,1889	20,98
12	11,8831	4,5732	0,2941	0,1500	16,67
14	13,8831	6,5732	0,2103	0,1073	13,24
16	15,8831	8,5732	0,1281	0,0654	10,52
18	17,8831	10,5732	0,0611	0,0312	8,35
20	19,8831	12,5732	0,0200	0,0102	6,64
22	21,8831	14,5732	3,68E-03	1,88E-03	5,27
24	23,8831	16,5732	2,76E-04	1,41E-04	4,19
26	25,8831	18,5732	5,03E-06	2,56E-06	3,33
28	27,8831	20,5732	9,89E-09	5,04E-09	2,64
30	29,8831	22,5732	5,74E-13	2,93E-13	2,10

BER and EVM at an amendment of C/N from 2÷30 dB and respectively S/N from 1,88-29,88 dB at  $\alpha=0,1$ .

The numeric results are shown graphically in Figs.3, 4 and 5, while in both cases (16-QAM/ 64-QAM) after the Viterbi and Reed-Solomon decoding BER is low ( $10^{-9} \div 10^{-11}$ ) - for S/N between 22 and 28 dB. That confirms itself through the

indication for Quality on Figs.1a and 2a, which reaches 99%.

For low values of S/N (1,88÷19,88 dB) BER is very high and the receiving is either with errors or not possible, and EVM amends from 7,3% to 0,92%. For  $S/N \geq 23$  dB EVM values are almost equal. At an  $S/N = 29,88$  dB for 16-QAM is  $EVM=2,38\%$  and for 64-QAM is  $EVM=2,1\%$ .

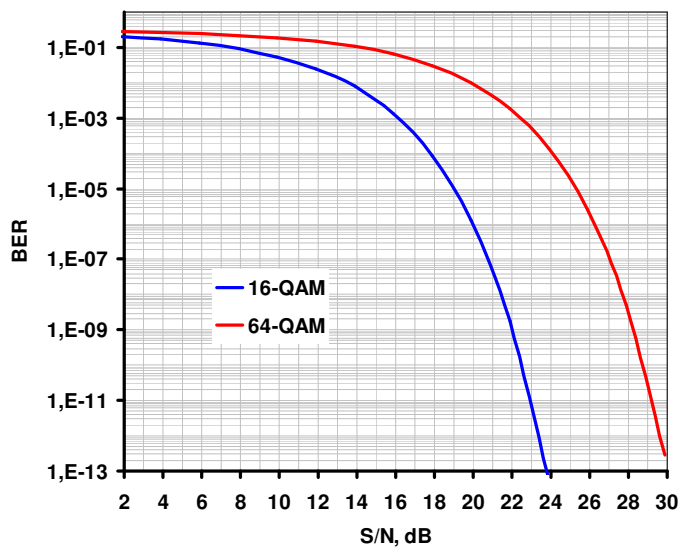


Fig.3.  $BER = func(S/N)$

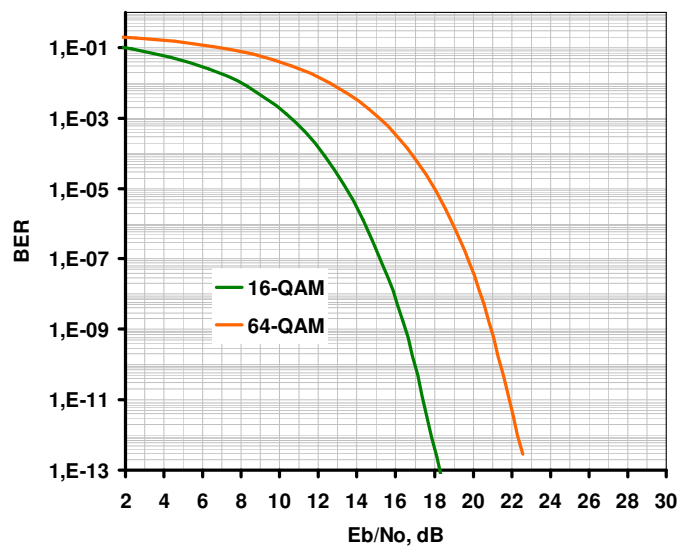


Fig.4.  $BER = func(E_b/N_0)$

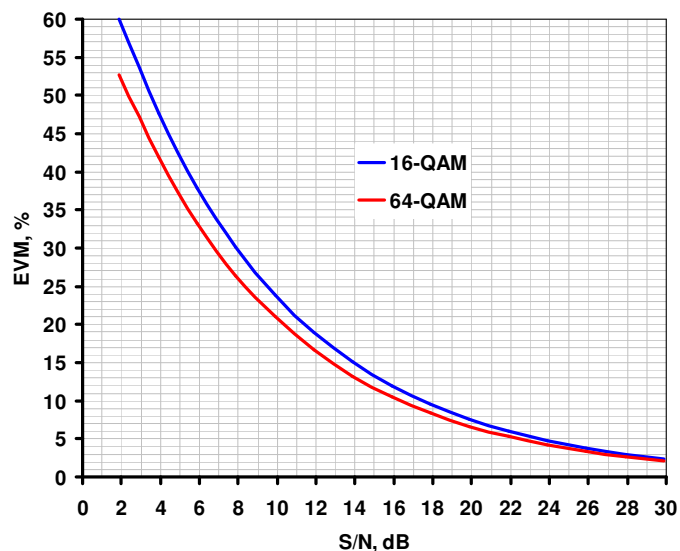


Fig.5.  $EVM = func(S/N)$

The operation mode is in both cases 8k and the modulation non-hierarchical. There is also a difference between the levels of received signals (level of ch.64 is less than a one of ch.40). Near to the ch.40 is transmitted an analog television channel (ch.41 with picture carrier 631,25 MHz - PAL-K), which in some DVB-T receivers causes interference that deteriorate the researched parameters and mostly *BER* (which increases). That makes receiving ch.40 impossible. Methods and means for the elimination of this problem and rising of *S/N* and reduction of *BER* are an object of review in the second paper at the conference.

#### IV. CONCLUSION

It should be mentioned that for QAM systems DVB only employs Gray coding within each quadrant, the quadrant boundaries are not Gray coded, and the mapping is partially differentially coded. Further work is required to establish the exact  $P_S$  to  $P_B$  relationship for this combination of mapping and coding [1], [2], [8], [9].

#### REFERENCES

- [1] DVB Document A014 Rev. 2, "MEASUREMENT GUIDELINES FOR DVB SYSTEMS", 2001.
- [2] ETSI EN 300 744 V1.4.1 (2001-01), Framing structure, channel coding and modulation for digital terrestrial television, 2001.
- [3] E. Benoit, *Digital television*, 3<sup>rd</sup> ed., Focal Press by Elsevier, Oxford, 2008.
- [4] K. Koichev, S. Sadinov, G. Traikovski. Defining the channel parameters and their influence on SER for OFDM transmission. ICEST 2007, Proc. of Papers, Ohrid, Macedonia, 2007, pp.433-436.
- [5] Technical reference for the RF engineer from the RF experts, TechNotes, AXCERA LLC, 05.12.1999.
- [6] <http://www2.rohde-schwarz.com>, application note, 2005-2011.
- [7] <http://www.wolframalpha.com>
- [8] R. Hranac, "Carrier-to-Noise versus Signal-to-Noise", Cisco Systems, sisco.com, 2003.
- [9] E. Lecomte, "DVB T/H Seminaire", Agilent technologies, 2008.

# Improving the reception of class DVB-T receivers

Oleg Borisov Panagiev<sup>1</sup>

**Abstract** - This paper presents some early results from experimental measurements in the service area of the DVB-T (8K) network. The aim of this study has been to determine the degradation suffered by digital signals in presence of analog PAL-K adjacent channel broadcasting. The result of this study is a circuit solution and curve that estimates protection ratios between analog and digital received power levels.

**Keywords** – DVB-T, PAL-K, receiver, adjacent channel interference, BER.

## I. INTRODUCTION

Recently digital terrestrial TV using the DVB-T standard [1] is gaining more and more interest in Europe and other parts of the world. This fact leads to a simulcast environment during a period of several years, depending on different countries and estimations, in which analog and digital services will be sharing the same spectrum frequency bands. During the transitional period due to the coexistence of both technologies, minimizing the digital service quality degradation caused from analog transmitters (and vice versa) in the same coverage areas will be one of the main factors to consider when planning digital services.

At the same time manufacturers offer DVB-T receivers with different application: domestic, portable, notebook, PC (Fig.1), built-in TV. Some are with low adjacent channel selectivity and lower sensitivity. This in many cases leads to poor or impossible to receive digital signals in the presence of a adjacent (upper) channel, low level of the received signal and others [2], [3].

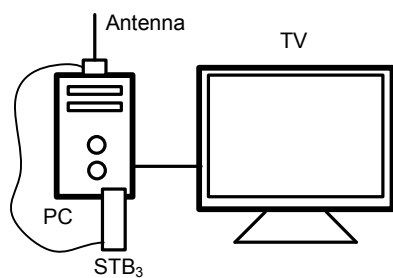


Fig.1. DVB-T reception with personal computer (PC) -TV tuner

In this paper are presented the results of theoretical and experimental investigations carried out to study the relationship between BER and protection ratio (S/N respectively) for portable reception in an urban environment inside the coverage of a Single Frequency Network (SFN). Two types of portable reception have been defined by the European Standards Institute (ETSI): Class A and Class B portable reception [4]. Class B portable reception is defined as indoor (inside a building) reception whereas Class A is

defined for outdoor reception using an omnidirectional antenna located at least at 1.5 meters above ground level [5]. The results presented here apply to both cases.

## II. EXPERIMENTAL AIMS AND TASKS

The formulation of the aims and the tasks is made on the basis of trials on several (different) DVB-T receivers. Further those would be called for shorter set-top-box (STB). Two of them are for connecting to a TV receiver (television set): one (STB1) is connected through an additional Eurocart-Eurocart or Eurocart-Chinch cable; the other one is connected directly to Eurocart of the TV receiver (television set), whereby it is supplied through an adapter. The infrared (IR) receiver is a separate module, and its connection with the DVB-T receiver is through a cable, which allows controlling the DVB-T receiver even though it is behind the TV receiver. The third STB3 is connected to a PC through a USB port. The dialog with it is fulfilled via the display (monitor) of the PC, as for the functioning of the DVB-T receiver we have to install certain software (sold in a CD together with the DVB-T receiver).

The three DVB-T receivers were connected respectively to a TV and a PC, then set and every digital TV channel that is broadcasted in the region of the city of Sofia was chosen. The signals were received by three different antennas:

- 1) Active antenna with G=45dB;
- 2) Yagi with G=18dB;
- 3) Whip

Each one of the antennas was successively connected to STB1, STB2 and STB3 and the levels of the signals with the respectively carrier frequency were measured by STB in % as well as with a level strength meter (LM) in dBμV. The results of the measurement are given in Table I.

*Note:* Antennas 1) and 2) are with a direct visibility to the transmitter centers (Kopitoto and the old TV tower).

After the measurements were done was established that STB1 and STB2 receive all digital channels without any problems, qua the strength and quality indicators are with evidence ≥50% independently of the antenna type. However, STB3 had a problem receiving on ch. 40 and ch. 52, although the conditions for receiving were the same as for STB1 and STB2.

TABLE I

Antenna	1)			2)			3)		
	Strength	Quality	Level	Strength	Quality	Level	Strength	Quality	Level
Channel	%	%	dBμV	%	%	dBμV	%	%	dBμV
40	61	99	91,2	63	99	68,1	60	99	58
41	-	-	96,8	-	-	74	-	-	63,8

<sup>1</sup>Oleg B. Panagiev is with the Technical University of Sofia, Bulgaria, E-mail: olcomol@yahoo.com

On the base of those results were formulated the aims and tasks of this research:

a) To be established the reason for not receiving ch.40 and ch.52;

b) To eliminate the problems with receiving ch.40 and ch.52, i.e. to be provided their receiving with STB3 and other similar to it [2], [3].

To accomplish those aims were made a number of measurements according to the standards [1], [4], [6] and the results were analyzed. As a result of that was made a conclusion, that such a problem could appear at a low selectivity at adjacent channel and by an interference of analog signal (PAL-K), when it is transmitted in higher adjacent channel (by measurements with a spectrum analyzer, was established, that such channels are available: ch.41 and ch.53 at standards G and K), (Fig.2).

To increase the selectivity at a adjacent channel (without changing anything in the STB3) was made the decision for creating a band filter with a certain slope and frequency band, which would be formed in a different module. Its assembly must be outside STB3 and to be applicable to different STBs with similar parameters.

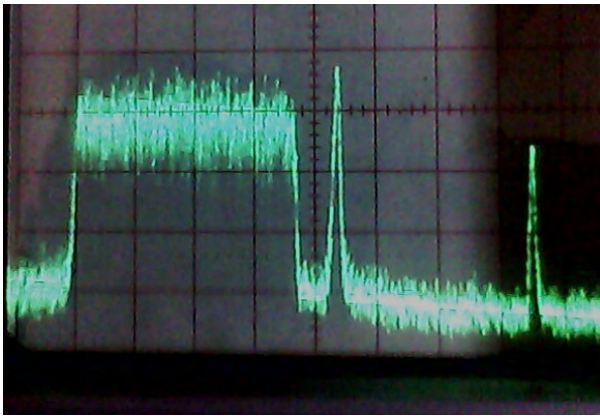


Fig.2. Spectrum of ch.40and ch.41

The first task was to design a BPF and simulate the influence of its parameters onto  $S/N$  and  $BER$ , respectively  $QoS$  (Quality of signal).

The second task is to be made an actual BPF to be assembled between the antenna and STB3.

The third task is to be checked the receiving of ch.40 and ch.52 (strength and quality) and to be made an analysis for the application of the proposed solution.

### III. CIRCUIT SOLUTIONS AND RESULTS

For researching the influence of the parameters of Band Pass Filter (BPF) on improving the quality of receiving for the aforementioned channels, as well as for optimal choosing of suitable scheme solution (Table II) was used an Ansoft Designer VS 2.0 [7]. A large number of simulations were made to determine the scheme solution and the number of links in filter, which give the necessary values of Protection ratio ( $PR$ ), [1], [4], [6], [8]. The theoretical and experimental researches were made with five values for  $PR$ , where in dB,

is the degradation in Power received between digital TV signal ( $P_{rD}$ ) and analog TV signal ( $P_{rA}$ ) and is given by the following formula [9]:

$$PR[dB] = P_{rD}[dBW] - P_{rA}[dBW] \quad (1)$$

Eq.1 could be written as follows, having in mind the connections between power and voltage [10] by impedance  $75\Omega$ :

$$PR[dB] = U_{rD}[dB\mu V] - U_{rA}[dB\mu V]. \quad (2)$$

The practical determination of  $PR$  is made due to Eq.2 for ch.40 and measured levels of two signals (digital and analog) via spectrum analyzer (SA) Promax AE-476 (Fig.3), according the requirements of [7].

*Note:* During the simulation was established that the solutions of the placed aims for ch.40 are also valid for ch.52, but at corresponding values of constructive elements (coils and capacitors) and frequency parameters of band pass filter.

The dependence of deterioration of the signal-to-noise ratio ( $\Delta S/N$ ) from the amendment of  $PR$  is shown on Fig.4. The resultant ratio signal-to-noise ( $S/N_i$ ) at a presence of interference from the analog channel in the digital one are calculated through the following formula:

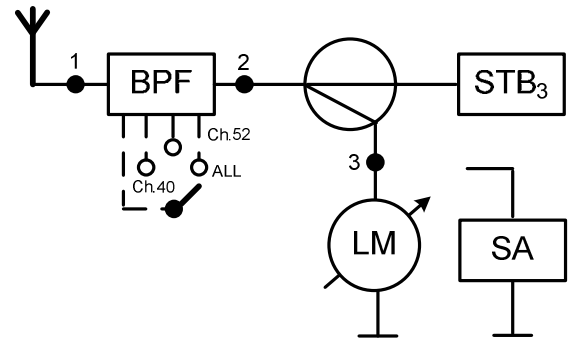


Fig.3. Measurement setup

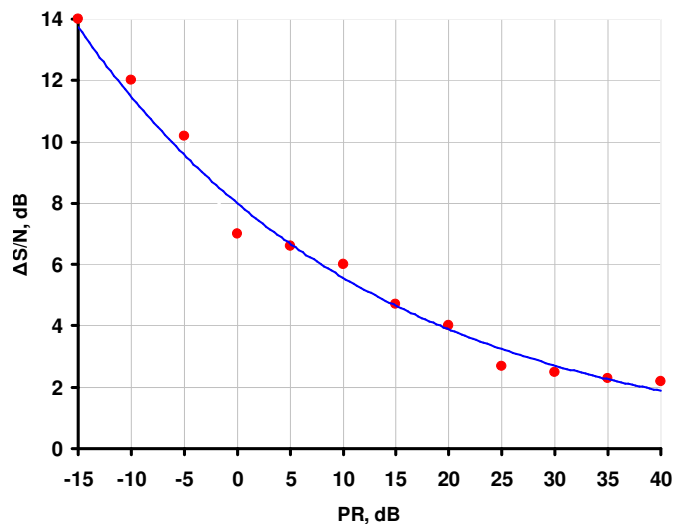
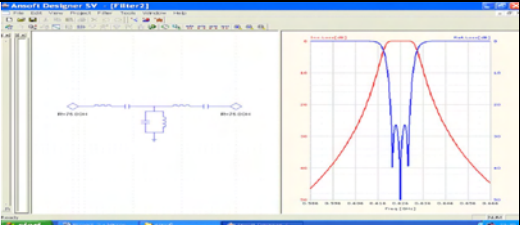
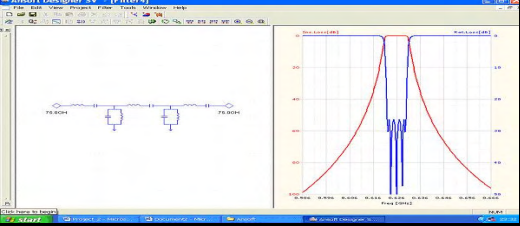
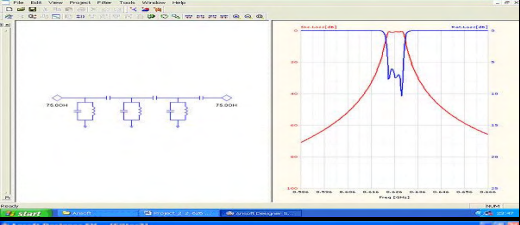
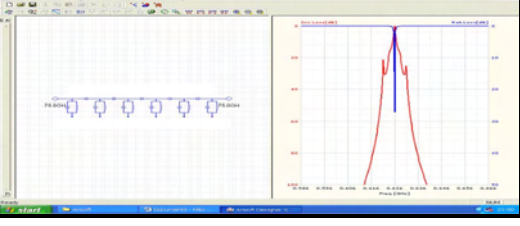


Fig.4.  $\Delta S/N[dB] = \text{func}(PR [dB])$

$$S/N_i [dB] = S/N [dB] - \Delta S/N [dB], \quad (3)$$

TABLE II

№	BPF	PR dB	BER					
			Experimental			Theoretical		
			1)	2)	3)	1)*	2)*	3)*
1	no	-6	0,0297955	0,0153234	0,0787797	0,0150099	0,0076909	0,0402155
2		1	0,0024338	0,0006898	0,0153234	0,0012177	0,000345	0,0076909
3		12	6,967E-05	8,322E-06	0,0015181	3,484E-05	4,161E-06	0,0007593
4		23	3,21E-07	1,009E-08	4,724E-05	1,61E-07	5,046E-09	2,362E-05
5		35	3,095E-08	5,401E-10	1,052E-05	1,547E-08	2,701E-10	5,259E-06

where  $S/N$  is signal-to-noise by the absence of interference from the analog PAL-K signal into the digital DVB-T signal (cannel).

The theoretical determination of BER is made through:

$$P_B = 0,89.[erfc(\sqrt{10^{(S/N_I - 7,43)/10}} / 7)] \times [1 - 0,44.erfc(\sqrt{10^{(S/N_I - 7,43)/10}} / 7)]. \quad (4)$$

The dependence of BER from PR is given in Fig.5 wherefrom is visible, that with antennas 1) and 2) and BPF the receiving of ch.40 is already possible, when  $PR \geq 15$ dB. At values less than 15dB, receiving with active filter (Fig.3) and any of the antennas is difficult (presence of sampling) or impossible (Fig.6).

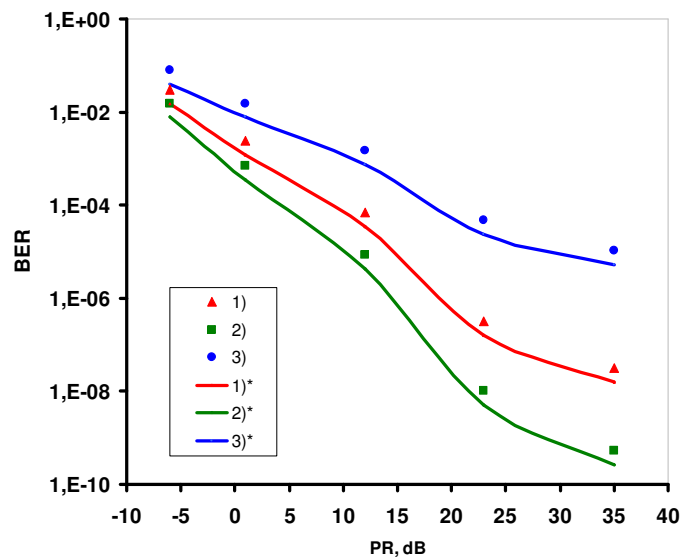
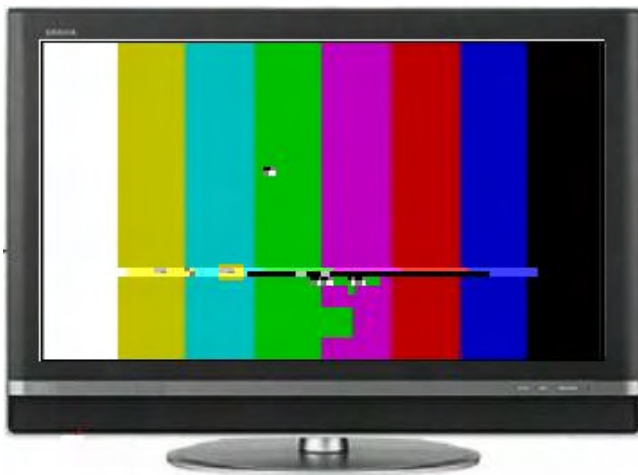


Fig.5. BER = func (PR [dB])



a)  $BER > 10^{-4}$



b)  $10^{-6} < BER < 10^{-4}$



c)  $BER < 10^{-6}$

Fig.6. Pictures with BPF for different PR and BER

#### IV. CONCLUSION

The offered solution for improving the receiving with DVB-T receivers (computer TV-tuners) gives the opportunity to achieve the (placed) aims. By the practical realization is needed to change the filters for ch.40 and ch.52 or to remove them, and points 1 and 2 (Fig.3) to be connected directly with each other in order to receive all other channels. On Fig.3 is offered a version with shift of the three regimes. BPF can be realized with varicaps so that the passing channel is changed electrically. However, this solution requires an additional supply and so is the practical realization elaborated. At a subsequent publication will be offered another version, by which the mentioned problems would be removed and the practical application of the offered solution will be facilitated.

#### REFERENCES

- [1] ETSI EN 300 744 v.1.4.1, Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for digital terrestrial television, 2001.
- [2] DIGIVOX mini Hybrid TV tuner and video capture unit. MICRO-STAR INT'L CO., LTD.
- [3] USB TV tuner AVerTV Hybrid Volar HD. AVer Media Technology.
- [4] TR 101 190: "Digital Video Broadcasting (DVB); Implementation Guidelines for DVB Terrestrial Services: Transmission Aspects". European Telecommunications Standards Institute (ETSI), 1997.
- [5] P. Angueira, M. Vález, D. De La Vega, G. Prieto, D. Guerra, J. M. Matías and J.L. Ordiales, "DTV reception quality field tests for portable outdoor reception in a single frequency network", IEEE Trans., Broadcasting, vol.50, no.1, pp.42-48, 2004.
- [6] Rec. ITU-R BT. 1368-2 "Planning Criteria for Digital Terrestrial Television Services in the VHF/UHF Bands", 2000.
- [7] <http://www.ansys.com/Products/Simulation+Technology/Electromagnetics/Ansoft+Designer>.
- [8] S. O'Leary "Digital/Analogue Co-channel Protection Ratio Field Measurements" IEEE Trans., Broadcasting, vol. 44, no. 4, pp. 540-546, 1998.
- [9] Regional Radiocommunication Conference for planning of the digital terrestrial broadcasting service (RCC-04), ITU-R Report, Geneva 2004.
- [10] D. Dobrev, L.Yordanova, *Receiving on radio- and TV programs by means satellite and cable*, Sofia, Electroninvest, 1996.

# Monolithic Integrated Antennas with High Radiation Efficiency

Hristomir Yordanov<sup>1</sup>

**Abstract**— The wireless chip-to-chip interface is an interesting alternative to wired digital buses. An important element from a wireless interconnect is the monolithic integrated antenna. On-chip antennas have to meet several restrictions—they have to cover very little chip area and they need to have high radiation efficiency in a large enough frequency range to allow for fast inter-chip communication. This work describes two methods for designing efficient on-chip antennas, namely using high-impedance substrate with standard thickness, or using very thin substrate.

**Keywords**— Embedded antennas, Near-field communication, On-chip integrated systems.

## I. INTRODUCTION

The wired inter-chip interconnects within a system are becoming a substantial bottle-neck for the chip-to-chip digital communications. The wired bus suffers frequency bandwidth limitations due to auto-interference effects like cross-talk and dispersion [1]. These effects can not be reduced by optimising the design of the printed circuit board, because the limitations of that technology have long been met. The standard approach for optimising inter-chip communication is based on using parallel lines and using software techniques like data buffering. The cost of these solutions is increased circuit board complexity. Parallel lines connecting multiple devices result in complicated board layout with very long design time—it is possible to have a multi-layer board design time up to several months.

An interesting alternative to wired chip-to-chip interface is the wireless communication. An efficient wireless interconnect has the potential to offer data rate higher than its wired counterpart, while keeping the system design much simpler and area-efficient [2].

The technical requirements for a wireless interconnect are specified by several conditions. First, the whole transceiver circuitry including the antennas must be integrated within the chip. Second, wireless link must provide sufficient bandwidth for fast enough communication. And third, the range of the link should be very short—a distance of several centimetres is to be covered. Therefore a suitable carrier frequency in the millimetre range can be used to design good chip-to-chip wireless link. The RF circuitry required for the transceivers can be readily implemented even in 90 nm CMOS technology [3], [4], the small wavelength allows for on-chip antenna integration and the high frequency allow for higher absolute bandwidth and therefore higher data rates.

<sup>1</sup>Hristomir Yordanov is with the Electromagnetic Compatibility Lab at the Technical University of Sofia, Faculty of Telecommunications, Kliment Ohridski Blvd. 8, Sofia 1000, Bulgaria, E-mail: h\_yordanov@tu-sofia.bg

There are two main problems in integrating the antenna within a CMOS chip. First is the area covered by the antenna. Normally this area is several square millimetres, which is unacceptable by integrated circuit manufacturers. The second problem is the antenna radiation efficiency. The silicon substrate used in CMOS technology is lightly p<sup>-</sup> doped, which decreases its resistivity to several Ohm.cm. The electromagnetic field within such substrate is subject to heavy dielectric losses due to this low resistivity, which reduces the antenna efficiency.

The first of these two problems can be tackled by double using available metallisation structures on the integrated circuit. For example the chip ground supply planes can simultaneously serve as antenna electrodes [5]–[7]. This has been achieved by cutting the top metallisation layer into patches and feeding the RF signal across the gap between the patches.

The efficiency of the integrated antenna can be optimised by two methods. One is to use high-resistivity substrate [6]. A drawback of this method is that the thermal conductivity of the substrate is also reduced and this changes the standard CMOS technology. Another interesting option, presented in detail in the current work, is to use very thin substrates. Current technologies like the ChipFilm technology, developed by the Institute for Microelectronics in Stuttgart, Germany, offers the possibility to fabricate integrated circuits on substrates as thin as 6 μm [8]–[10]. Such a substrate increases the antenna efficiency by reducing the volume where the electromagnetic field is subject to a loss. A drawback of this solution is the introduction of several additional steps to the standard CMOS technology.

## II. ANTENNA MODE

A cross-section view of the integrated antenna using the circuit ground plane as electrodes is presented in Fig 1. The antenna is fabricated either on high-resistivity (about 1 kΩ.cm) substrate with thickness of 675 μm, or on thin low-resistivity substrate, as discussed in the previous section. The active elements are fabricated atop the substrate. Several metallisation layers follow, where the on-chip interconnects are fabricated. The top metallisation layer holds the ground supply plane. This plane is cut into patches and an RF generator is connected across the resulting gaps, exciting the antenna.

Since the patches serve as ground planes also for the power supply for the CMOS circuitry underneath them, a low-frequency galvanic connection must be provided between them. The block inductors, providing a DC connection between the patches are not shown in the figure.

The antenna structure consists of two or more patches, separated by narrow gaps. If the length of the gaps is comparable with the wavelength, they can be treated as slot lines. If these

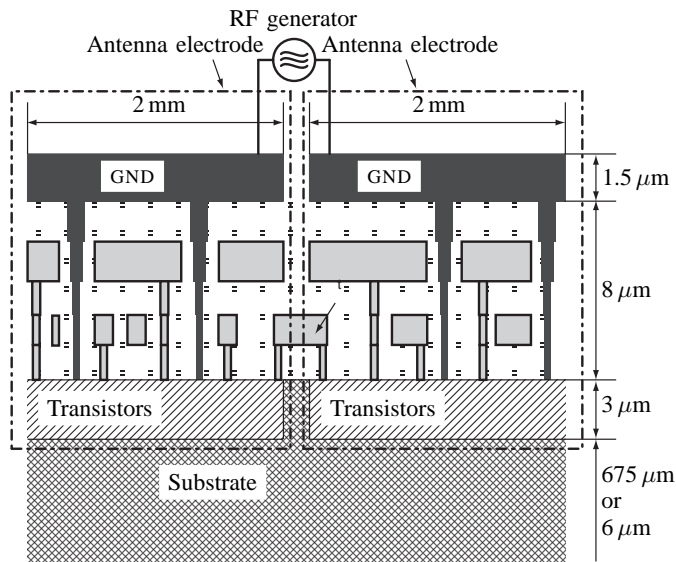


Fig. 1. Detailed view of the cross-section of the integrated on-chip antenna, using the ground planes as antenna electrodes. The separated areas of the ground planes have to be connected to each other using inductive connections. The RF generator is also integrated in the CMOS circuit. Figure is not to scale.

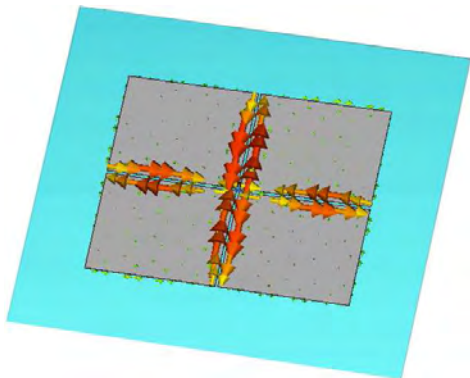


Fig. 2. Current distribution in an  $2 \times 2$  patch antenna configuration.

lines are terminated with a open or short-circuit, a standing wave pattern will be formed along the line. This standing wave provides a time-varying electric polarisation, which is a source of radiation [11]. Figure 2 shows the current distribution in a four-patch antenna configuration.

The guided wavelength in a slot line is given by [12]

$$\lambda_g = \frac{\lambda_0}{\sqrt{\frac{1}{2}(\epsilon_{r,Si} + 1)}}, \quad (1)$$

where  $\lambda_0$  is the free-space wavelength and  $\epsilon_{r,Si}$  is the relative permeability of the substrate. For a carrier frequency of  $f = 60$  GHz the corresponding wavelength is  $\lambda_g = 1.8$  mm. Therefore an open-circuited slot 1.35 mm long is a  $3/4\lambda_g$  resonator. Simulations of a two-patch structure, presented in Fig. 3 show that the length of the slot should be a bit shorter, namely 1.1 mm, accounting for the effective slot elongation due to the stray capacitance of the open slot line end. The current distribution, plotted in the figure, shows variation in the current density along the gap.

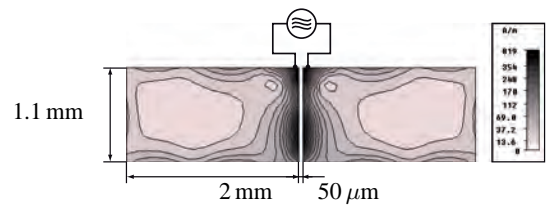


Fig. 3. Top view and current distribution of a two patch antenna, operating at 66 GHz.

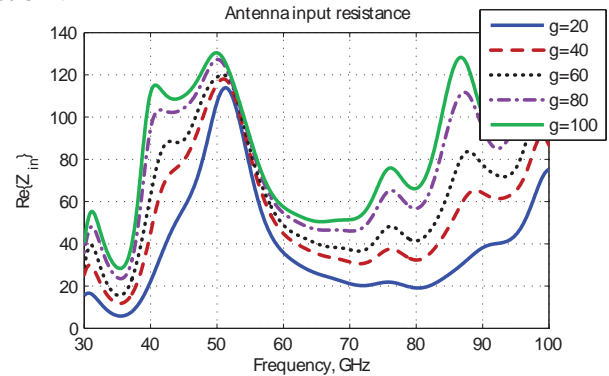


Fig. 4. Real part of the input impedance of the antenna from Fig. 3 for different slot widths  $g$ .

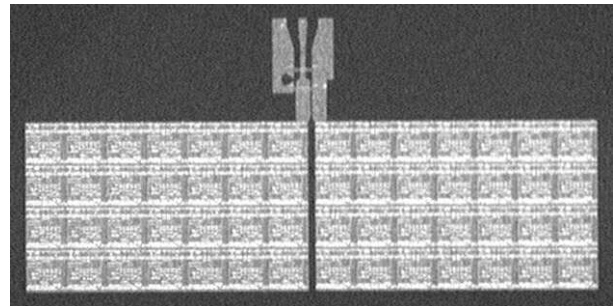


Fig. 5. Photograph of the manufactured antenna.

The input impedance of the antenna depends on the gap width. A plot of the real part of the two-patch slot antenna from Fig. 3 is shown in Fig. 4. The input resistance is smaller for a narrower slot.

The antenna has been manufactured on high-impedance substrate. A photograph of the structure is shown in Fig. 5. The results of the investigation of that type of antennas are presented in [5], [6].

### III. ANTENNAS ON THIN SUBSTRATE

The ChipFilm technology provides the possibility of manufacturing integrated circuits on substrate with thickness as low as 6 μm, as discussed in the introduction. Numerical experiments have been performed to assess the radiation efficiency of antennas on thin substrate.

The antenna loss and the antenna radiation are both modelled as resistors connected in series in the equivalent circuit of the antenna. Therefore it is not possible to identify the antenna efficiency by investigation of the input impedance. The efficiency can be computed numerically by comparing the quality factor of a lossless antenna model and the one of



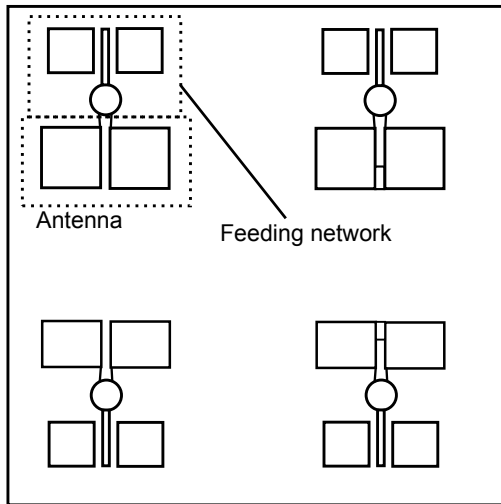


Fig. 6. A top view of the experimental setup.

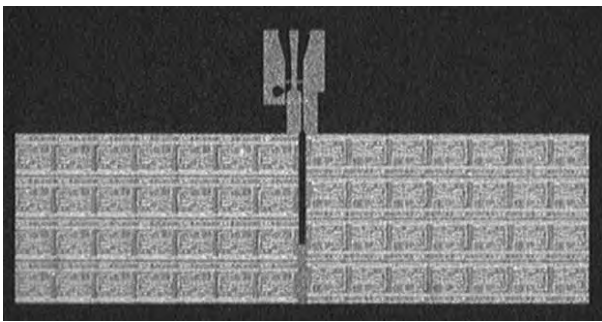


Fig. 7. A photograph of a short-circuited antenna.

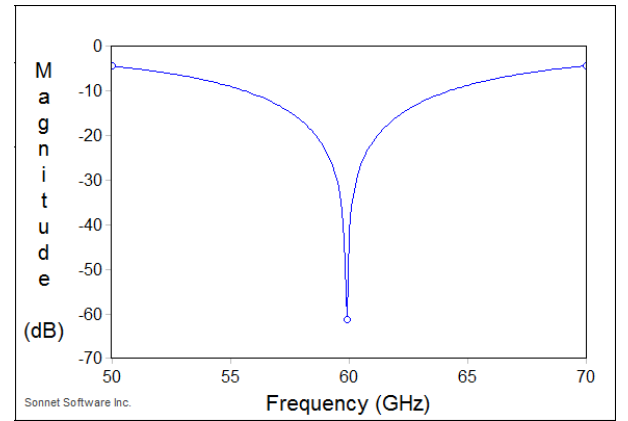


Fig. 8. Return loss of the lossy integrated antenna.

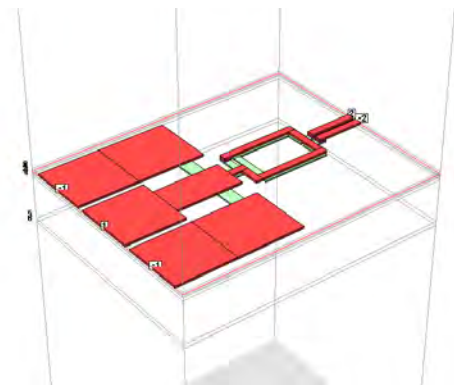


Fig. 9. A 3D model of the balun transformer.

#### IV. EXPERIMENTAL SETUP

a lossy antenna. The quality factor of any two-port is defined as

$$Q = \frac{P_{\text{Active}}}{2\pi E_{\text{Stored}}} = \frac{1}{\Delta f}, \quad (2)$$

or as a ratio between the active input power to the stored energy per cycle, which is inversely proportional to the  $-3$  dB bandwidth. Since the antenna input power is the sum of the radiated power and the power lost as heat, or

$$P_{\text{Active}} = P_{\text{rad}} + P_{\text{loss}}, \quad (3)$$

and the antenna efficiency is the ratio of the radiated to the input power,

$$\eta_{\text{Ant}} = \frac{P_{\text{rad}}}{P_{\text{Active}}}, \quad (4)$$

we can compute the antenna efficiency as the ratio of the  $-3$  dB bandwidth of a lossless and a lossy antenna,

$$\eta_{\text{Ant}} = \frac{\Delta f_{\text{lossless}}}{\Delta f_{\text{lossy}}} \quad (5)$$

The simulation results show antenna efficiency of  $\eta_{\text{Ant}} = 74.8\%$ . A plot of the return loss of the lossy antenna is shown in Fig. 8.

The model of antennas on thin substrate is to be verified by measurement. An experimental setup is prepared, which allows the measurement of the input impedance of the slot antenna. The setup contains four antennas, each equipped with a feeding network, as shown in Fig. 6. There are two types of antennas that will be investigated. The first type is the slot antenna, described in the previous section. The second type is the same slot antenna, but short-circuited at the end. Such a short-circuit provides a DC connection between the two patches, as they need to serve as a CMOS circuit ground plane. A photograph of such antenna is shown in Fig. 7. Various antenna lengths for both antenna types have been designed.

The feeding network contains a balun transformer, as the antenna is symmetrical, so that the feeding line should be balanced, whereas the measurement equipment provides unbalanced port connections. The balun is designed using a transformer. The offered technology provides two metallisation layers, so a stacked transformer design has been selected, as the stacked transformers provide lower insertion loss than the single-layered ones [13]. Figure 9 shows a 3D model of the designed balun.

The calibration of the measurement system must be performed using the Thru-Reflect-Line (TRL) [14] technique in order to de-embed the balun characteristics from the measurement results. This technique allows the calibration of

two-port measurement systems using three different, but not standardised, port loads: direct connection of the measurement ports at the reference plane (Through), connection of high-reflection-factor load (Reflect) and connecting the ports with a short line (Line).

Therefore calibration structures have also been designed on a different chip. The structures allow for the measurement of the input impedance of the antennas at their input ports, while de-embedding the influence of the balun transformers.

## V. SUMMARY

This work discusses the design of on-chip integrated antennas with high radiation efficiency and low area requirements. There are two possibilities for minimising the antenna losses—one is using high-resistivity substrate and the other is using very thin substrate. An experimental setup has been prepared to verify the characteristics of integrated antennas on thin substrate. The antenna input impedance and radiation efficiency have been computed.

## ACKNOWLEDGEMENT

This work has been supported by the Marie Curie Programme.

## REFERENCES

- [1] H. Yordanov, M. T. Ivrlač, A. Mezghani, J. Nossek, and P. Russer, "Computation of the impulse response and coding gain of a digital interconnection bus," in *24<sup>th</sup> Annual Review of Progress in Applied Computational Electromagnetics ACES*, Niagara Falls, Canada, Apr. 2008.
- [2] K. K. O, K. Kim, B. A. Floyd, J. L. Mehta, H. Yoon, C.-M. Hung, D. Bravo, T. O. Dickson, X. Guo, R. Li, N. Trichy, J. Caserta, W. R. B. II, J. Branch, D.-J. Yang, J. Bohorquez, E. Seok, L. Gao, A. Sugavanam, J.-J. Lin, J. Chen, and J. E. Brewer, "On-chip antennas in silicon ICs and their application," *IEEE Trans. Electron Devices*, vol. 52, no. 7, pp. 1312–1320, Jul. 2005.
- [3] C. Cao and K. K. O, "A 140-GHz fundamental mode voltage-controlled oscillator in 90-nm CMOS technology," *IEEE Microw. Wireless Compon. Lett.*, vol. 16, no. 10, pp. 555–557, Oct. 2006.
- [4] T. Yao, M. Q. Gordon, K. K. W. Tang, K. H. K. Yau, M.-T. Yang, P. Schvan, and S. P. Voinigescu, "Algorithmic design of CMOS LNAs and PAs for 60 GHz radio," *IEEE J. Solid-State Circuits*, vol. 42, no. 5, pp. 1044–1057, May 2007.
- [5] H. Yordanov and P. Russer, "Integrated on-chip antennas for chip-to-chip communication," in *Proceedings of the IEEE Antennas and Propagation Society International Symposium, 2008*, San Diego, CA, 2008.
- [6] —, "Integrated on-chip antennas using CMOS ground planes," in *Proceedings of the 10<sup>th</sup> Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems*, New Orleans, LA, 2010, pp. 53–56.
- [7] P. Russer, N. Fichtner, P. Lugli, W. Porod, J. Russer, and H. Yordanov, "Nanoelectronics-based integrated antennas," *IEEE Microwave Magazine*, vol. 11, Dec. 2010.
- [8] H. Rempp, J. Burghartz, C. Harendt, N. Pricopi, M. Pritscow, C. Reuter, H. Richter, I. Schindler, and M. Zimmermann, "Ultra-thin chips on foil for flexible electronics," in *Proc. of IEEE Intl. Solid-State Circ. Conf., 2008*, San Francisco, CA, Feb. 2008, pp. 334–617.
- [9] H. Richter, H. Rempp, M.-U. Hassan, C. Harendt, N. Wacker, M. Zimmermann, and J. Burghartz, "Technology and design aspects of ultra-thin silicon chips for bendable electronics," in *Proc. of IEEE Intl. Conf. on IC Design and Technology, 2009*, Austin, TX, May 2009, pp. 149–154.
- [10] J. Burghartz, W. Appel, C. Harendt, H. Rempp, H. Richter, and M. Zimmermann, "Ultra-thin chips and related applications, a new paradigm in silicon technology," in *Proc. of ESSCIRC, 2009*, Athens, Nov. 2009, pp. 28–35.
- [11] P. Russer, *Electromagnetics, Microwave Circuit and Antenna Design for Communications Engineering*, 2nd ed. Nordwood, MA: Artec House, 2006.
- [12] S. B. Cohn, "Slot line on a dielectric substrate," *IEEE Trans. Microw. Theory Tech.*, vol. 17, pp. 768–778, Oct. 1969.
- [13] T. O. Dickson, M.-A. LaCroix, S. Boret, D. Gloria, R. Beerkens, and S. P. Voinigescu, "30-100-GHz inductors and transformers for millimeter-wave (Bi)CMOS integrated circuits," *IEEE Trans. Microw. Theory Tech.*, vol. 53, pp. 123–133, Jan. 2005.
- [14] D. M. Pozar, *Microwave Engineering*, 3rd ed. Ney York, NY: John Wiley & Sons, 2005.

# SCP-RPSC - The New Technology for Microwave Broadband Mobile Communications

Veselin Demirev<sup>1</sup>

**Abstract** – A retrospective review of a new technology, named SCP-RPSC, is given in this report. It is particular useful as the radio front end of the mobile microwave broadband communication systems with terrestrial and satellite positioning. A list of possible applications is given too.

**Keywords** – SCP, RPSC, RPSC-MA, broadband microwave mobile communications.

## I. INTRODUCTION

The terrestrial and satellite broadband mobile communications are currently a strong growth market, driven chiefly by major projects to deploy vast regional or worldwide networks. They need new and wider frequency bands, available at higher frequencies – up to millimetre wave frequency bands. One of the biggest technical problems of these mobile networks is the way of access to the satellite or terrestrial base stations, particular the used antenna systems. The need to change the polarization, to track Low Earth Orbiting Satellites (LEO,s) or High Altitude Platforms (HAPS), to select one of several Geo Stationary Orbit Satellites (GEO,s) positions, as well as the requirements for two way broadband mobile communications at low price and mass market production leads to unsolved by traditional antennas problems. Their solution needs entirely new approach, which is subject of the last decade research activity of the author. The name of the new technical solution is Spatial Correlation Processing – Random Phase Spread Coding (SCP-RPSC). A retrospective review of the research step by step approach, used by the author, is given in the report with the main bibliography for details.

## II. SCP TECHNOLOGY

The main objectives of the SCP technology [1, 3, 5, 6, 7, 10, 20, 21] are:

- To receive one or more radio signals coming from one or several spatially distributed sources (satellites, base stations), insuring high gain of the antenna systems and using fixed or mobile receiving terminals, equipped with SCP signal processing systems.
- To ensure spatial selectivity high enough to cancel the same frequency channel interference, coming from

different space directions, using simple one-channel receiver and signal processing techniques.

The objectives stated above are achieved by a patented method for radio communications, which proposes application of additional pilot signal transmitted in the band of information signals and available in the receiver by one of the known methods for access. The SCP receiver terminal is equipped with antenna array with random phase aperture excitation. The phase shifts among the signals, coming from the antenna elements, are random at the antenna output, regardless of the information source direction. These random phase spread signals correlate with the recovered pilot signal, phase spread in the same manner, in a signal recovery unit (Fig.1). The result of the correlation process between pilot and information signals is the recovered information signal at base band.

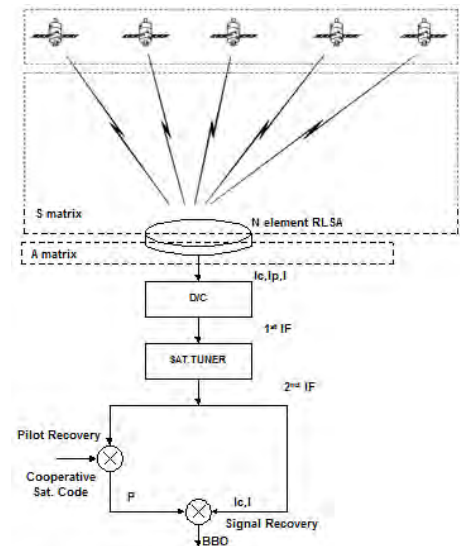


Fig.1. Block scheme of a SCP-CDMA system

One of the main parts of the SCP system is the random phased antenna. In principle all kind of antenna arrays could be used, but for Ku and Ka bands particular suitable is the Radial Line Slot Antenna (RLSA). Until now it is used as phased array for fixed satellite reception.

The main features of the SCP approach are:

- Simple, cheap and flat passive radial line antenna, suitable for mass production in Ku and Ka frequency bands.
- One channel convenient microwave receiver with simple signal processing.
- Omni directional for the cooperative satellite or terrestrial base station, but with high figure of merit G/T.
- Selection of different satellites and polarizations by PN-codes.

<sup>1</sup>Veselin Demirev is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: demirev\_v@tu-sofia.bg.

- Applications in existing S-DVB systems with minor modifications of the ground transmitters, compatible with the existing satellite transponders.
- Multi-beam and soft handover features.

### III. RPSC TECHNOLOGY

The idea to use SCP principle in transmit mode [2, 9, 15, 16] was born during the SCP project research. The transmitting antennas, as well as the receiving random phase antenna arrays in SCP technology are pure passive, without any active or nonreciprocal elements. The specific SCP processing is situated in the receiver (Fig.2). According to the basic electromagnetic antenna laws the replacement of the passive transmitting antenna with passive random phase antenna array in the transmitter, and vice versa in the receiver should not change the system working principles and system parameters. The transmitted by the random phase antenna array signals have specific phase spread. It can be considered as random spatial coding.

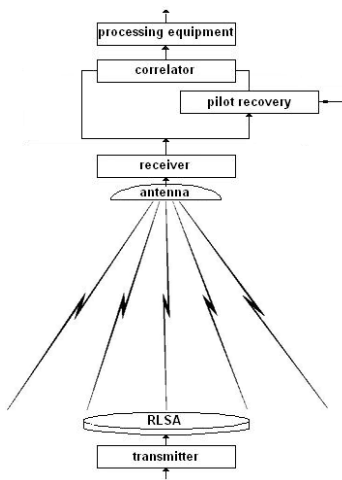


Fig.2. Block scheme of a RPSC system

The main features of the RPSC approach are:

- Providing of full duplex radiocommunication system with one simple and cheap transmit-receive antenna.
- The transmitted random poly-phase spread signals will not cause significant harmful interference to the conventional satellites, using the same frequency channels. The interference will be similar to that, caused by the sidelobes of an antenna array with random elements distribution.
- The transmitted random poly-phase spread signals are uniformly radiated in the space above the antenna. Several satellites, equipped with the same SCP receivers and providing space diversity, receive them. The knowledge of the receiving satellites positions for the transmitting equipment is not necessary (as it is for a conventional satellite earth station).
- The transmitted random poly-phase spread signals have low detection probability for the conventional

microwave receivers, leading to low active jamming probability.

- The SCP-RPSC approach could be a breakthrough technology, leading to unpredictable increase of the frequency reuse factor in satellite and terrestrial wideband networks. Close situated subscriber terminals could communicate with base stations, using the same frequency channel without interference. The isolation between the terminals will be provided by their specific random phase spread coding.

### IV. SCP-RPSC APPLICATIONS

#### A. Satellite Digital Video Broadcasting (DVB-S)

Proposal for a SCP-CDMA GEO satellite system, suitable for DVB-S communications in Ku-band for fixed and mobile terminals, is given in [4]. The proposed algorithm for system parameters evaluation, based on link budget calculations, gives good results – Figure of merit (G/T) better than 14 dBi/K for 60 cm antenna diameter at very low prize (in order of several \$). Similar proposal for quasi GEO satellites at elliptical polar orbits is given in [8].

#### B. Space Links

##### • Inter Satellite Links (ISL)

The space segment of the future global satellite systems for broadband communications can be designed in number of ways, depending on the orbital type of the satellites and the payload technology available on board. The use of different satellite orbits to provide complementary services, each optimized for the particular orbital type, is certainly feasible. Satellites can be used to connect with each other and the ground networks, through the use of Feeder Lines, Inter – Satellite Links or Inter-Orbit Links, which when combined with on-board routing facilities, can be used to form a network in the sky. The unique properties of the SCP-RPSC approach will give a new support for the future broadband LEO,s communication systems in the service feeder lines, inter-satellite and inter-orbit lines domain. The possible applications of the SCP-RPSC technology in these microwave lines of several different types LEO,s constellations are considered in the report [18]. A review of the possible advantages, supported by a critical analysis, is given too.

##### • Feeder Lines

The company O3b, supported by Google, is building a new network of MEO satellites with steerable Ka-band beams to provide lower-cost, fiber-grade access for cellular backhaul and IP backbone trunking in traditionally underserved areas. The main O3b network parameters and architecture are presented in the report [32]. Information concerning the satellite constellation and orbit, satellite and ground antenna beam steering, as well as inter-satellite handover procedures are given. The possible applications of

SCP-RPSC technology in O3b MEO satellite system are discussed too.

- *SPS Lines*

One of the most important use of satellite technologies in the future will be in Solar Power Satellites [SPS]. The concept of generating solar power in space for wireless transmission to receivers on the ground has been discussed in details during the last four decades. All of the sophisticated SPS systems will need broadband wireless communications for telemetry and control purposes among the different parts of their architectures. Another important problem of the future SPS systems will be the transmission of video and telemetric information among SPS mounting robots and satellite or ground based control centers. The unique properties of the SCP-RPSC approach will give a new support for the future SPS mobile broadband communication systems [17].

### C. High Altitude Platform Systems

A new radio technology to realize the last mile access to the broadband fixed networks, named High Altitude Platform Systems (HAPS), is discussed in the reports [14, 24, 29]. Such a mode of service delivery offers advantages as coverage can be rapidly set-up over any location and can be just as easily removed or relocated; high elevation angles can be achieved to the mobile users; efficient frequency re-use schemes can be employed to maximize network capacity; the round-trip delay is relatively short; the cost is considerably less than terrestrial or satellite counterparts. The goal of the reports [14, 24, 29] is to discuss the possibilities and the advantages of the implementation of SCP-RPSC technology in HAPS communications, particularly as subscriber terminal front end equipment. The proposals deal with Line of Sight (LOS) mm-wave propagation environment, which is accepted by the communication community as the only way to communicate in these frequency bands. However, in high building city environment most of the terminal links will be shadowed, which will need more and more new base stations. A possible solution is the Non-LOS mm-wave systems, working properly in high building city environment. The possibilities of SCP technology to create simultaneous several narrow virtual antenna beams could be a good solution of the problem, leading to effective use of the reflected beams by gathering the signals in phase at baseband [31].

### D. WIMAX

The goal of the reports [19, 25] is to discuss the possibilities and the advantages of the implementation of SCP-RPSC technology in *Wi-MAX* communications. The implementation of this technology in subscriber terminals is discussed first. After that the possible base station applications are treated too. The applications of SCP-RPSC technology simultaneous at base station and terminal stations are possible, but they will need additional research and investigations.

### E. Military applications

A review of the mobile satellite communication systems with military applications is given in report [11]. The profits of SCP-RPSC technology for such kind systems are listed and analyzed too.

### F. Telemedicine

The benefits for telemedicine systems, using SCP-RPSC mobile satellite communications, are given in the report [12].

### G. Aeronautical and Global Navigation Satellite Systems (GNSS)

The benefits for aeronautical systems and GNSS, using SCP-RPSC mobile satellite communications, are given in the reports [22, 23, 30].

## V. IMPROVEMENTS OF REGULATORY STATUS OF SATELLITE SERVICE USING VEHICLE-MOUNTED ANTENNAS AND RPSC MULTIPLE ACCESS TECHNIQUES

Satellite connectivity while driving traditionally has been possible by using handheld personal terminal equipment with low gain omni directional antennas. Recently, the new satellite interactive broadband communication systems use high gain satellite tracking antennas, installed on vehicles. Vehicle-Mounted Earth Stations (VMES) currently can operate on conventional Ku-band frequencies (14 GHz Uplink, 11-12 GHz Downlink) but only on a secondary basis. This means VMES can not claim interference protection from primary services such as fixed satellite systems and Earth Station on Vessels (ESV). A co-primary allocation of VMES in the conventional Ku-band would be in the public interest, as it would address a growing commercial demand for on the move services. However, a co-primary allocation would also have to be conditioned on strict adherence to interference avoidance mechanism, which in the best way obviously is satisfied by the RPSC technology [26, 28].

The Radio-communication Sector of ITU is now seeking submissions from industry and governments on various technical, regulatory and economic ideas in order to increase the efficient use of satellite orbits and frequencies. The SCP-RPSC could be a breakthrough technology, leading to unpredictable increase of the frequency reuse factor in satellite and terrestrial wideband networks. Close situated subscriber terminals could communicate with terrestrial or satellite base stations, using the same frequency channel without interference. The isolation between the terminals will be provided by their specific random phase spread coding, due to their specific random design. We can consider this principle of operation as a new multiple access approach, named by us Random Phase Spread Coding - Multiple Access (RPSC-MA) [27].

## VI. CONCLUSIONS

The practical implementations of SCP-RPSC principles will drastically change the existing paradigm in the mobile microwave broadband satellite and terrestrial communication business in general. Many of the existing problems of the proposed systems, dealing with frequency and orbital resource sharing, beam pointing, beam shadowing, terrorist jamming etc., will be solved successfully.

## REFERENCES

- [1] V. Demirev, "Method and System for Space Diversity Communications", Patent WO/2003/013022, PCT/BG2002/000016, 3.02.2003, H04B1/07 (2006.01), H04B 7/02 (2006.01), Priority №105671/04.07.2001, BG.
- [2] V. Demirev, "Radiocommunication Method and System with poly-phase spread coding", Patent WO/2003/107552, PCT/BG2003/000027, 27.12.2003, H04B1/69 (2006.01), H04B 7/06 (2006.01), Priority №106819/13.06.2002, BG.
- [3] V. Demirev, V., "SCP technology – the new challenge of broadband satellite communications", ICEST'04, Conference Proceedings, pp. 159-162, Bitola, Macedonia, 2004.
- [4] V. Demirev, A. Efremov "SCP-CDMA GSO's system proposal", ICEST'04, Conference Proceedings, pp. 163-166, Bitola, Macedonia, 2004.
- [5] V. Demirev, "The probability theory with application in SCP technology", ICEST'04, Conference Proceedings, pp. 167-168, Bitola, Macedonia, 2004.
- [6] В. Демирев, В. Витков, С. Каменополски, А. Ефремов, "Изследване на някои параметри на технологията SCP с MATLAB симулации", Телеком,04, Трудове на конференцията, том 1, стр.321 -326, Варна, България, 2004.
- [7] V. Demirev, A. Efremov, "Some Important Parameters of the SCP Technology", ICEST'05, Conference Proceedings, V.2, pp.516-518, Nis, Serbia and Montenegro, 2005.
- [8] R. Markov, V. Demirev, "Application of SCP technology in Quasi-GEO satellite systems", ICEST'05, Conference Proceedings, V.2, pp.519-521, Nis, Serbia and Montenegro, 2005.
- [9] V. Demirev, "Review of SCP-RPSC technology", ICEST'05, Conference Proceedings, V.2, pp.630-633, Nis, Serbia and Montenegro, 2005.
- [10] В. Демирев, Д. Илиева, "Информационен капацитет по Shannon на SCP система, ограничена по топлинен шум", Телеком, 05, Трудове на конференцията, том 1, стр.95-100, Варна, България, 2005.
- [11] В. Демирев, "Спътниковите комуникации – тенденции и приложения", СЮ Journal, 11, стр.34-38, 2006.
- [12] V. Demirev, "Broadband satellite SCP-RPSC Communications – the new chance for the telemedicine", CEMA, 07, Conference Proceedings, pp. 17-21, Sofia, Bulgaria, 2006.
- [13] В. Демирев, А. Ефремов, Е. Михайлова, "Информационен капацитет по Shannon на ограничена по интерференция SCP система", Телеком,06, Трудове на конференцията, стр.272-277, Варна, България, 2006.
- [14] В. Демирев, В., М. Николова, "Приложение на технологията SCP-RPSC в широколентови HAPS", Телеком,06, Трудове на конференцията, стр.283-288, Варна, България, 2006.
- [15] В. Демирев, "Приложение на вероятностната теория в технологията SCP-RPSC", Телеком,06, Трудове на конференцията, стр.278-282, Варна, България, 2006.
- [16] V. Demirev, "SCP-RPSC technology – the new challenge in the broadband satellite communications", Caspian Telecoms,07, Conference Proceedings, pp.178-181, Istanbul, Turkey, 2007.
- [17] V. Demirev, "Application of SCP-RPSC Mobile Communications in SPS Technology", ISRSSP,07, Conference Proceedings, pp. 129-132, Sofia, Bulgaria, 2007.
- [18] V. Demirev, "SCP-RPSC Technology in the Feeder Lines of the LEO,s Communication Systems", CEMA,07, Conference Proceedings, pp. 1-5, Sofia, Bulgaria, 2007.
- [19] В. Демирев, "Приложение на технологията SCP-RPSC в микровълновите Wi-Max системи", Телеком, 07, Трудове на конференцията, 29, Варна, България, 2007.
- [20] V. Demirev, "Review of SCP Test Set-up and Results – I", CEMA, 08, Conference Proceedings, pp.7-10, Athens, Greece, 2008.
- [21] V. Demirev, "Review of SCP Test Set-up and Results – II", CEMA, 08, Conference Proceedings, pp. 11-13, Athens, Greece, 2008.
- [22] В. Демирев, "Приложение на технологията SCP-RPSC в широколентовата мобилна спътникова система MOWGLY", Телеком, 08, Трудове на конференцията, стр.121-126, Варна, България, 2008.
- [23] В. Демирев, "Приложение на технологията SCP-RPSC в широколентовите спътникови комуникации за въздушния транспорт", Бултранс,10, Трудове на конференцията, стр. 34-37, Созопол, България, 2010.
- [24] В. Демирев, "Основни характеристики на микровълновите HAPS системи, базирани на технологията SCP-RPSC", Телеком, 10, Трудове на конференцията, стр.172-179, София, България, 2010.
- [25] V. Demirev, "SCP-RPSC WI-MAX systems – the new approach for the next generation microwave access strategies", Caspian Telecoms,08, Conference Proceedings, Istanbul, Turkey, 2008.
- [26] V. Demirev, "SCP-RPSC Technology – The Perfect Solution of the Broadband MSS Problems", Computer and Communication Engineering, vol.4, N1, pp.43-46, 2010.
- [27] V. Demirev, "Random Phase Spread Coding Multiple Access - the New competitor of CDMA in the broadband wireless networks", Journal of Applied Electromagnetism, vol.13, Number 1(June 2011), pp. 26-32, 2011.
- [28] V. Demirev, "The regulatory aspects of SCP-RPSC technology – could they solve the VMES problems", Journal of Applied Electromagnetism, vol.13, Number 1(June 2011), pp. 33-38, 2011.
- [29] В. Демирев, "Широколентови радиокомуникационни системи, базирани на летателни апарати", Бултранс,2011, Трудове на конференцията, стр. 43-46, Созопол, България, 2011.
- [30] В. Демирев, "Приложение на технологията SCP в системите за спътникова навигация", Бултранс,2011, Трудове на конференцията, стр. 47-50, Созопол, България, 2011.
- [31] V. Demirev, "SCP Rake Receiver - The Possible Solution of N-LOS HAPS and WIMAX mm-Wave Networks", CEMA,11, Conference Proceedings, pp. 18-21, Sofia, Bulgaria, 2011.
- [32] В. Демирев, "Приложение на технологията SCP-RPSC в спътниковата MEO система O 3 b на Google", Телеком,11, Трудове на конференцията, София, 2011.

# Study on Hybrid FSO/RF Systems Availability Depending on the Meteorological Conditions

Tsvetan Mitsev<sup>1</sup>, Maxim Shupak<sup>2</sup> and Boncho Bonev<sup>3</sup>

**Abstract** – The investigation for increasing the reliability of operation of an FSO communication system through its integration into a hybrid with a wireless radio frequency system is done. By using numerical experiments, the systems' capabilities are compared, depending on the meteorological conditions. An algorithm for switching the constitutive systems, depending on the specific meteorological condition is presented. It is based on the received power of each subsystem (FSO and RF) of the hybrid FSO/RF communication system.

**Keywords** – Free space optics, FSO, radio frequency, hybrid systems, telecommunications.

## I. INTRODUCTION

In the last decade the Free Space Optics (FSO) communication systems became alternative of the fiber optic systems. In these years the technological progress leads to their serious development that allows their wider usage. The four channel FSO system works successfully on distance of more than 28 km with data rate 2.5 Gbps per channel or 10 Gbps overall [1].

In USA, more than 90 % of business costumers are situated on about one mile distance from some high-speed fiber-optic network, but only 19% have access to it. That is mainly caused by high cost of new networks installation. This gives to FSO systems very big potential market. Their wider usage in telecommunications, for example by Internet providers, is limited by relatively low and dependent on distance link availability. That low availability is connected with comparatively rapid and casual changes of atmospheric transparency that are due to unpredictable changes of meteorological conditions during the time. If it is found a way to increase link reliability of FSO from typical 97-98 % to at least 99.9 %, the problem will be solved.

Besides FSO systems, that operate in visible and near infrared diapason, the wireless communication systems use millimeter radio waves also. A practical solution for increasing link availability of FSO is its integration whit a radio frequency (RF) communication system in a hybrid FSO/RF system. The radio subsystem has lower data rate, but has higher stability against bad meteorological conditions like thick fog.

<sup>1</sup>Tsvetan Mitsev is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: mitzev@tu-sofia.bg.

<sup>2</sup>Maxim Shupak is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria.

<sup>3</sup>Boncho Bonev is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: bbonev@tu-sofia.bg.

This paper presented analytical expressions and results for estimation of the power budget of systems, operating in optical wavelength and millimeter frequency ranges. The potentialities of these systems are compared by using a numerical experiment. An algorithm for switching of FSO and RF subsystems depending on meteorological conditions is suggested. Numerical and graphic results of the investigation are presented.

## II. THEORETICAL ANALYSIS

The electromagnetic waves (EM) are absorbed in atmospheric gases, especially in water molecules, carbonic dioxide, oxygen and ozone, when they propagate in Earth's atmosphere. For decrease the influence of this effect the wireless communication systems use frequencies (wavelengths) situated in so-called "transparency windows of the atmosphere".

The EM waves are also absorbed and scattered quite frequency selectively by atmospheric aerosol, anthropogenic or natural, like fog, cloud, smoke and hydrometeors (rain, snow etc.). The attenuation for some wavelengths is so great, that sometimes the connection can interrupt completely even if the link distance is short. When a hybrid communication system is used its link availability has to be higher than of each containing subsystem.

There are some main physical reasons for higher functionality of hybrid FSO/RF communication systems. In frequency range above 10 GHz radio waves are hardly influenced by the hydrometeors and mainly by the rain, because it can be more intensive by the snow. Because rain drops size (0,1 mm to 7 mm) is in the same order as wavelength, the radio wave attenuation in rain is considerable. The link availability of a FSO system is influenced mainly by the fog, what is closely related with visibility  $S_m$  [2, 3]. The fog particles have size of the order of microns and part of microns (typical 1  $\mu\text{m}$  to 20  $\mu\text{m}$ ), i.e. they are commensurable to optical radiation wavelength and therefore the optical wave attenuation in presence of fog is hardly significant. The probability of simultaneously presence of fog and rain is very small, for that reason it can be expected that for a hybrid FSO/RF communication system the disadvantages of its two parts will compensate each other and the overall availability of the system will increase significantly.

The functionality of the communication systems, working in wavelengths of millimeter and centimeter ranges, is determined by some factors as: the effective isotropic radiated power – EIRP, that depends on transmitter power, feeder losses and antenna gain; propagation losses, including free space losses, diffraction losses, attenuation caused by hydrometeors and gaseous molecules; receiver sensibility –

given as minimal power in the input of the receiver that guaranteed fixed value of *BER* (it is often chosen  $10^{-6}$ ).

The atmospheric gaseous absorption losses are well described in the literature and they depend mainly on frequency [4]. The rain attenuation depends on frequency and rain rate and can be obtained by using the methods described in [5]. The scattering losses are due to troposphere's scattering caused by troposphere's density changes and rain scattering.

The power in the input of a radio receiver  $P_r$  [dBm] can be calculated by expression

$$P_r = P_t + G_t - L_t - L_f - L_a - L_r + G_r, \quad (1)$$

where  $P_t$  [dBm] is the transmitter's power;  $G_t$  ( $G_r$ ) [dBi] – transmitter (receiver) antenna gain;  $L_t$  ( $L_r$ ) [dB] – transmitter's (receiver's) tract losses;  $L_f$  [dB] – free space losses;  $L_a$  [dB] – atmospheric losses (in atmospheric gases and hydrometeors) [4, 5].

The antenna gain  $G$  and free space losses can be calculated by relations

$$G = \frac{4\pi\eta A_g}{\lambda^2} \quad (2)$$

and

$$L_f = \left( \frac{4\pi z}{\lambda} \right)^2, \quad (3)$$

where  $\eta$  is the coefficient of antenna area usage, that usually has values from 0,55 to 0,75;  $A_g$  [m<sup>2</sup>] is geometrical area of the antenna aperture;  $\lambda$  [m] is wavelength in free space;  $z$  [m] is distance between the transmitter and the receiver (link distance).

The signal-to-noise ratio  $SNR$  can be calculated by using determined by Eq. (1) value of received power -

$$SNR = P_r - S_r + A \quad [\text{dB}]. \quad (4)$$

We need to know the receiver's sensibility  $S_r$  [dBm] and to choose a proper value of power reserve  $A$  [dB]. The sensibility describes receiver's internal noises for given frequency band (the manufacturer's catalog value can be used). The power reserve is necessary for the cases when unpredictable losses appear. Usually its typical value is 10 to 15 dB.

The link budget for FSO system was described in our previous works [6, 7]. In this work we present only final results what we use in the numerical experiments. For simplification of used mathematical model we assume that: the region of Earth's atmosphere that is part of the FSO communication channel is homogeneous; the laser beam is with Gaussian distribution in transmitter's aperture and the conditions in the atmosphere are suitable for its preservation during the propagation; we introduce a correction for current laser beam radius that takes account of additional extension of the laser beam compared with theoretical; the noise due to the background radiation is neglected [9]; analog modulation of the laser optical beam is preserved in the transmitter.

Therefore the average value of received signal optical power in the input of photo detector is

$$\Phi_{PD} = \frac{1}{2} \Phi_L \tau_t \tau_r \exp \left\{ -z[\text{km}] \frac{3,92}{S_m[\text{km}]} \left( \frac{\lambda[\mu\text{m}]}{0,55} \right)^{-0,5853 \sqrt{S_m[\text{km}]}} \right\} \cdot \frac{1 - \exp \left( -2 \frac{R_r^2}{\rho_z^2(z)} \right)}{1 - \exp(-2)}, \quad (5)$$

where  $\Phi_L$  [W] is the laser power;  $\tau_t$  and  $\tau_r$  are respectively the transmitter's and the receiver's optical systems transparency;  $z$  is the link distance;  $S_m$  is the meteorological visibility of the atmosphere;  $\lambda$  is the wavelength;  $R_r$  [m] is the transmitter's antenna radius;  $\rho_z$  [m] is current Gaussian laser beam radius. When  $\rho_z$  is calculated the initial radius of laser beam  $\rho_0$ , the wavelength  $\lambda$  and the link distance  $z$  are used. The above mentioned correction is also applied.

The signal-to-noise ratio in the output of the receiver can be calculated with taking in account quantum noises, thermal noises and dark current of photo-detector by using the formula

$$SNR = \frac{(\Phi_{PD} \cdot R_D)^2}{2 \cdot e \cdot \Delta f \cdot I_D + 2 \cdot e \cdot \Delta f \cdot I_S + \frac{4 \cdot k_B \cdot T \cdot \Delta f}{R_L}} \quad (6)$$

In eq. (6)  $R_D$  [A/W] is the integral sensibility of photo-detector;  $e$  [C] is the electron's charge;  $\Delta f$  [Hz] is the frequency band of the photo-detector;  $I_D$  [A] is the dark current of the photo-detector;  $I_S$  [A] is the signal current;  $k_B$  [J/K] is the Boltzmann constant;  $T$  [°K] is the temperature of the detector;  $R_L$  [ $\Omega$ ] is the load in the photo-detector's circuit.

In the graphics we use the logarithmical values of signal-to-noise ratio  $SNR$  [dB].

Very important part of the hybrid FSO/RF system functionality is the procedure of subsystems switching and the choice of a parameter and thresholds that will be used in this switching. There are various variants. In dependence on the chosen value of *BER* that is related to the applied modulation technique, we can determine necessary  $SNR$ . Further by using this signal-to-noise ratio we can calculate the minimal value of the received power that guaranteed system's functionality. The easiest way to realize a system, what takes decision for switching of subsystems is to use the received power. In this case there is not a need of additional signal transformations and calculations. Furthermore it is good to use two different threshold levels and the level for switching RF→FSO has to be somewhat greater than the level for FSO→RF switching. In this way we avoid the more frequently subsystems switching. This can cause significant data losses, because of lower data capacity of RF system. After switching, the inactive subsystem has to be in waiting regime.

The "International table for relation between visibility, meteorological conditions and attenuation" [3] can be used for fast estimation of the attenuation in one hybrid FSO/RF communication system in different weather conditions. Further the switching terms can be defined in one model FSO/RF system in dependence on the visibility [8]. The two different thresholds can be defined.



In a hybrid FSO/RF system the only uncontrolled variability in the link budget equation is the communication channel transparency. It depends on concrete meteorological conditions. A possible algorithm for switching and functionality control of the subsystems is by estimation of this variability. Then after comparison with previously given values or with dynamic determined ones, obtained by studying, the algorithm takes decision for switching. This is the approach we choose, and the switching algorithm is given on Fig.1.

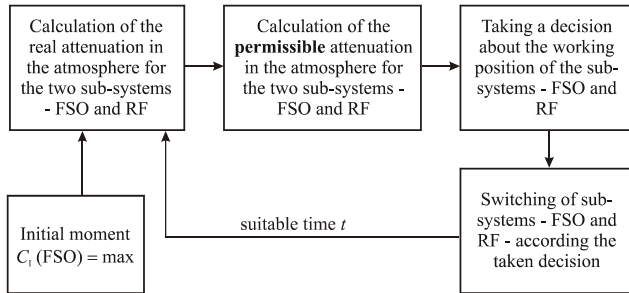


Fig. 1. Subsystems switching algorithm

### III. NUMERICAL EXPERIMENTS

For our numerical experiments we use the parameters' values typical for the real communication system. We fix a link distance  $z = 1,5$  km because the typical hybrid communication system operates on distances in order of 1 km when the rain rate is from 2 mm/h to 100 mm/h or the meteorological visibility is from 0,1 km to 20 km. When we study the SNR dependence on the link distance  $z$ , we assume  $z \leq 10$  km in case of clear atmosphere ( $S_m = 20$  km).

The other parameters are as follow: for the RF subsystem we choose  $f = 2.4$  GHz, with  $P_t = 5$  dBm,  $G_r = G_t = 15$  dBi,  $L_r = L_t = 3$  dB,  $L_f = 103,566$  dB,  $S_r = -90$  dBm and  $f = 60$  GHz, with  $P_t = 10$  dBm,  $G_r = G_t = 41$  dBi,  $L_r = L_t = 3$  dB,  $L_f = 131,525$  dB,  $S_r = -85$  dBm; for FSO subsystem -  $\lambda = 532$  nm, with  $\Phi_L = 0,5$  mW,  $\tau_t = 0,7$ ,  $\tau_r = 0,9$ ,  $R_r = 8$  cm,  $\rho_0 = 2$  cm,  $\theta = 2$  mrad,  $R_d = 0,24$  A/W,  $R_L = 5$  k $\Omega$ ,  $\Delta f = 1$  GHz,  $I_d = 1.5$  nA and  $\lambda = 850$  nm, with  $\Phi_L = 0,5$  mW,  $\tau_t = 0,7$ ,  $\tau_r = 0,9$ ,  $R_r = 8$  cm,  $\rho_0 = 2$  cm,  $\theta = 2$  mrad,  $R_d = 0,5$  A/W,  $R_L = 5$  k $\Omega$ ,  $\Delta f = 1$  GHz,  $I_d = 2$  nA. The frequencies for RF are chosen because they aren't object of license.

Calculated values of SNR for two subsystems of the hybrid FSO/RF system are shown on Fig. 2, 3 and 4.

Fig.2 shows that even very weak rain influences seriously 60 GHz RF subsystem functionality. In other way the FSO subsystem keeps its functionality even in heavy rain conditions for two investigated wavelengths.

The drawings in Fig. 3 show that in fog conditions the FSO subsystem performance decreases significantly without dependence on wavelength even for visibility  $S_m$  in order of 1 to 2 km. In other hand the RF subsystem will work stable even when the visibility  $S_m$  is about 30 m, although in this case SNR is relatively low.

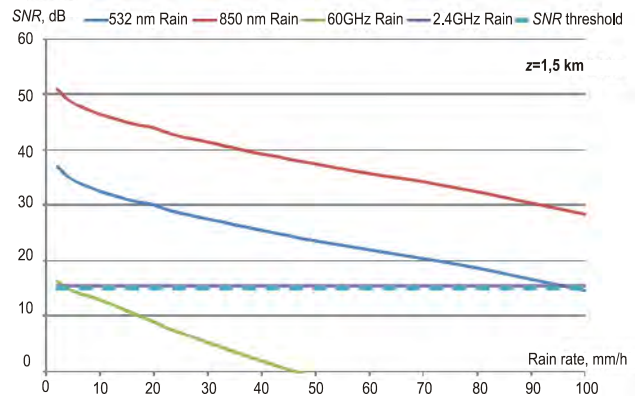


Fig. 2. Hybrid FRO/RF system SNR dependence on the rain rate

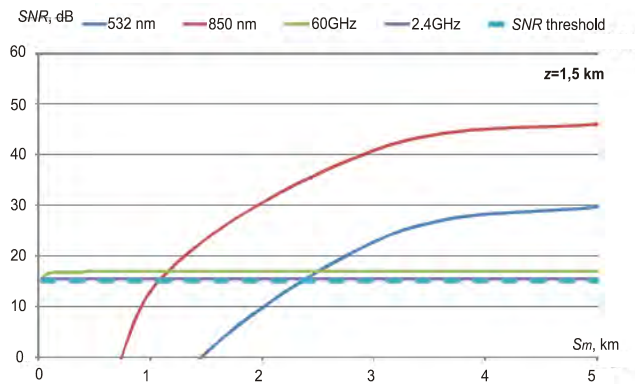


Fig. 3. Hybrid FRO/RF system SNR dependence on the visibility

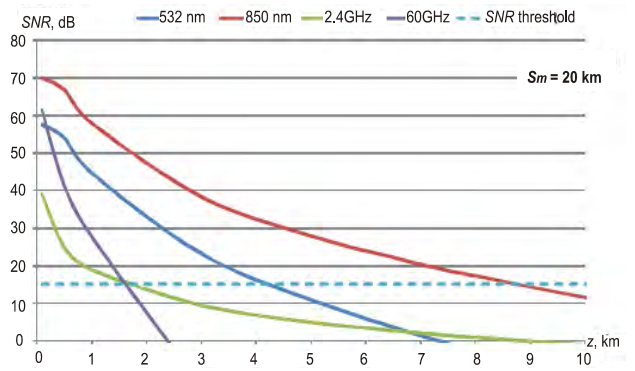


Fig. 4. Hybrid FRO/RF system SNR dependence on the link distance

Fig.4 shows that when the atmosphere is clear ( $S_m = 20$  km), the FSO subsystem has serious advantage over the RF one even for link distances from 4 km to 9 km.

### IV. RESULTS ANALYSIS AND CONCLUSIONS

From applied drawings can be viewed that a heavy rain influences the FSO subsystem functionality much lower than the RF one. As expected the results show increase of optical power through the photo detector's aperture even in high

values of rain rate. When RF subsystem uses frequency of 2.4 GHz the *SNR* is not affected both by the rain and fog but despite this it stands relatively low. It varies significantly only with increases of distance. The calculations give us a possibility to assess that in frequencies of 60 GHz the rain attenuation is considerable, compared to other frequencies, and it doesn't allow high speed communication in long distances. However the fog attenuation in this frequency is relatively low even when the visibility is very low. At the same time, despite relatively low fog attenuation for wavelength 850 nm according to 532 nm, we can't speak for independently working FSO system at these conditions. The very high attenuation of optical waves in fog then the visibility  $S_m < 0,5$  km, is obvious.

For distances up to 2 km RF system works satisfactorily with data transfer, but for communications on distances from 2 to 10 km, FSO systems are more suitable. We can remark that they show better functionality in high visibility conditions. Therefore a hybrid FSO/RF system will rely more on the FSO subsystem. The determination of concrete intervals for work of each subsystem is object of further investigation. The desired system configuration depends on the meteorological conditions and the wanted availability.

It is obvious that a hybrid FSO/RF system contains parts that compensate their disadvantages each other. Therefore the overall availability time for a hybrid system will be significantly greater than for each containing subsystem one by one.

In our work we have chosen to use a *SNR* as criterion for link quality. This is motivated with our desire to avoid connection with used modulation technique. Nowadays when digital systems are used, the more attractive quantity is *BER*, what allows development of this investigation. The theoretical model can be also developed with taking into account some specific effects and phenomena like type of fog – continental or seaside for example. It is also possible to be studied a hybrid system by using real statistical, meteorological data and then to be formulated more accurately the interruption intervals and link availability for the concrete geographical region.

## ACKNOWLEDGEMENT

This work was supported under *Project Nr DUNK-01/03 – 12.09 “University Scientific and Research Complex for Innovation and Transfer of Knowledge in Areas of Micro/Nano Technologies and Materials, Power Effectiveness and Virtual Engineering”*.

## REFERENCES

- [1] Johnston D., “Scientists complete laser link between Lab, Mount Diablo,” Newline, Lawrence Livermore National Laboratory, 2002.
- [2] Kolka Z., Wilfert O., Biolkova V., Reliability of Digital FSO Links in Europe, *International Journal of Electronics, Circuits and Systems*, Volume 1, Number 4, p.236-240, 2007.
- [3] Willebrand H., Baksheesh G., *Free-Space Optics: Enabling Optical Connectivity in Today's Networks*, Sams Publishing, Indianapolis, 2002.
- [4] ITU-R Recommendation P.676-7 (02/07) Attenuation by atmospheric gases
- [5] ITU-R Recommendation P.838-3 (03/05) Specific attenuation model for rain for use in prediction methods
- [6] Mitsev Ts., K. Dimitrov, B. Bonev, Influence of Laser Beam Divergency on Free Space Optic Systems Functionality, *TELECOM 2008*, 9-10 October, St. Constantine, Varna, Bulgaria, 2008.
- [7] Mitsev Ts., K. Dimitrov, N. Kolev, Reliability Testing of Free Space Optical Systems in Laboratory Conditions, *XLIV Intern. Scientific Conf. ICEST 2009*, 26 to 27 June, Veliko Tarnovo, Bulgaria, 2009.
- [8] Akbulut A., Ilk H.G., Arı F., Design, Availability and Reliability Analysis on an Experimental Outdoor FSO/RF Communication System, *Ankara University, Faculty of Engineering*, 2005.
- [9] Bonev B., Relative Influence of Some Stochastic Factors on Bit-Error Rate of Ground-to-Ground Free Space Optics, *XLII International Scientific Conference on Information, Communication and Energy Systems and Technologies, ICEST 2007*, Vol. 1, pp. 203 – 206, June 2007, Ohrid, Macedonia.

# Cylindrical Mesh TLM Model of Probe-Coupled Cavity Loaded with Planparallel Dielectric Layers

Tijana Dimitrijevic<sup>1</sup>, Jugoslav Jokovic<sup>1</sup> and Bratislav Milovanovic<sup>1</sup>

**Abstract** – This paper describes an experimental verification of a compact wire model implemented into the 3-D TLM cylindrical mesh for the purpose of an efficient analysis of a probe-coupled cylindrical microwave cavity loaded with planparallel dielectric layers. An implementation of the compact wire model into a cylindrical TLM mesh is based on wire structures parameters calculation in conditions of variable cross-section of the TLM nodes through which a wire conductor passes due to the nature of a cylindrical grid along the wire path. Results reached by the TLM based approach for characterization of a cylindrical metallic cavity loaded with a planparallel dielectric of the permittivity corresponding to the water have been compared with the measured results.

**Keywords** – Cavity resonators, Electromagnetic analysis, Dielectric layer, Probe antennas, Wire model, Cylindrical grid.

## I. INTRODUCTION

Extensive use of microwave energy in communication, industry, science and medicine has led to the development of a number of different microwave devices based on microwave metallic cavities [1-3]. Among them, the most popular ones are resonant applicators classified as either single or multimode cylindrical metallic cavities, partially loaded with dielectric slabs, widely used in the processes of material heating and drying. They come in various shapes and sizes based on the electromagnetic (EM) properties, geometry and volume of dielectric materials. The knowledge of the mode tuning behavior in a cavity under loading condition (i.e. physical and electrical parameters of the load) forms an integral part of the studies in microwave heating and can significantly help in designing these applicators.

Electromagnetically-based numerical TLM (*Transmission-Line Matrix*) time-domain method can be successfully used to investigate an influence of different EM and geometric parameters of dielectric materials used as a load in microwave cavity applicators on cavity's resonant frequencies [4-7]. Desired mode distribution in the modelled cavity can be established exciting a particular field component through an impulse excitation. However, this way of setting up the wanted TE or TM mode obviously differs from the experimental procedure [7] where a probe, placed inside a cavity, is used as an excitation. Consequently, numerical results obtained in case of an impulse excitation would be different from the experimental ones, in terms of resonant frequency values and an EM field level. Further, since the

reflection and transmission characteristics are common parameters in the cavity exploration, input and output ports of real microwave cavity devices are generally realized by coaxial probe that ensures coupling with corresponding electromagnetic (EM) field component [1]. For that reason, physical and electrical probe parameters form an integral part of microwave technique studies regarding EMC (electromagnetic compatibility) problem of a probe-coupled cavity.

The TLM enhancement in form of the compact model for wire structures has been developed [8, 9], yielding a significant improvement in the required computer resources compared to the traditional TLM method. The so-called TLM wire node has been implemented into the uniform TLM mesh based on a rectangular grid, as mean cross-section dimensions of nodes through which a wire conductor runs are always constant, allowing to easily preserve distributed capacitance and inductance of a wire per unit length. However, if a rectangular uniform mesh is used to model a cylindrical structure [10, 11], a curved boundary would have to be described in a step-wise fashion which might result in a deviation of resonant frequency values as well as in excitation of unwanted modes. A numerical error could be reduced by applying the TLM mesh of a higher resolution, which would result in increasing the simulation time. Moreover, a mesh resolution increasing is limited since an implementation of the compact wire model into the rest of the TLM mesh is based on arbitrary ratio between radius of a wire and dimensions of nodes through which a wire conductor passes. Consequently, a higher resolution of an applied rectangular TLM mesh enables a cylindrical cavity to be precisely modelled only if probes of a relatively small radius are used [10, 11].

These limitations of the rectangular TLM mesh for the purpose of the modelling of a probe-coupled cylindrical cavity were overcome by implementation of the TLM wire node into the cylindrical TLM grid. This solution has enabled the precise modelling of cylindrical cavity boundaries independently of a mesh resolution applied. However, in case of probes that are radially placed, mean cross-section dimensions of cylindrical TLM nodes along the wire path are variable from one node to another, leading to different wire network properties between nodes. For that reason, an additional connecting procedure for wire segments belonging to TLM nodes with different cross-sections has been implemented [12].

In this paper, an efficiency of the TLM wire model adapted to cylindrical mesh has been verified on the example of a probe-coupled cylindrical metallic cavity loaded with planparallel dielectric layer placed at the arbitrary height from a cavity bottom. The transmission coefficient based on the coupling of two wire probes inserted into the cavity has been considered numerically and experimentally.

<sup>1</sup>Tijana Dimitrijevic, Jugoslav Jokovic and Bratislav Milovanovic are with the University of Nis, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, Nis 18000, Serbia, E-mail: [tijana, jugoslav, bata]@elfak.ni.ac.rs

II. MODELING PROCEDURE

In the conventional TLM time-domain method, an EM field strength in three dimensions, for a specified mode of oscillation in a 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. An efficient computational algorithm of scattering properties, based on enforcing continuity of the electric and magnetic fields and conservation of charge and magnetic flux [13], is implemented to speed up the simulation process. EM properties of different mediums in the cavity are modelled by using a network of interconnected nodes, a typical structure known as the symmetrical condensed node – SCN [13]. 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 may be incorporated into the TLM network 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  $(\varphi, r, z)$  can be used for the modelling purpose. The coordinate system used and the port designations are shown in Fig. 1. 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.

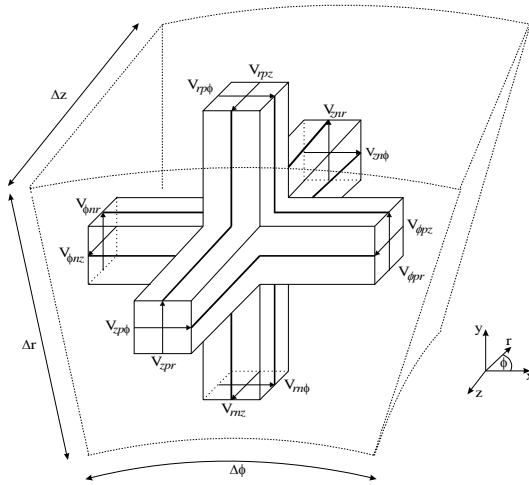


Fig. 1. A cylindrical SCN

The TLM wire node in a cylindrical grid 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. 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 [8].

An equivalent radius of the fictive cylindre in a cylindrical grid for calculating the capacitance and inductance,  $r_{Cr}$  and

$r_{Lr}$ , respectively, for a wire segment running along radial direction are  $r_{Cr} = k_{Cr} \Delta r_c$  and  $r_{Lr} = k_{Lr} \Delta r_c$ , where  $\Delta r_c$  represents a 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  $r_{Cr}$  and  $r_{Lr}$  are factors empirically obtained by using known characteristics of the TLM network.

Distributed capacitance and inductance per unit length, needed for modelling of wire segments, may be expressed as:

$$C_{wr} = \frac{2\pi\epsilon}{\ln(r_{Cr}/r_w)}, \quad L_{wr} = \frac{\mu}{2\pi} \ln(r_{Lr}/r_w) \quad (6)$$

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, for a radial wire conductor in a cylindrical grid, 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. 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. 2). Therefore, an additional connecting procedure for wire segments with different link-lines admittances has been implemented into the existing TLM-based software [12].

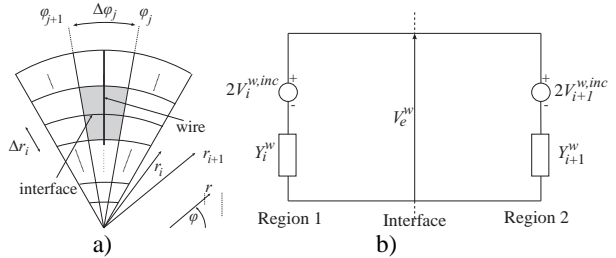


Fig. 2. a) TLM nodes in  $r\varphi$  plane through which wire runs and b) an interface between two nodes

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:

$$V_e^w = 2V_{i+1}^{w,inc} \frac{Y_{i+1}^w}{Y_i^w + Y_{i+1}^w} + 2V_i^{w,inc} \frac{Y_i^w}{Y_i^w + Y_{i+1}^w} \quad (7)$$

$$V_i^{w,ref} = \frac{Y_i^w - Y_{i+1}^w}{Y_i^w + Y_{i+1}^w} (V_i^{w,inc} - V_{i+1}^{w,inc}) + V_{i+1}^{w,inc} \quad (8)$$

$$V_{i+1}^{w,ref} = \frac{Y_i^w - Y_{i+1}^w}{Y_i^w + Y_{i+1}^w} (V_i^{w,inc} - V_{i+1}^{w,inc}) + V_i^{w,inc} \quad (9)$$

where  $V_e^w$  is an equivalent voltage at the interface,  $V_i^{w,inc}$  and  $V_{i+1}^{w,inc}$  are the incident voltages.

Such compact wire model allows for simple incorporation of voltage/current sources and lumped loads and takes into account the physical dimensions of wire probes [9], determined only by TLM mesh resolution.

When modelling of cavities containing lossy loads is concerned, implementation of losses in the TLM model is carried out by introduction of stubs with losses in the nodes where scattering is going on. Stubs with losses may be considered as infinitely long transmission lines, or equivalently, as lines terminated with its characteristic impedance. They can be used to model either electric or magnetic losses. In case of the symmetrical condensed node, stubs with losses are directly implemented in the scattering procedure, including coupling with the corresponding EM field component.

If  $\sigma_{ek}$  and  $\sigma_{mk}$  represent effective electric and magnetic conductance, respectively, in  $k$ -direction, where  $k = (\varphi, r, z)$ , elements in the TLM node used for modelling of losses are defined as:

$$G_{ek} = \sigma_{ek} \frac{\Delta i \Delta j}{\Delta k}, R_{mk} = \sigma_{mk} \frac{\Delta i \Delta j}{\Delta k} \quad (10)$$

where  $(\Delta i, \Delta j, \Delta k) = (r\Delta\varphi, \Delta r, \Delta z)$ .

Starting from  $\varepsilon_k^* = \varepsilon_0 \varepsilon_{rk} - j \frac{\sigma_{ek}}{\omega}$ ,  $\mu_k^* = \mu_{rk} - j \frac{\sigma_{mk}}{\omega}$  [13], it is possible to define a loss tangent at the appropriate frequency as:

$$\tan \delta_{ek} = \frac{\sigma_{ek}}{2\pi f \varepsilon_0 \varepsilon_{rk}}, \tan \delta_{mk} = \frac{\sigma_{mk}}{2\pi f \mu_0 \mu_{rk}} \quad (11)$$

Finally, corresponding equations for reflected total voltages and currents in corresponding direction have to be modified in case of modelling of mediums with losses [13].

### III. RESULTS AND ANALYSES

The proposed TLM model based on the cylindrical grid and enhanced with the TLM wire node has been used for modelling of a loaded probe-coupled cylindrical cavity with dimensions  $a = 7$  cm and  $h = 14.24$  cm, chosen to follow the experimental ones [7]. Wire probes, representing a feed and receiving probe, were placed at the height  $l = 7.4$  cm from the bottom of the cavity, along the radial direction and opposite to each other (Fig. 3). In order to model real coaxial cable characteristics, the probes were connected, through the TLM wire port, to the real voltage source and resistances of  $50\Omega$ . In order to illustrate effect of probe length changing to EM field distribution, analyses have been carried out for different length of probes  $d_1 = d_2 = d = 4$  and  $5$  cm, respectively.

Two types of the loaded cavity have been considered. One represents the cavity containing only one layer of water placed at  $h_1=10$ cm from the bottom of the cavity (Fig.3a), whereas the other contains two water layers, one at the bottom and the other at  $h_1=10$ cm from the cavity bottom, (Fig. 3b). The height of the each layer is  $h_2=3$ cm. The permittivity of the water was taken into account ( $\varepsilon_r = 77 - j6$ ). For modelling purpose a cylindrical TLM grid  $(\varphi \times r) = (36 \times 28)$  was used, whereas in  $z$ -direction dimensions of nodes are different, due to inhomogeneity of the medium inside the cavity. In order to achieve time synchronization of network, the TLM node in dielectric was set to be  $\sqrt{\varepsilon_r}$  times less compared to the nodes dimension in the rest of the cavity filled with air.

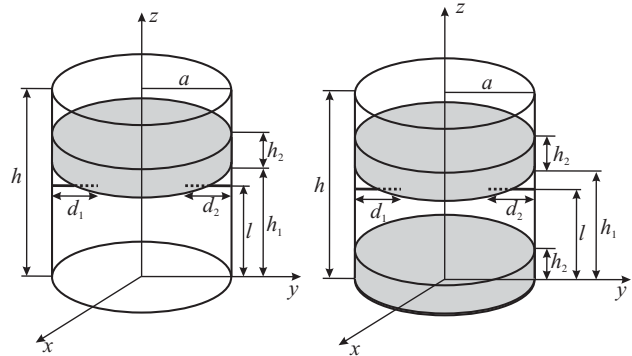


Fig. 3. A probe-coupled loaded cylindrical cavity with a) one water layer and b) two water layers

Obtained numerical results, representing transmission coefficient in the frequency range of interest, have been experimentally verified in both cases of a loaded cavity and for variable lengths of probes. Figs. 4. and 5. present comparative results in terms of varying of probe length and load conditions. Apparently, a very good agreement between numerical and experimental results has been achieved, confirming that cylindrical mesh based TLM model can be used for analyses of probe-coupled cavity loaded with planparallel dielectric layers.

### IV. CONCLUSION

This paper presents an efficiency of the compact wire model implemented into the 3-D TLM cylindrical mesh for the purpose of the analysis of a probe-coupled loaded cylindrical microwave cavity. Due to cylindrical grid structure and empirical nature of the compact model, this implementation has to take into account a change of wire parameters with a variable cross-section of the TLM nodes through which a radially placed wire conductor passes. The model accuracy has been experimentally verified on an example of a probe-coupled cylindrical cavity containing one and two layers of water as a dielectric load, for different probe length values.

Considered configuration of a probe coupled cavity loaded with dielectric layers placed at different heights from the cavity bottom is of a great importance in a realization of microwave resonant applicators, widely used for thermal processing of materials.

### ACKNOWLEDGEMENT

This work was supported by Serbian Ministry of Education and Science within the project TR 32054.

### REFERENCES

- [1] T. V. C. T. Chan, H. C. Reader, *Understanding microwave heating cavities*, Boston, London: Artech House, 2000.
- [2] C. Balanis, *Advanced Engineering Electromagnetics*, New York, John Wiley & Sons, 1989.

- [3] S. Amnartpluk, C. Phongcharoenpanich, S. Kosulvit, and M. Krairiksh, "A power divider using linear electric probes coupling inside conducting cylindrical cavity", *Int. Symp. on Circuits and Systems* 3, 2003, pp. 419-422.
- [4] B. Milovanovic, N. Doncov, and A. Atanaskovic, "Tunnel type microwave applicator analysis using the TLM method", *Proc. of the 4th International Workshop on Computational Electromagnetics in the Time Domain: TLM/FDTD and Related Techniques, CEM-TD 2001*, Nottingham, United Kingdom, 2001, pp.77-84.
- [5] B. Milovanovic and 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*, A Publication of the International Microwave Power Institute, VA, USA, 37, 2002, pp.237-247.
- [6] J. Joković, B. Milovanović, T. Randelović, "TLM Modelling of Microwave Applicator with an Excitation Through The Waveguide", *Microwave and Optical Computer Technology Letters*, John Wiley&Sons, 48, 2006, pp.2320-2326.
- [7] S. Ivkovic, B. Milovanovic, A. Marincic, and N. Doncov, "Theoretical and experimental investigations of resonance frequencies in loaded cylindrical microwave cavity", *Proc. of the 3rd IEEE TELSIS'97 Conference*, Nis, Yugoslavia, 1997, pp.306-309.
- [8] V. Trenkic, *The development and characterization of advanced nodes for TLM method*, Ph.D. Thesis, University of Nottingham, 1995.
- [9] V. Trenkic, A.J. Wlodarczyk, and R. Scaramuzza, "Modelling of coupling between transient electromagnetic field and complex wire structures", *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, 12, 1999, pp.257-273.
- [10] J. Joković, B. Milovanović, and N. Dončov, "TLM analysis of cylindrical metallic cavity excited with a real feed probe", *International Journal of RF and Microwave Computer Aided Engineering*, John Wiley&Sons, USA, 16, 2006, pp.346-354.
- [11] 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, 18, 2008, pp.295-302.
- [12] Bratislav Milovanović, Nebojša Dončov, Jugoslav Joković, Tijana Dimitrijević, "EM Field Monitoring in Circular Cavity Using Wire Compact Model Implemented in Cylindrical TLM Mesh", *Proc. of the TELSIS 2009 Conference*, Niš, Serbia, October 7-9, 2009, pp. 339-342
- [13] C. Christopoulos, *The Transmission-Line Modelling (TLM) Method*, Series on Electromagnetic Wave Theory, IEEE/OUP Press, 1995.

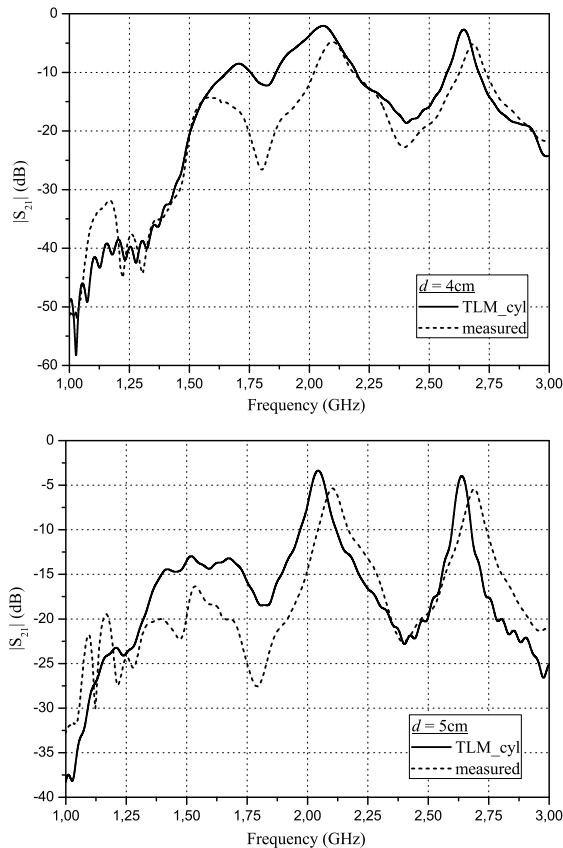


Fig. 4. Transmission coefficient magnitude in the cylindrical cavity loaded with one water layer

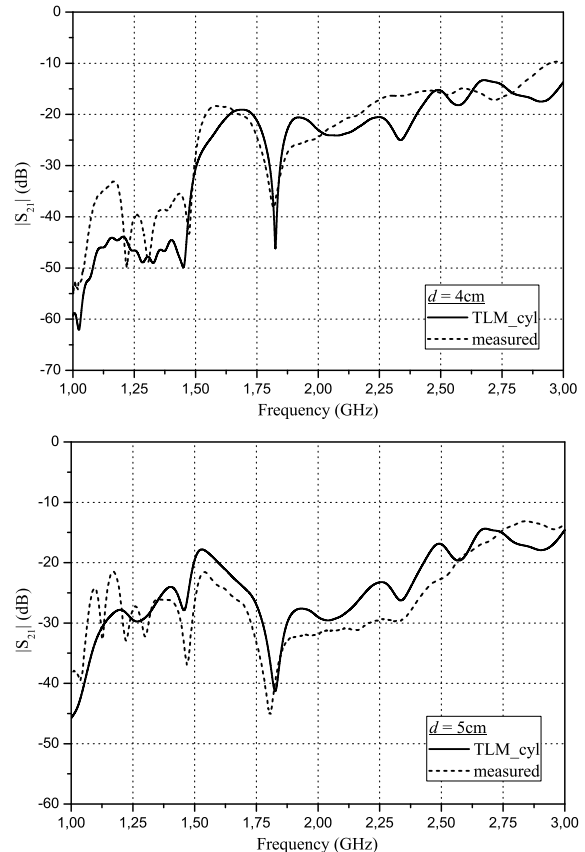


Fig. 5. Transmission coefficient magnitude in the cylindrical cavity loaded with two water layers

# Implementation of pseudo random noise generator in FPGA for Free Space Optics BER testing

Nikolay Kolev<sup>1</sup> and Tzvetan Mitzev<sup>2</sup>

**Abstract** – The way to determine the quality of communication link is to measure its Bit Error Ratio (BER). BER most often is measured by the so called brute force method. It is sending random bits through the system and calculating BER. The devices which measure BER are called Bit Error Rate Testers. They consist of pseudo random noise generator and comparing scheme. This report presents the creation of FPGA based system for testing BER in Free Space Optics Systems. The paper offers block and principle realisation of PRNG algorithm. There is synthesized HDL code for Spartan 6 FPGA chip. Our generator is tested in real conditions with oscilloscope. There is shown computer simulation and practical results.

**Keywords** – Bit Error Rate, Free Space Optics, Pseudo Random Noise Generator, FPGA

## I. INTRODUCTION

One of the most important parameter, for quality assessment in telecommunication systems is BER. It is a measure of the percentage of bits that a system does not transmit or receive correctly. The experiments connected with BER give us the information about the maximal bandwidth and quality of the FSO communication link. The device which measure BER is called Bit Error Rate Tester (BERT). There are a lot of companies that make BERTs such as Tektronix, Agilent, Rohde & Schwarz. Because of price of these commercial systems we decide to build our own BERT. The main block of BERT is the pseudo random noise generator which is explained in this work.

Random number generators are mainly used in telecommunications, cryptographic algorithms and authentication protocols. There are many different ways to generate pseudo random bit sequence. Most popular are ADC method, shift register method and software method [1]. ADC method can be achieved by feeding sinusoidal signal to the ADC, while at the output of the IC there is a random digitized signal. With ADC can be achieved hi speeds, but fast ADC's are expensive. In Software method we are restricted form computer's communication ports. We choose to work with shift register method because of the low price and hi speed of FPGA. In this method we have a couple of D type flip flops and one XOR logic element connected in series. The clock generator moves the logic levels through the scheme. In this generator the output is the last D type flip flop. In software generators we increment fast one or couple registers,

it depends of the case. The random bits or bytes are obtained by reading registers in definite period of time. We decide to work with shift register method because of simple hardware implementation. This technique has good statistical properties and leads to very efficient hardware implementations.

FPGA are reconfigurable silicon chips which can be programmed in various different hardware configurations [2]. In this way we obtain completely different hardware implementations. The specific thing in this IC's is the ability to perform parallel processes unlike the MCU's. Each processing task is assigned to a different section of the chip, and can function autonomously without any impact from other logic blocks. The performance of single part of the application is not affected when we increase the number of processes. The FPGA's have a couple of advantages which are connected with better performance, low cost and good reliability. Due to the parallelism of these chips their computing power is almost equal to the Digital Signal Processor's computing power. The response times of each I/O pin is faster.

In this work FPGA implementation of PRNG is presented. This report is addition of our paper which presents Experimental setup for BER measuring of Free Space Optical System [3]. Due to the needs of increasing the throughput of our system we replace the existing discrete components based generator with FPGA chip. The advantages of new system are: easy reprogramming, high frequencies (500+ MHz), minimizing the physical dimensions.

## II. SCHEMATIC DESIGN

The block scheme of generator is shown in fig.1. The operating principle is explained in the next paragraph.

The principle scheme of the pseudo random noise generator can be seen on fig.2. It possible to consist of several D type flip flops connected in series, one XOR logic gate and one

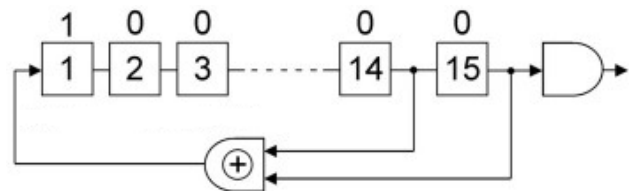


Fig. 1. Block scheme of random number generator

inverter. In shown case there are four D type flip flops which Q outputs are connected to D inputs of next flip flop. The two inputs of the logic gate XOR are connected to the last Q output of the scheme, through inverter, and between two middle flip flops.

<sup>1</sup> Nikolai Kolev is with the Faculty of Telecommunications at Technical University, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: kolev@tu-sofia.bg.

<sup>2</sup> Tsvetan Mitzev is with the Faculty of Telecommunications at Technical University of Sofia. E-mail: mitzev@tu-sofia.bg

The four D type flip flops form 4 stages shift register. At every tick of the clock signal the logic level on every input is shifted one step left. In this situation the output of the XOR gate is a source of pseudo random bit sequence. The function of inverter is to start the generator, because in the beginning at every inputs and outputs there are only zeroes. The XOR gate need different levels to produce ones at his outputs, that's why there is inverter at the one of two inputs of the XOR gate [4].

The period of repetition depends on number of D type flip flops. In particular case it can be calculated by formula 1, where "n" is number of D type flip flops.

$$2^n - 1 \tag{1}$$

In the shown scheme in fig.2 Feedback polynomial is (2), the period of repetition is 15 because  $2^4 - 1 = 15$

$$x^4 + x^3 + 1 \tag{2}$$

For example if the number of bits are 16 then the period of repetition will be 65536, because  $2^{16} - 1 = 65535$ . If the numbers of bits are 19, the period of repetition is 524287 and so on. So we can connect the necessary number of D type flip flops for desired period of repetition.

For implementation of the generation algorithm it is necessary to produce confirmation code for the FPGA chip. This code makes individual connections between the different parts in the chip. For code configuration tool we used Xilinx ISE Web Pack tool which is freely accessed in Xilinx web site [5]. For top layer of our project we used schematic design fig.2.

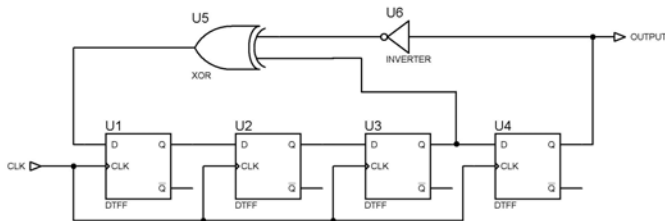


Fig.2. Principle realization of PRNG

The HDL code is too big to be shown in paper, that's why we will show a essential part of it in listing 1. In the beginning of code there is definition of standard libraries like ieee.std, ieee.numeric and so on. After that we have definition of inputs and outputs of the chip. At the next lines there is the configuration of the divider of the clock generator. After that we have defining inputs and outputs of the logic elements D flip flop, inverter and XOR gates of the scheme. At the end of the code there is shown interconnections between the logic elements and I/O pins.

Listing1 HDL Code for Spartan 6 board.

```
library ieee;
```

```
use ieee.std_logic_1164.ALL;
use ieee.numeric_std.ALL;
library UNISIM;
use UNISIM.Vcomponents.ALL;
```

```
entity tesALTYS is
    port ( clk : in  std_logic;
          Led5 : out std_logic;
          Led6 : out std_logic;
          Led7 : out std_logic);
end tesALTYS;
.....
end component;
attribute CLKFXDV_DIVIDE of DCM_CLKGEN :
component is "2";
attribute CLKFX_DIVIDE of DCM_CLKGEN :
component is "1";
attribute CLKFX_MD_MAX of DCM_CLKGEN :
component is "0.000";
attribute CLKFX_MULTIPLY of DCM_CLKGEN :
component is "4";
attribute CLKIN_PERIOD of DCM_CLKGEN :
component is "0.0";
attribute SPREAD_SPECTRUM of
DCM_CLKGEN : component is "NONE";
attribute STARTUP_WAIT of DCM_CLKGEN :
component is "FALSE";
attribute BOX_TYPE of DCM_CLKGEN :
component is "BLACK_BOX";
component FD
    generic( INIT : bit := '0');
    port ( C : in  std_logic;
          D : in  std_logic;
          Q : out std_logic);
end component;
attribute BOX_TYPE of FD : component is
"BLACK_BOX";

component INV
    port ( I : in  std_logic;
          O : out std_logic);
end component;
attribute BOX_TYPE of INV : component is
"BLACK_BOX";

component XOR2
    port ( I0 : in  std_logic;
          I1 : in  std_logic;
          O : out std_logic);
end component;
.....
begin
    Led5 <= Led5_DUMMY;
    Led6 <= Led6_DUMMY;
```



```

Led7 <= Led7_DUMMY;
XLXI_7 : DCM_CLKGEN
.....
.....
XLXI_8 : FD
  port map (C=>XLXN_51,
           D=>XLXN_5,
           Q=>XLXN_4);

XLXI_9 : INV
  port map (I=>XLXN_4,
           O=>XLXN_5);

XLXI_10 : FD
  port map (C=>XLXN_4,
           D=>XLXN_7,
           Q=>XLXN_6);
XLXI_11 : INV
  port map (I=>XLXN_6,
           O=>XLXN_7);

XLXI_12 : FD
  port map (C=>XLXN_6,
           D=>XLXN_10,
           Q=>XLXN_9);

XLXI_13 : INV
  port map (I=>XLXN_9,
           O=>XLXN_10);

XLXI_14 : FD
  port map (C=>XLXN_9,
           D=>XLXN_12,
           Q=>XLXN_11);

XLXI_15 : INV
  port map (I=>XLXN_11,
           O=>XLXN_12);

XLXI_16 : FD
  port map (C=>XLXN_11,
           D=>XLXN_14,
           Q=>XLXN_13);

XLXI_17 : INV
  port map (I=>XLXN_13,
           O=>XLXN_14);

XLXI_18 : FD
  port map (C=>XLXN_13,
           D=>XLXN_16,
           Q=>XLXN_15);
    
```

For our development we used evaluation board Altys Spartan 6 from DIGILENT (fig.3) [6] . It is very powerful board with rich peripheral devices such as LAN interface unit, a couple of switch units and micro switches, LED's, USB, UART, HDMI, external RAM memory. We choose to work with ALTYS because of it high operating frequencies 500+MHz and existence of Gigabit Ethernet.

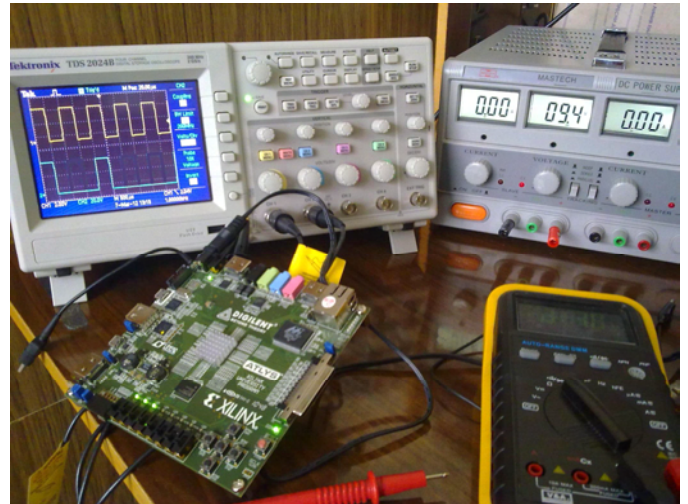


Fig.3 ALTYS Spartan 6 Board

### III. EXPERIMENTAL SETUP

The block scheme for measuring BER in FSO atmosphere channel can be seen at fig.4. It consists of Rx and Tx optical drivers, FSO atmosphere channel, counter device, BER measuring scheme and pseudo random noise generator. The information is passed through atmosphere channels forward and backward. At the end of the backward channel there is BER measuring scheme, which compare signals from the generator and atmosphere channel. If signals are different the scheme formed logical ones. Otherwise the scheme form logic zeroes, which mean no error [3] .

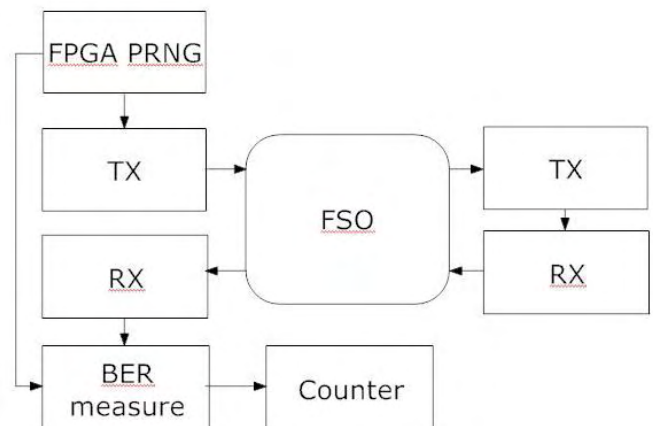


Fig.4 FSO system

We plan to combine our FSO system fig.5 and FPGA based PRNG for measuring BER in real conditions. We want to bind the results for BER with atmospheric effects such as fog, rain, snow. For reading the intensity of meteorological effects we will use meteorological station.



Fig.5. Scheme for measuring BER in atmosphere channel

IV. SIMULATION

Fig.6 and Fig.7 shows computer simulations and real oscillogram. In the upper part of the figures there is a PRNG signal, while in the bottom of the figures there is a clock signal [7].

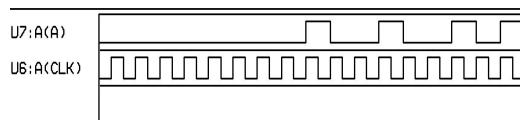


Fig.6. Computer simulation of PRNG signal

V. CONCLUSION

The described algorithm can be used for learning the dependence of the level of bit error in Free Space Optics systems [8]. This system is easy reprogramming, it can generate pseudo random bit sequence at frequency of 500Mhz, it have small physical dimensions. In future we plan to realize all blocks in fig. 3 with FPGA.

Research results presented in this publication are funded by Internal competition of TU-Sofia-2011, contr. Nb112pd033-7.

REFERENCES

- [1] Beker H., P.Fred, The Protection of Communications. Wiley-Interscience. p. 212. 1982.
- [2] Pong P. Chu FPGA Prototyping by VHDL Examples: Xilinx Spartan-3 Version (Feb 4, 2008)
- [3] Kolev N., K. Dimitrov Y. Velchev T. Mitsev Experimental Setup for BER Measuring of Free Space Optical System , pp.214-217, Nis, Serbia , 2011.
- [4] Anil K. Maini Defence Digital Electronics Principles, Devices and Applications Research and Development Organization (DRDO), India
- [5] www.xilinx.com
- [6] <http://www.digilentinc.com/Products/Detail.cfm?NavPath=2,400,836&Prod=ATLYS>
- [7] Hartmann A. K. A practical guide to computer simulations World Scientific Publishing Company; Pap/Cdr edition (March 30, 2009)
- [8] Stamatios V. Kartalopoulos Optical Bit Error Rate: An Estimation Methodology Wiley-IEEE Press 2004

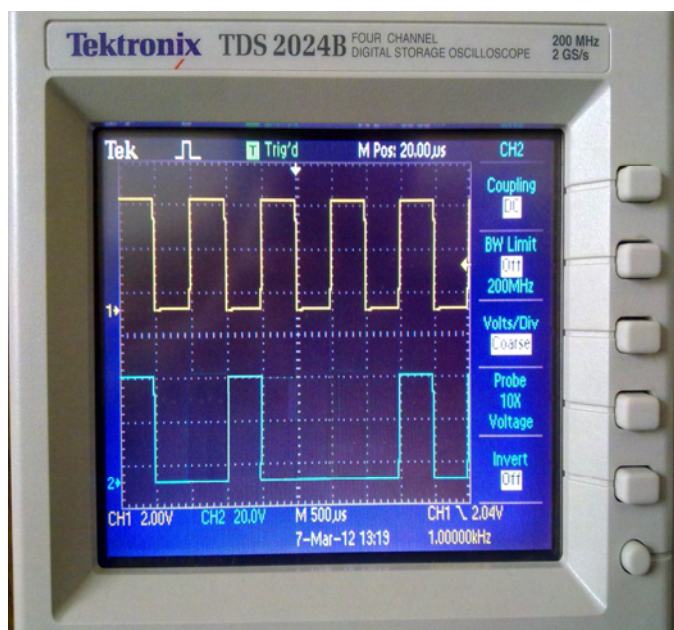


Fig.7. Oscillogram of PRNG signal

# Experimental Estimation and Correction of the Methods for Radio Waves Attenuation Prediction in Rain

Boncho Bonev<sup>1</sup>, Kliment Angelov<sup>2</sup> and Emil Altimirski<sup>3</sup>

**Abstract** – An experimental estimation of main methods for prediction of rain attenuation of radio waves is presented. Data for attenuation values and rain rates from real radio links, worked in Bulgaria are used for the study. Then the real and predicted attenuations are analyzed by using existing methods. Corrections in existed methods for increase their precision for the territory of Bulgaria are suggested.

**Keywords** – Rain attenuation, gaseous attenuation, microwave, prediction methods.

## I. INTRODUCTION

The human race progress sets new challenges to the communication technologies, related mainly with transfer of much more information. The radio links that work in frequency band from 2,4 to 5,6 GHz are not able to respond to these challenges. This requires the usage of higher frequencies in the present-day communication systems. In this case the other problems, like hydrometeors and atmospheric gaseous attenuation exist. The losses caused by atmospheric gases and hydrometeors are the main reason for link distance limitation when frequencies over 10 GHz is used [1]. This influence is significant especially for frequencies above 45-50 GHz when the link distance can reach only 1-2 km with the link availability of 99,99% or greater [1].

This paper presents an analysis of calculated and measured rain attenuation and suggests a correction of coefficients used in rain prediction models [2]. The experimental data for radio wave attenuation are from 4-year measurements over radio link that has worked in Sofia region [3].

## II. THEORETICAL ANALYSIS

The radio waves with frequencies above 10 GHz are significantly absorbed and scattered by hydrometeors like rain, snow, fog, clouds etc. The attenuation when the heavy rain occurs is so great, that sometimes the connection can interrupt completely even if the link distance is short.

The rain attenuation  $L_{rain}$  is caused mainly by rain absorption and is expressed as [4]

$$L_{rain} = L_{sp\_rain} \cdot r \cdot d / 1000, \quad (1)$$

where  $L_{sp\_rain}$  is the specific rain attenuation in dB/km and can be calculated by using the ITU model [4] as follows

$$L_{sp\_rain} = k(f) \cdot RR^{\alpha(f)}. \quad (2)$$

In Eq. (2)  $RR$  is the rain rate in mm/h,  $k(f)$  and  $\alpha(f)$  are the frequency and polarization dependent constants.

In Eq. (1), the distance  $d$  is in meters and  $r$  is a correction, what renders an account of the fact that the rain falls only on part of the link distance and can be calculated by the expression [4, 5]

$$r = (1 + d / 35000 \exp(-0.015RR))^{-1}. \quad (3)$$

In our analysis we use these formulas to calculate theoretical rain attenuation for examined radio link. We also use them to obtain the value of specific rain attenuation based on experimental data. Then by using of the smaller squares method we obtain new coefficients for rain attenuation prediction model.

## III. RESULTS AND ANALYSIS

We examine two radio links working at frequencies 11,15 GHz and 19,15 GHz with vertical polarization on distance of 7,1 km in Sofia region. For our study we use statistical data for rain rate in Sofia region based on 4-year measurement. By using this data we calculate theoretical rain attenuation with Eq. (1) – (3). The values of  $k(f)$  and  $\alpha(f)$  are obtained from [2] and are given in Table I.

TABLE I

Frequency, GHz	$k(f)$	$\alpha(f)$
11,15	0,018396	1,155685
19,15	0,0878735	0,991755

We also use the statistical data of rain attenuation for these radio links and compare and analyze theoretical and experimental rain attenuation – yellow and red drawing on Fig. 1 and Fig. 2.

<sup>1</sup>Boncho Bonev is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: bbonev@tu-sofia.bg.

<sup>2</sup>Kliment Angelov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: kna@tu-sofia.bg.

<sup>3</sup>Emil Altimirski is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria.

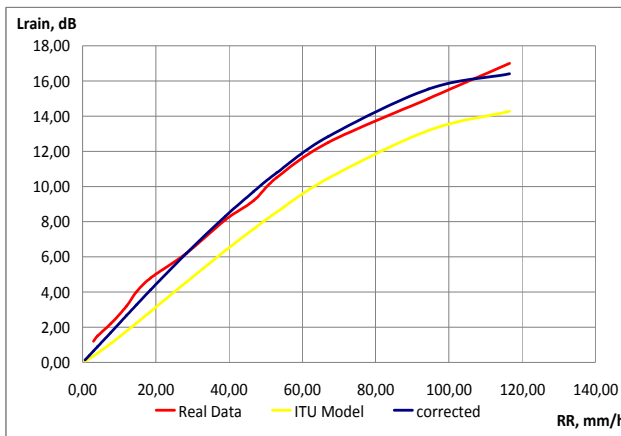


Fig. 1. Rain attenuation for 7,1 km radio link,  $f=11,15$  GHz

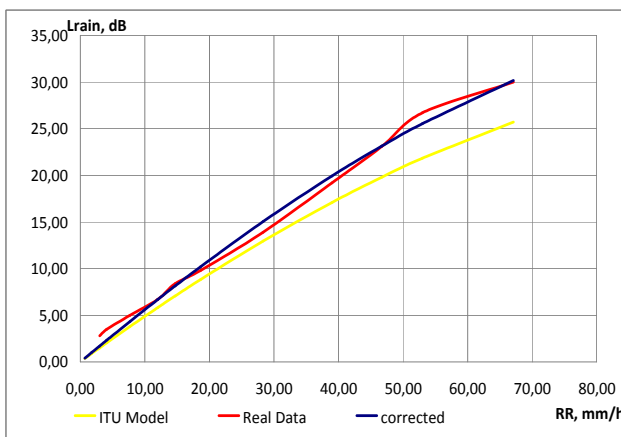


Fig. 2. Rain attenuation for 7,1 km radio link,  $f=19,15$  GHz

We assume the rain rate statistic is the same for the every part of link distance, because the measurement period is long enough.

Fig.1 and Fig. 2 show the similar disposition of the drawings for theoretical and real rain attenuation of studied radio link. Obviously the real attenuation is 1 dB for light rain and 3-4 dB for heavy rain bigger than the theoretical for two studied frequencies. The correction obtained from Eq. (3) is not the frequency dependent and the difference can be searched out in the specific rain attenuation calculation – Eq. (2). One of the possible reasons can be found in the rain drops size distribution that is different for the rainfalls occur in different time but with the same rain rate. That for example causes serious differences in optical wave attenuation [6] and also can lead to different specific attenuation for the given frequencies. That give us the ground to precise the coefficients  $k(f)$  and  $\alpha(f)$  in Eq. (2).

First it is calculated the specific attenuation value according the real measured rain attenuation and then by using the same model based on Eq. (2) is obtained the new values of  $k(f)$  and  $\alpha(f)$ . The smallest squares method is used [7].

The corrected values of rain attenuation are calculated and are shown with the blue line on Fig. 1 and Fig. 2. The corrected values of the coefficients are given in Table II.

TABLE II

Frequency, GHz	$k(f)$	$\alpha(f)$
11,15	0,03485	1,04390
19,15	0,096502	1,00477

#### IV. CONCLUSIONS

From applied drawings can be viewed that in a heavy rain conditions the calculated by ITU model rain attenuation and the real one are rather different. This can cause mistakes in the link budget calculation. One of possible reasons is the rain drops size distribution that can be different for the different regions of a given country what is in some of climate regions according ITU prediction model [8]. Therefore there is a need of specifying the rain attenuation prediction models and one of possible ways is by defining more accurately the coefficient in these models. Another way is to divide the climate regions on sub-regions according to the specific meteorological conditions. This demands more experimental data for rain attenuation and rain rate in the typical regions like flat country, mountain etc. and is the object of our further studies.

#### ACKNOWLEDGEMENT

This work was supported under Project Nr *DUNK-01/03 – 12.09 “University Scientific and Research Complex for Innovation and Transfer of Knowledge in Areas of Micro/Nano Technologies and Materials, Power Effectiveness and Virtual Engineering”*.

#### REFERENCES

- [1] V. Kvicera and M. Grabner, Results of long-term concurrent measurement of rain rate and rain attenuation at 38 GHz, Proceedings of 2002 URSI General Assembly, Maastricht, Netherlands, 17-24 August 2002.
- [2] ITU-R Recommendation P.838-3 (03/05) Specific attenuation model for rain for use in prediction methods
- [3] Alexandrova E., V. Sviatogor, V. Pozhidaev, A. Kavecki, Results from experimental investigation on rain attenuation over terrestrial radio links at frequencies 11,5, 19,3 and 29,3 GHz, “Electropromishlenost i priborostroene”, vol. 7, 1989, pp. 15-19
- [4] J. Seybold, Introduction to RF propagation, John Wiley & Sons Inc., Hoboken, New Jersey, 2005.
- [5] R. L. Freeman, Radio System Design for Telecommunications, John Wiley & Sons Inc., Hoboken, New Jersey, 2007.
- [6] Colvero, C., M. Cordeiro, J. von der Weid, FSO Systems: Rain, Drizzle, Fog and Haze Attenuation from Different Optical Windows Propagation, International Microwave and Optoelectronics Conference 2007, October 29 - November 1, 2007, Salvador, Brazil.
- [7] Ferdinandov, E., B. Pachedjjeva, Probability and Statistical Methods in Communications – Part I, Siela, Sofia, 2005.
- [8] ITU-R Recommendation P.837-5. Characteristics of precipitation for propagation modeling.

# Multiresolution Analysis of Multiple Reflections in Transmission Lines

András Fehér<sup>1</sup>, Ádám Békefi<sup>2</sup> and Szilvia Nagy<sup>3</sup>

**Abstract** – Wavelet analysis is applied to time-domain signals of conducted measurements on cables with multiple reflections in order to detect repeating patterns in the time-frequency domain. In conducted radio frequency measurements the reflection is one of the most varying components of the measurement uncertainty.

**Keywords** – Reflection, wavelet analysis, measurement, measurement uncertainty.

## I. INTRODUCTION

In conducted radio measurements multiple reflections due to connections are always present, even if matched cables are used, and this effect plays an important role in the measurement uncertainty. In this article we present an analyzing method based on wavelet transform to study the nature of multiple reflections in a cable. The method is also tested with an artificially mis-matched cable, which has a part with 75 Ω impedance between two 50 Ω impedance pieces, without impedance matching.

## II. ON WAVELET ANALYSIS

Wavelet analysis or multiresolution analysis [1–3] is a widely used tool in data processing, especially in image compression and noise reduction [4,5], but it can be also used for solving differential equations [6–8]. The results of the wavelet analysis can also detect patterns.

Let  $f$  be a function of the space of the square integrable functions  $L^2(\mathbb{R})$ . Wavelet analysis can be introduced e.g., as a generalization of the windowed Fourier transform

$$\mathcal{F}^{WFT}\{f\}(\tau, \omega) = \int w(t - \tau)f(t)e^{-i\omega t} dt, \quad (1)$$

and its discretised version

$$\mathcal{F}_{jk}^{WFT}\{f\} = a_{jk}^{WFT} = \int w(t - \tau_j)f(t)e^{-i\omega_k t} dt. \quad (2)$$

Here  $w(t)$  is the window function,  $\tau$  and  $\omega$  are the time and angular frequency of the transformed signal, and  $j$  and  $k$  are the indices of the values resulting from the discrete transformation, belonging to  $\tau_j = j \cdot \tau_0$  and  $\omega_k = k \cdot \omega_0$ , respectively. The reproduction, or synthesis of the original signal from its continuous transformed counterpart can be written as a double integral, similar to the original inverse Fourier integral, whereas the synthesis of the discrete windowed Fourier transformed signal from its coefficients is to be calculated as

$$f(t) \propto \sum_j \sum_k \bar{a}_{jk}^{WFT} \cdot \bar{w}_{jk}(t), \quad (3)$$

where  $\bar{w}_{jk}(t)$  is the basis function of the transformation

$$\bar{w}_{jk}(t) = \overline{w}(t - \tau_j) \cdot e^{i\omega_k t}, \quad (4)$$

with overline meaning the complex conjugation. Using this notation, transformation (2) can be rewritten in a shortened form

$$\mathcal{F}_{jk}^{WFT}\{f\} = a_{jk}^{WFT} = \int w_{jk}(t)f(t)dt. \quad (5)$$

The wavelet transform has similar formula, except that the transformation function is different, i.e., in case of discrete transformation

$$\mathcal{F}_{jk}^{WaT}\{f\} = d_{jk} = \int \psi_{jk}(t)f(t)dt. \quad (6)$$

where the wavelets

$$\psi_{jk}(t) = 2^{j/2} \psi(2^{-j}t - k) \quad (7)$$

Similarly to the window functions, there are many types of wavelets, but once its type is chosen, all the wavelets are generated as scaled and shifted versions of one function, the mother wavelet  $\psi(t)$ . This means that while in case of the windowed Fourier transform, the envelop of the transforming function remains the same (only shifted in time), and increasing frequency manifests in more oscillations within the window function, in case of the wavelet transform the shape of the transforming function remains similar, only shrunk with the increasing frequency. A demonstration for the basis function shape can be seen in Fig. 1.

The inverse transformation takes the form

<sup>1</sup>Andras FEHÉR, Széchenyi István University, Radio Frequency Test Laboratory, Egyetem tér 1, Győr, Hungary, H-9026 E-mail: afeher@sze.hu. Web: <http://rf.sze.hu>

<sup>2</sup>Ádám BÉKEFI, Széchenyi István University, Department of Telecommunications, Egyetem tér 1, Győr, Hungary, H-9026 E-mail: bekefiadam@index.hu.

<sup>3</sup>Szilvia NAGY, Széchenyi István University, Department of Telecommunications, Egyetem tér 1, Győr, Hungary, H-9026 E-mail: nagysz@sze.hu.

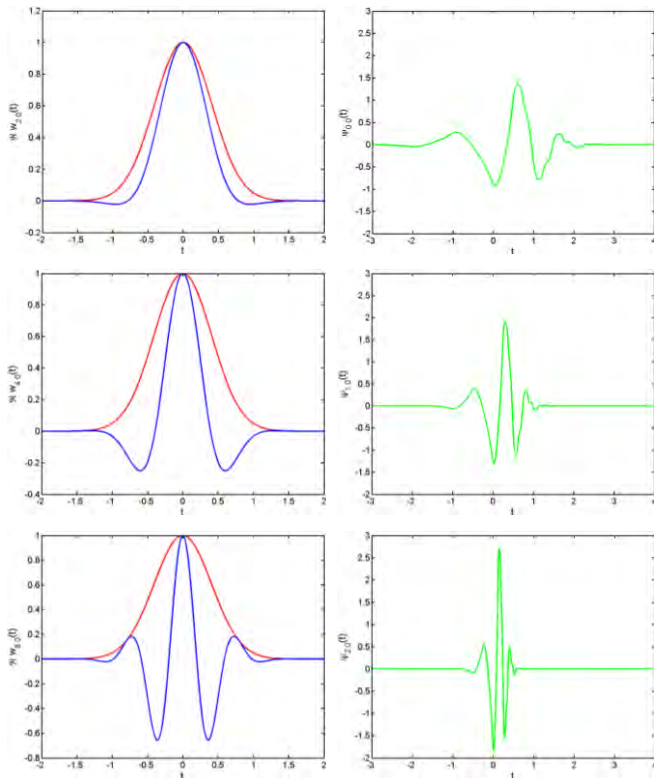


Fig. 1. Basis functions of a windowed Fourier transform (blue lines) with Gaussian window (red lines) and a wavelet (Daubechies-4, green lines) at three frequencies or resolution levels.

$$f(t) = \sum_{k=-\infty}^{\infty} c_{0k} \phi_{0k}(t) + \sum_{j=0}^{\infty} \sum_{k=-\infty}^{\infty} d_{jk} \psi_{jk}(t) \quad (8)$$

In this equation the first summation is necessary to set the DC or low frequency components of the signal  $f(t)$ , whereas the second part makes the higher frequency refinements. The basis elements  $\phi_{0k}(t)$  of the lowest frequency part are the scaling functions. They behave similarly to wavelets in Eq. (7). With increasing the frequency index  $j$ , the corresponding frequency doubles, as it can be seen from (7) and Fig. 1. It can also be observed in (7) that by increasing  $j$  by 1, the shift distance is halved. This property is very favourable if sharp edges, quick changes have to be reproduced, as near the changes the wavelet coefficients  $d_{jk}$  are large, and they are negligible in regions where smooth changes are present only. Such functions are often present both in image processing and in one dimensional data analysis, and they cannot be treated properly with windowed Fourier analysis, discrete cosine transform, etc., where the window width is constant, and usually much larger than the edge which is studied by it.

In practical applications, where a one or two dimensional digital (sampled and quantized) signal is analyzed, the highest frequency corresponds to the sampling frequency. During the analysis the vector (matrix in 2D) is transformed according to Fig. 2, consecutively. The frequency domain is always halved,

the high-pass part will belong to the actual wavelets, the low pass one to the scaling functions; this can be further analyzed. Downsampling is needed, thus the total number of expanding coefficients remain constant after each step;  $c_{j-1,l}$  and  $d_{j-1,l}$  are half as long as the starting  $c_{j,l}$ .

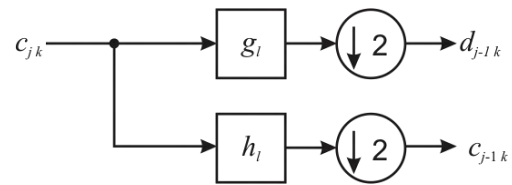


Fig. 2. Schematic diagram of one step in the wavelet analysis. Filter coefficients  $g_l$  and  $h_l$  represent the high-pass and low-pass filters, respectively; the circles mean downsampling by 2. The starting vector  $c_{j,k}$  can be either the original sampled signal or the low-pass output of the previous step.

The synthesis or reconstruction step is opposite to the analysis step plotted in Fig. 2.

### III. STUDY OF MULTIPLE REFLECTIONS IN COAXIAL CABLES

In order to study multiple reflections we have prepared a wrong connecting cable from Hirschmann KOKA 709 (75  $\Omega$ ), and H155 (50  $\Omega$ ) low loss coaxial cables. 50  $\Omega$  instruments were used for the measurement, and the connecting ends of the cable under test (CUT) were the H155 type lines, the middle part was substituted by 1.62 m of KOKA 709 line. As a reference high precision cable with attenuator was applied.

As a first step, the transmission characteristics of the cable was determined by a network analyzer; its parameter S21 can be seen in Fig 3. A clear resonance valley is present at the first marker, near 105 MHz.

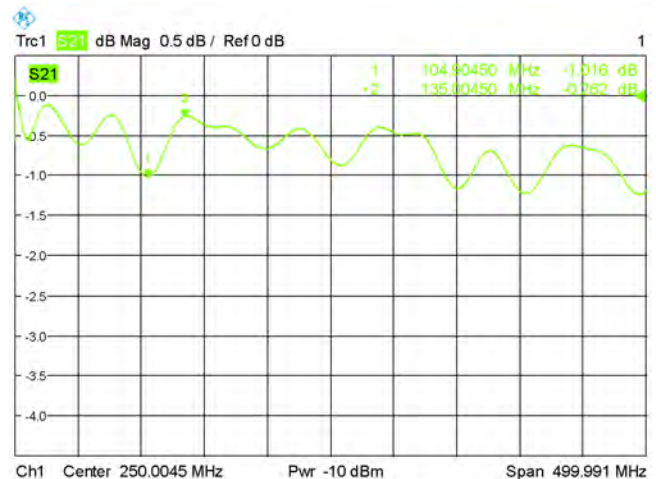


Fig. 3. S21 of the studied cable measured by Rohde&Schwarz ZVL Network Analyzer (9 kHz to 6 GHz)

As a test signal we have applied a carrier signal modulated by a 30 ns burst with a period of 200 ns. The carrier frequencies were near the 105 MHz point as the transmission

parameter varied there rather quickly, thus the modulated signals two sidebands had really different propagation conditions. Wavelet analysis is efficient where the quick changes in the time domain are present, hence the usage of the short bursts as modulator signals, practically the edges and their near environments are interesting. The cable was measured by a 5 GS/s Tectronix oscilloscope, an example with its reference signal can be seen in Fig. 4. The modulator and modulated signals were not synchronized in order to be able to study different relative phases, thus different shapes of

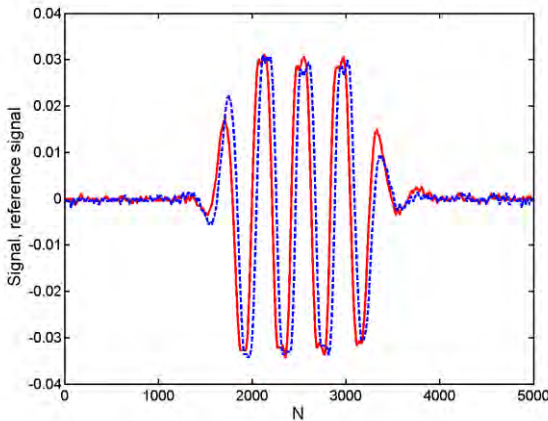


Fig. 4. An example of the measured signal with a carrier frequency of 120 MHz coming through the cable with reflection points (dashed blue line) and a reference cable (continuous red line). The horizontal axis is the number of the sampling.

signals.

Automatic measurement environment was prepared to gather sufficiently large number of data vectors, at the frequencies 93 MHz and 120 MHz. one hundred of measurements were carried out and wavelet transformed. The resulting coefficients varied a lot, as it is demonstrated in Fig. 5.

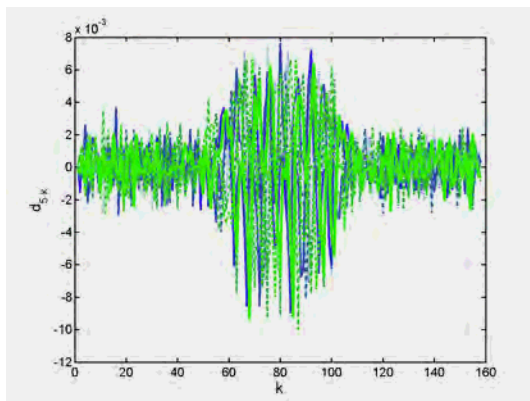


Fig. 5. The 5<sup>th</sup> wavelet transforms of 10 different signals from the cable with multiple reflection points.

In order to see the trends, the square of the coefficients  $d_{jk}$  were summarized for all the signals. The result's square root was normalized by the number of measurements,

$$\langle d_{jk} \rangle = \frac{\sqrt{\sum_{n=1}^{N_m} (d_{jk}^{[n]})^2}}{N_m}, \quad (9)$$

and plotted in Fig. 6 for  $N_m = 10$  and  $N_m = 100$ . The upper index  $n$  means the serial number of the measurements,  $N_m$  the total number of measurements.

Fig. 7 contains the results for various resolution levels, i.e., for various frequency components. Ten analysing steps ( $j = 0 \dots 9$ ) were carried out, but only those are given where interesting characteristics can be seen. The lower frequency

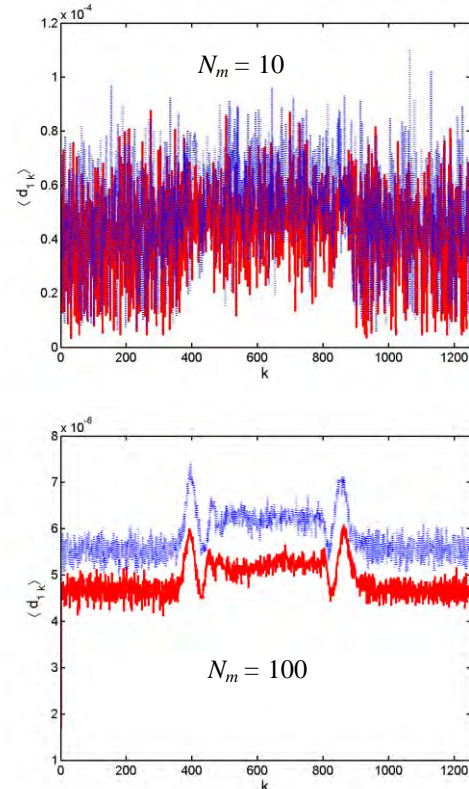


Fig. 6. The effective wavelet coefficients for 10 (upper subfigure) and 100 measurements. Data of cable with multiple reflection points is plotted with blue dashed line, the reference with red continuous line. Resolution index  $j = 1$ .

components do not differ significantly from the reference.

#### IV. CONCLUSION

The normalized wavelet coefficients of the reference and the multiple reflected signals can be distinguished in higher frequencies, however, the low frequency terms are in approximately the same, at least in average. The sampling time  $t_s = 400\text{ps}$  of the two-channel oscilloscope is just about 4 to 5 refinement steps away from the 93-120 MHz carrier frequency's characteristic time, thus the high frequency effects of the reflections and nonlinearities should be found in the first couple of steps, thus our results meet the expectations. Also,  $t_s$  becomes commensurable with the 30 ns

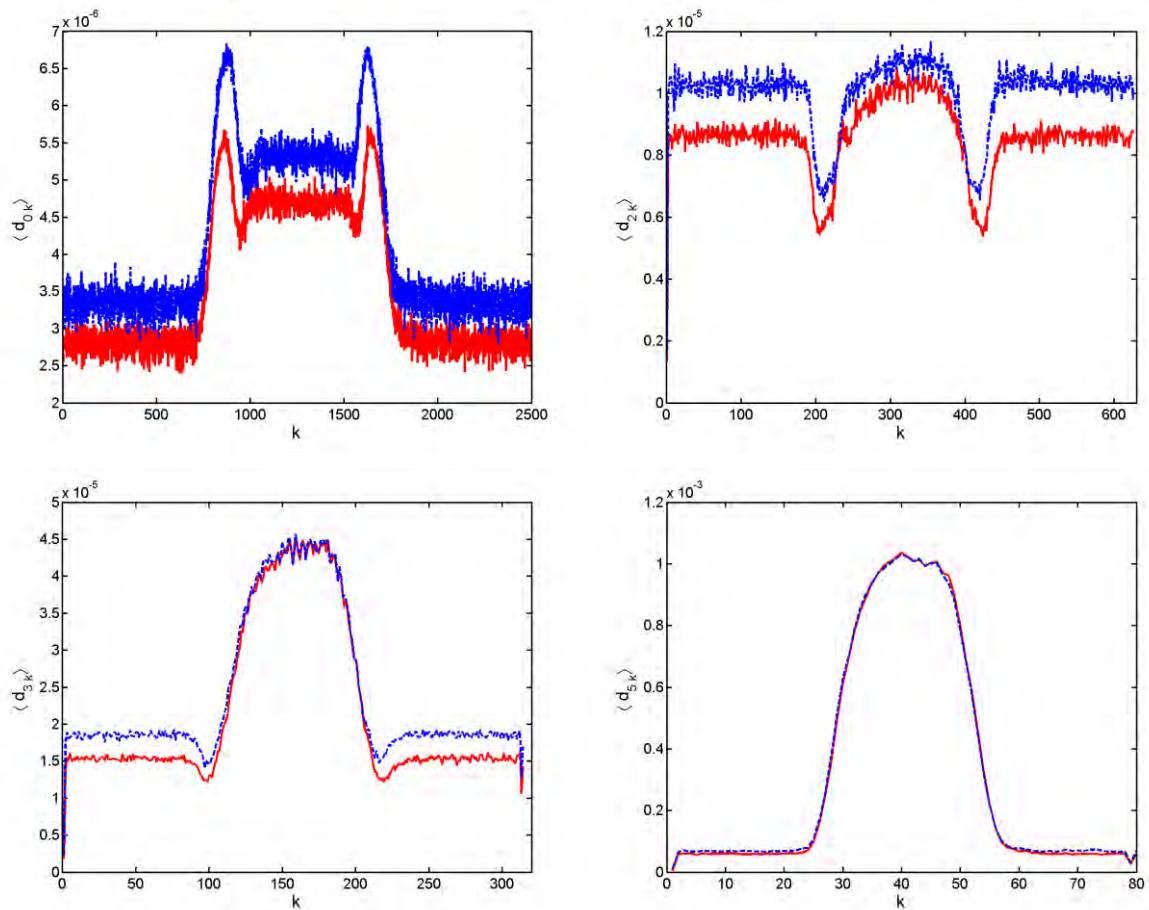


Fig. 7. The effective wavelet coefficients after 100 measurements.

burst time after 6-7 refinement steps, thus the lower frequency terms will contain mostly the burst itself. Multiple reflection in conducted measurements can be characterized by a significant increment in the average fine resolution wavelet coefficients.

ACKNOWLEDGEMENT

This work was supported by the Hegedüs Gyula fellowship and the projects TÁMOP-4.2.1./B-09/1/KONV-2010-0003, and TÁMOP 4.1.1.A-10/1/KONV-2010-0005 of the Széchenyi István University.

REFERENCES

[1] I. Daubechies, *Ten Lectures on Wavelets*, CBMS-NSF Regional Conference Series in Applied Mathematics 61 (SIAM, Philadelphia, 1992).

[2] C. K. Chui, *An Introduction to Wavelets*, (Academic, San Diego, 1992).  
 [3] S. Mallat, *A theory for multiresolution signal decomposition: the wavelet representation*; IEEE Trans. Pattern Anal. Mach. Intel., vol. 11, pp. 674-693 (1989).  
 [4] J. Kaftan, A. A. Bell, C. Seiler, T. Aach, *Wavelet based denoising by correlation analysis for high dynamic range imaging*; IEEE International Conference on Image Processing, Cairo, pp. 3857-3860. (2009).  
 [5] C. Christopoulos, A. Skodras, and T. Ebrahimi, IEEE Trans. Consum. Electron. vol. 46, p. 1103 (2000).  
 [6] K. Urban, *Wavelet Methods for Elliptic Partial Differential Equations*, Oxford University Press, Oxford, (2009).  
 [7] W. Dahmen, *Wavelet methods for PDEs—Some recent developments*, J. Comput. Appl. Math. vol. 128, p. 133 (2001).  
 [8] J. Pipek, Sz. Nagy, *Refinement trajectory and determination of eigenstates by a wavelet based adaptive method*, J. Chem. Phys. vol. 125, 174107 (2006).  
 [9] M. Kuczmann, A. Iványi, *Finite Element Method in Magnetics*, Academic Press, Budapest, 2008, ISBN: 978 963 05 8649 8.



# Ad-Hoc Supported, Connection Fault-Tolerant Model for Mobile Distributed Transaction Processing

Tome Dimovski<sup>1</sup> and Pece Mitrevski<sup>2</sup>

**Abstract** – Mobile embedded systems increasingly use transactions for applications like mobile commerce, banking or commercial applications. We make a short review of the Connection Fault-Tolerant Model for mobile distributed transaction processing. We made simulation experiments study. The performance analysis shows that ad-hoc communication supports considerably improve the transaction commit rate.

**Keywords** – Mobile transaction, mobile computing environment.

## I. INTRODUCTION

The increasing emergence of mobile devices contributes to rapid progress in wireless technologies. Mobile devices interacting with fixed devices can support applications such as e-mail, mobile commerce (m-commerce), mobile banking etc. But there are many issues that are challenging and need to be resolved before enabling mobile devices to take part in distributed computing. For distributed systems the transaction is a set of operations that fulfill the following condition: either all operations are permanently performed, or none of them are visible to other operations (known as the atomicity property). In the execution of transactions the key issue is the protocol that ensures atomicity.

The mobile environment is characterized with mobile devices with limited resources like processing, storage, energy capacity and continuously varying properties of the wireless channel. This increases the time needed for mobile hosts to execute transactions and can even lead to execution failure. *Mobile hosts (MHs)* naturally show frequent and random network disconnections. The limitations and characteristics of the mobile environment make it harder to design appropriate and efficient commit protocols. A protocol that aborts the transaction, each time the MH disconnects from the network, is not suitable for mobile environments because it is part of the normal mode of operation. Disconnections need to be tolerated by the protocol.

The *two-phase commit (2PC)* protocol [1] that allows the involved parties to agree on a common decision to commit or abort the transaction even in the presence of failures is the most commonly used protocol for fixed networks but is unsuitable for mobile environments. There are several other protocols, [3-5], [7], for transaction execution in distributed mobile environment, but almost all consider limited number of communication models.

<sup>1</sup>Tome Dimovski is with the Faculty of Technical Sciences at University “St. Kliment Ohridski” of Bitola, Ivo Lola Ribar bb, Bitola 7000, Macedonia, E-mail: tome.dimovski@uklo.edu.mk.

<sup>2</sup>Pece Mitrevski is with the Faculty of Technical Sciences at University “St. Kliment Ohridski” of Bitola, Ivo Lola Ribar bb, Bitola 7000, Macedonia, E-mail: pece.mitrevski@uklo.edu.mk.

The main contribution of this paper is performance analysis of the Connection Fault-Tolerant Model [8] for mobile distributed transaction processing. We study the impact of the level of ad-hoc communication support on the mobile transaction commit rate.

The paper is organized as follows. Section II gives a survey of related work. In Section III we present model of the mobile environment, and in Section IV transaction model. In Section V we present our Connection Fault-Tolerant Model. In Section VI, we present simulation results and analysis. Section VII discusses conclusion.

## II. RELATED WORK

All *Transaction Commit On Timeout (TCOT)* [3] protocol is based on timeout approach for Mobile Database Systems, which can be universally used to reach a final transaction termination decision in any message oriented system. This protocol limits the amount of communication between the participants in the execution of the protocol. TCOT does not consider mobile hosts as active participants in the execution of transactions.

The basic idea of the *Two-Phase Commit Protocol for Mobile Wireless Environment (M-2PC)* [4] is to adapt the 2PC protocol for mobile systems with distributed transactions. Mobile hosts are active participants in execution of transaction and they are sending confirmation that the work is done to the agent or to the fixed device in order to save energy. This protocol does not provide adequate management of mobility and failures caused by the network disconnection.

*Fault-Tolerant Pre-Phase Transaction Commit (FT-PPTC)* [5] protocol provides mechanisms for dealing with disturbances in the systems in mobile environment. The protocol supports heterogeneous mobile databases. FT-PPTC implements distributed transaction in two phases: pre-phase, one which is covering the mobile hosts and the main phase which refers to the fixed part of the network. Mobile hosts are active participants in execution of transaction. No mechanisms are developed for competition in the mobile distributed transactions. FT-PPTC doesn't provide adequate management of mobility because when mobile hosts are disconnected from fixed network for a long time they can block resources on the fixed participants. This leads to an increased number of mobile transaction aborts.

## III. MODEL OF THE MOBILE ENVIRONMENT

In this paper we consider system model for a mobile distributed environment consisting of a set of *mobile hosts (MHs)* and a set of *fixed hosts (FHs)*, presented in Fig. 1. The model has two main parts: the fixed part of the network and a

mobile part of the network. Communication between the fixed part and the mobile part of the network is conveyed via *Mobile Support Stations (MSS)*. MSSs are connected to the fixed part of the network via wired links. MHs can cross the border between two different geographical areas covered by different MSSs.

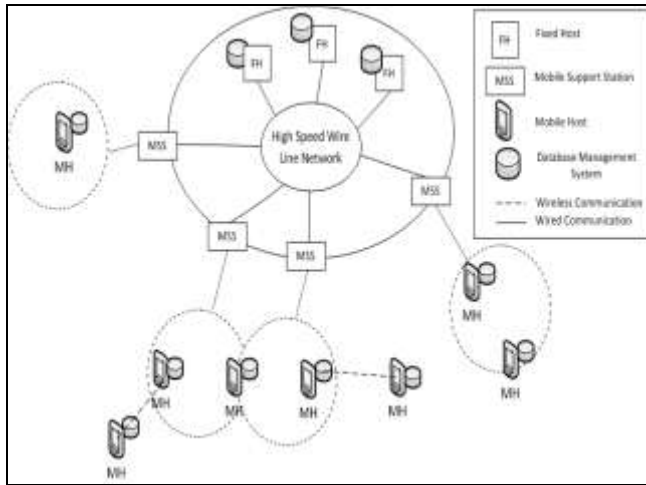


Fig. 1. Model of the Mobile Environment

In the considered system model, first, MHs can communicate with the FHs through MSS via wireless channels only when they are located within the MSS coverage area. Second, the MHs can ad-hoc communicate with neighboring MHs via wireless channels. When MHs enter a geographical area that is out of coverage of any MSS, to access database servers on fixed network they may connect through a neighboring MH which is in the covering area of any MSS.

In our scenario, we consider mobile distributed environment where MHs can communicate with each other, or with FHs through MSS, and MHs can ad-hoc communicate with neighboring MHs to reach fixed network.

We assume that database servers are installed on each FH, and each MH has a mobile database server installed.

#### IV. TRANSACTION MODEL

A distributed transaction where at least one MH participates is called a *Mobile Transaction (MT)*. We identify a MH where a transaction is issued as a *Home-MH (H-MH)*. Participating MHs and FHs in the execution of a mobile transaction are called *participant MHs (Part-MH)* and *participant FHs (Part-FH)*.

In our model we assume the existence of a *Coordinator (CO)* which is responsible for coordinating the execution of the corresponding transaction. The CO is responsible for storing information concerning the state of the transaction execution. Based on the information collected from the participants of the transaction, the CO takes the decision to commit or abort the transaction and informs all participants about its decision. The CO should be executed on the fixed host or hosts. That means logs will be kept more safely.

#### V. CONNECTION FAULT-TOLERANT MODEL

##### A. Overview

Some of the most frequent failures in mobile environments are communication failures. When MHs are in motion, they may exit the geographical area that is covered by some MSS and the resources of the fixed participants may potentially be blocked for an undefined period of time. If MHs do not reestablish connection with any MSS the transaction is aborted.

To minimize the number of mobile transaction aborts by tolerating failures caused by network disconnections we proposed Connection Fault-Tolerant (CFT) Model [8] for distributed transaction processing in mobile computing environment. The CFT Model ensures the atomicity property.

The CFT model considers two communication scenarios:

1. The first scenario is a *Standard communication scenario* when MHs can directly connect to fixed network through MSSs.

2. The second scenario is an *Ad-hoc communication scenario* when MHs cannot directly connect to the fixed network through any MSS. In this scenario MHs can ad-hoc communicate with neighboring MHs to reach the fixed network.

In the Standard scenario, similar to [6], to minimize the use of the wireless communication and conservation of the resources of MHs we assign a *Mobile Host Agent (MH-Ag)*, which we add in the fixed network, to each MH. We assume that in the execution of a transaction MH-Ag is representing the MH in the fixed network and it acts as an intermediary between MH and the transaction CO. All communications between MH and CO go through the MH-Ag. The MH-Ag is responsible for storing all the information related to the states of all MTs involving the MH. In the fixed network, a server or servers can be designated, where MH-Ag are created for each participating MH.

In our CFT model we define additional function to MH-Ag that we called *Decision Algorithm (DALg)*. DALg is used during the execution of a transaction when MH-Ag cannot directly or ad-hoc communicates with its MH for a defined period of time. DALg's task is to check if *Transaction Processing Fragment (TPF)* function is **write** (insert/update/delete) or **read**. If TPF function is **write**, DALg saves the TPF in FIFO (First In First Out) queue list and makes a decision for MH to send "Yes" vote to the transaction CO. When the connection between MH and corresponding MH-Ag is reestablished, MH-Ag's first task is to send all saved TPFs to the corresponding MH. If TPF function is **read**, DALg will wait for connection reestablishment between MH and corresponding MH-Ag, for a defined period of time. If the connection is not reestablished in the specified time period, DALg makes a decision for MH to send "No" vote to the transaction coordinator.

The second communication scenario is when MHs cannot directly connect to the fixed network, or MH-Ag cannot directly communicate with its MH through any MSS. In that case, they try to connect through ad-hoc communication with any neighboring MH which is in the covering area of any

MSS. To allow this, we assign a *MH-Relay Agent (MH-RAg)* to each MH. The *MH-RAg* is responsible for ensuring relay wireless link between neighboring MHs. This means that MH which is out of the coverage area can connect to his MH-Ag of the fixed network via *MH-RAg* of the neighboring MH which is in coverage area of any MSS.

*B. Connection Fault-Tolerant Model operation*

In this section we make a short review of the operation of the CFT model that has been presented in [8]. Fig. 2 illustrates the execution of a mobile transaction for the proposed model, but without the functions of the Decision Algorithm because in this paper our interest is concentrated on the Ad-hoc communication impact on mobile transaction execution.

If H-MH is connected to the fixed network through some MSS, it initiates a mobile transaction by sending transaction processing fragment (TPF) to the coordinator through its corresponding MH-Ag which acts as mediator between coordinator and MH.

Transaction coordinator computes the execution time-out ( $E_t$ ), which is a time limit for all participants to complete the execution of the TPFs and send a VOTE to coordinator. After that coordinator sends  $E_t$  and TPFs to all participating FHs and MH-Ags, asks them to prepare to commit the transaction, and enters the wait state. MH-Ags try to send TPFs to their participating MHs. If MH-Ag cannot communicate with a MH through standard communication, it tries to connect through ad-hoc communication to any neighboring MH which is in the covering area of any MSS. That function is assigned to the MH-Relay Agent.

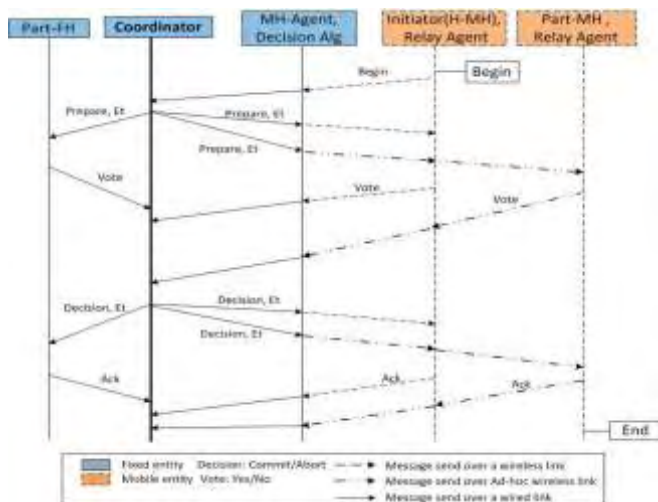


Fig. 2. Mobile transaction processing

When the participants receive the “prepare” message, they check if they could commit the transaction. If so, participants send YES vote to the coordinator. MH sends vote to coordinator through its corresponding MH-Ag via standard or ad-hoc communication.

After the coordinator has received vote from every participant, it decides whether to commit or abort the transaction. If, for any reason, even one of the participants votes “No” or execution time-out expires, the coordinator decides to abort the transaction and sends “Abort” message to all participants. Otherwise, if all received votes are “Yes” and execution time-out is not expired, the coordinator decides to commit the transaction and sends “Commit” message, with reset execution time-out to all participants. The participants need to acknowledge the coordinator decision before execution time-out expires.

VI. SIMULATION RESULTS AND ANALYSIS

The simulation experiments study the impact of ad-hoc communication in CFT Model on the mobile transaction commit rate. For the simulations, we used SimPy [9], a process-based discrete-event simulation package based on standard Python programming language [10]. Table I summarizes our simulation parameters.

TABLE I  
SIMULATION SETTINGS

Parameter	Value
Number of Part-MHs	5-10
Fragment execution time (MH)	0.5s
Fragment execution time (CO)	0.3s
Transmission delay (wireless link)	0.4s
Transmission delay (wireless ad-hoc link)	0.9s
Transmission delay (wired link)	0.2s
Disconnection Rate	0 - 95%
Ad-hoc support	0 - 80%

Hence, disconnection rate is defined as the ratio of time where the participating MH is disconnected from the fixed network, against the total simulation time. Ad-hoc support is the ratio of time where ad-hoc communication is available between MHs, against the total simulation time. It is hard to quantify the level of ad-hoc support between MHs in mobile distributed environment. In some parts of the wireless network ad-hoc support can be lower compared to other. For that reason, in our simulation we define three groups that represent different parts of the wireless network with different level of ad-hoc support. Every MH in the wireless network is a member of one of the defined groups.

A simulation run is set to simulate 10 hours. Transactions are generated with exponentially distributed in-between intervals, with an average of 30 seconds. We assume that all transactions are of similar length, but experience different connection conditions.

The main parameter for performance evaluation is the mobile transaction commit rate. Figs. 3, 4 and 5 show mobile transaction commit rate against different disconnection rates, and different ad-hoc support values. Transaction time-out is set to 5 seconds. For results shown in Fig. 3 we assume that every MH is a member of the same group. It is evident that ad-hoc support in the CFT model considerably improves the

transaction commit rate. The ad-hoc communication impact is higher for networks where disconnection rate is higher.

of committed mobile transactions and shows resilience to connection failures of the mobile devices.

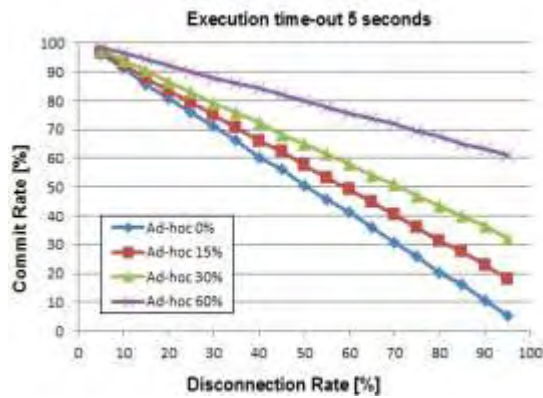


Fig. 3. Impact of ad-hoc communication on commit rate (single group)

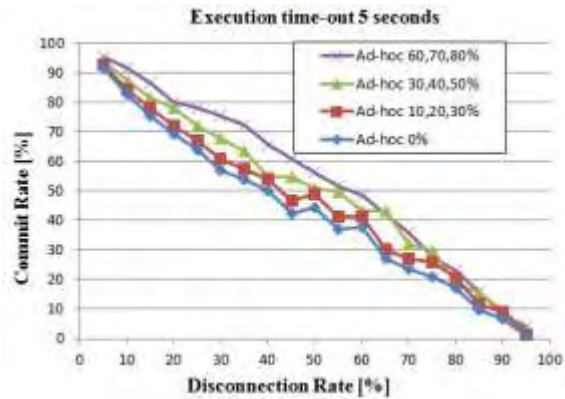


Fig. 5. Impact of ad-hoc communication on commit rate (three groups)

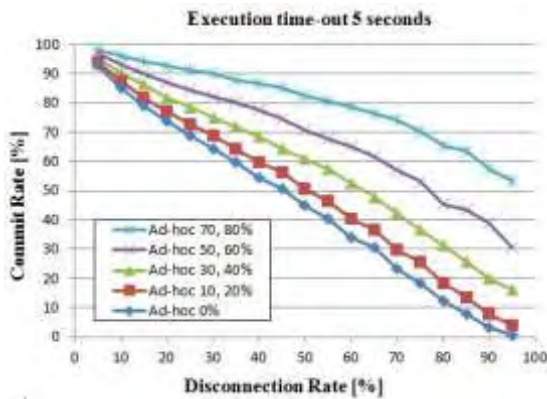


Fig. 4. Impact of ad-hoc communication on commit rate (two groups)

Fig. 4 shows simulation results where MHs are classified in two different groups that have different levels of ad-hoc support. It is evident that ad-hoc support improves the transaction commit rate, but the percentage of improvement is lower compared to the previous scenario where the level of ad-hoc support was the same in each part of the wireless network. Compared to the previous scenario, commit rate slightly decreases when the disconnection rate rises all the way from 60 up to 95 seconds.

To present the influence of ad-hoc support in a highly dynamic wireless network, we classify all MHs in three groups that have different levels of ad-hoc support. From the chart in Fig. 5, one can conclude that commit rate increment is not evident as before, e.g. the highest improvement of the commit rate is about 11%.

## VII. CONCLUSIONS

In this paper we made a short review of the operation of the Connection Fault-Tolerant Model [8] for mobile distributed transaction processing. It is developed to increase the number

The performance analysis shows that ad-hoc support in the CFT model considerably improves the transaction commit rate. It is evident that the impact of ad-hoc support is higher for less dynamic wireless networks.

In our future work we plan to evaluate the performance of a model that includes a decision algorithm, as well, for which we expect to have positive impact on mobile transaction commit rate.

## REFERENCES

- [1] J. Gray, "Notes on Data Base Operating Systems", Operating Systems, An Advanced Course, 1978.
- [2] N. Santos, P. Ferreira, "Making Distributed Transactions Resilient to Intermittent Network Connections", Proceedings of the 2006 International Symposium on on World of Wireless, Mobile and Multimedia Networks. IEEE Computer Society, Washington, 2006.
- [3] V. Kumar, "A Timeout-Based Mobile Transaction Commitment Protocol", Proceedings of the East-European Conference on Advances in Databases and Information Systems, 2000.
- [4] N. Nouali, A. Doucet, H. Drias, "A Two-Phase Commit Protocol for Mobile Wireless Environment", Proc. 16th Australasian Database Conference, 2005.
- [5] B. Ayari, A. Khelil, N. Suri, "FT-PPTC: An efficient and fault-tolerant commit protocol for mobile environments", Proc. of SRDS, 2006.
- [6] L. Xiang, Z. Yue-long, C. Song-qiao, Y. Xiao-li, "Scheduling Transactions in mobile distributed real-time database systems", Journal of Central South University of Technology, 2008.
- [7] S.A. Moiz, M.K. Nizamudin, "Concurrency Control without Locking in Mobile Environments", First International Conference on Emerging Trends in Engineering and Technology, Nagpur, Maharashtra, 2008.
- [8] T. Dimovski, P. Mitrevski, "Connection Fault-Tolerant Model for distributed transaction processing in mobile computing environment", Information Technology Interfaces, 33<sup>rd</sup> International Conference, Dubrovnik, Croatia, 2011.
- [9] URL < <http://simpy.sourceforge.net/>>
- [10] URL < <http://www.python.org/>>

# Fast Synthesis of High Order Microwave Filters by Coupling Matrix Optimization

Marin V. Nedelchev, Ilia G. Iliev

*Abstract:* This paper presents optimization method for synthesis of generalized high order microwave filters with arbitrary topology. The method utilizes local optimizer for coupling matrix determination. The synthesis procedure converges very fast as for a initial point is used a vector based on the Chebyshev all pole filter for the same degree of the filter. To validate the proposed synthesis method two numerical examples for resonant filters are computed. The frequency responses from the synthesis procedure and the theoretical responses show excellent agreement.

*Keywords:* microwave filter, Chebyshev filter, Nelder-Mead optimization, coupling matrix.

## I. INTRODUCTION

Microwave coupled resonator filters play important role in the modern communication systems. The constraint RF/microwave spectrum requires high attenuation in the stop band and low insertion loss in the passband of the filters. These requirements can be met only by cross-coupled microwave filters, realizing attenuation poles on finite frequencies. Cross-coupled resonator filters allow using various topologies with variety of frequency responses. The microwave filter modelling is very important for the fast and accurate design.

Key point in the obtaining of the coupling matrix corresponding to the practical filter topology is to convert its transversal form to folded form using matrix rotations. Most of the matrix rotation sequences are given in [4]. It is noticed that this method for synthesis suffers from generality, because the matrix rotations cannot be derived for every one practical filter topology. Some of the matrix rotation sequences cannot converge in order to find the coupling matrix. Some of the disadvantages in this method are solved if arrow form of the coupling matrix is used [5] or Pfitzenmeir method is used [6].

In many practical cases, it is necessary to define the filter topology in order to satisfy some manufacturing or space requirements. In this case, the exact solution is hard to be found utilizing the conventional synthesis methods.

One possible general solution to the filter design for arbitrary topology is to apply direct local optimization over the coupling matrix with successive starting point. In the basic papers proposed optimization method for coupling matrix synthesis [7,8], the starting vector is set to arbitrary values. This makes the local optimization very unstable method for cost function minimization. Another method is to use global optimization method for finding the coupling matrix for certain filter topology.

Marin Veselinov Nedelchev and Ilia Georgiev Iliev –are with Dept. of Radiocommunication and Videotechnologies in Faculty of Telecommunication in TU –Sofia, N8, Kliment Ohridski bul., 1700 Sofia, Bulgaria. E-mail: [mnedelchev@tu-sofia.bg](mailto:mnedelchev@tu-sofia.bg), [igiliev@tu-sofia.bg](mailto:igiliev@tu-sofia.bg).

They perform robust optimization, no matter about the starting point. Unfortunately the global optimizers such as genetic or stochastic have very slow convergence to the cost function minimum.

This paper presents optimization method for synthesis of high microwave filters with arbitrary topology. The method uses Nelder-Mead local optimizer for coupling matrix determination. The synthesis procedure converges very fast as for a initial point is used a vector based on the Chebyshev all pole filter for the same degree of the filter. The cost function is based on amplitude of the transmission and reflection coefficient zeros and their values at the cut-off frequencies and the reflection coefficient maxima. To validate the proposed synthesis method two resonant filters are designed with asymmetrical responses. The frequency responses from the synthesis procedure and the theoretical responses show excellent agreement.

## II. RESONATOR FILTER CHARACTERISTICS

The synthesis procedure starts with the low-pass prototype with normalized angular frequency of passband  $\omega=1$ . The transfer and reflection coefficients may be expressed as a ratio of two N-th degree polynomials as follows:

$$S_{21} = \frac{P_N(\omega)}{E_N(\omega)}, S_{11} = \frac{F_N(\omega)}{\varepsilon E_N(\omega)} \quad (1),$$

where  $\omega$  is real angular frequency and  $\varepsilon = \left(1/\sqrt{10^{RL/10}} - 1\right) \cdot (F_N(\omega)/P_N(\omega))\big|_{\omega=1}$ ,  $RL$  is the prescribed value of the return loss in  $dB$ , in the passband of the filter. It is assumed that all polynomials are normalized to their highest degree coefficient.

The method of computing the numerator of the reflection coefficient is outlined in [3].

$$F_N(\omega) = \frac{1}{2} (G_N(\omega) + G'_N(\omega)), \quad (2),$$

where both polynomials can be represented by two polynomials:  $G_N(\omega) = U_N(\omega) + V_N(\omega)$  and  $G'_N(\omega) = U_N(\omega) - V_N(\omega)$ . Both polynomials  $U_N(\omega)$ ,  $V_N(\omega)$  can be arranged according to the Cameron's recursive procedure in [3]. Obviously the roots of  $U_N(\omega)$  corresponds reflection zeros, and the roots of  $V_N(\omega)$  correspond to the in-band reflection maxima.

It can be easily found that the transfer coefficient may be expressed in the following way[3]:

$$S_{21}^2(\omega) = \frac{1}{1 + \varepsilon^2 C_N^2(\omega)} \quad (3)$$

where  $C_N(\omega)$  is the filtering function. For general Chebyshev characteristics, the filtering function is in the form:

$$C_N(\omega) = \cosh\left(\sum_{n=1}^N a \cosh(x_n)\right) \quad (4),$$

where  $x_n = \frac{\omega - 1/\omega_n}{1 - \omega/\omega_n}$ , where  $\omega_n$  is the angular frequency of the prescribed transmission zero.

In order to obtain the coupling matrix, it is necessary to consider the equivalent circuit of general coupled resonator filter shown on Fig.1.

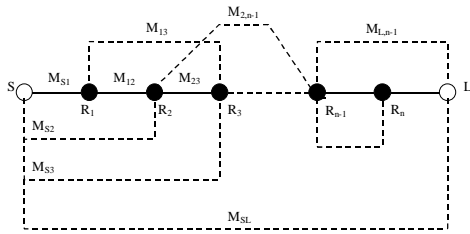


Fig.1. General coupled resonator filter

The equivalent circuit consists of  $N$  series coupled resonators with frequency independent couplings  $M_{ij}$  ( $i \neq j$ ), between the  $i$ -th and  $j$ -th resonators. The circuit is driven by voltage source  $E$  with internal normalized resistance  $R_1=1$  and loaded to normalized impedance  $R_2=1$ . The resonant frequency of each resonator  $f_{oi}$  is represented by the self-coupling coefficient  $M_{ii}$  and the center frequency of the filter. The transmission and reflection coefficients of a lossless filter of  $N$ -th order depend only of the coupling matrix  $|M|$  (7):

$$S_{21} = -2j[A]_{N+2,1}^{-1}, \quad S_{11} = 1 + 2j[A]_{11}^{-1}, \quad (6),$$

where  $[A] = -j[R] + \omega[W] + [M]$ , and  $[R]$  is  $(N+2) \times (N+2)$  matrix, which elements are zeroes except  $R_{11} = R_{N+2,N+2} = 1$ .  $[W]$  is a  $(N+2) \times (N+2)$  matrix, where the main diagonal elements are unity except  $W_{11} = W_{N+2,N+2} = 0$ . All remaining elements of  $[W]$  are zeroes.  $[M]$  is the coupling matrix, symmetrical around the main diagonal.

### III. SYNTHESIS OF MICROWAVE FILTER WITH COUPLING MATRIX OPTIMIZATION

The cost function used in the optimization process is based on the zeroes and poles of the filtering function  $C_N$ , assuming that the number of poles is  $P$  and zeroes  $N$  [8]:

$$Cost = \sum_{i=1}^N |S_{11}(\omega_{zi})|^2 + \sum_{i=1}^P |S_{21}(\omega_{pi})|^2 + \left( |S_{11}(\omega = -1)| - \frac{\epsilon}{\sqrt{\epsilon^2 + 1}} \right)^2 + \left( |S_{11}(\omega = 1)| - \frac{\epsilon}{\sqrt{\epsilon^2 + 1}} \right)^2 + \sum_{i=1}^{N-1} \left( |S_{11}(\omega = \omega_{mi})| - \frac{\epsilon}{\sqrt{\epsilon^2 + 1}} \right)^2 \quad (7).$$

In the cost function  $\omega_p$  are the prescribed transmission zeros,  $\omega_z$  are the zeroes of the reflection coefficient, and  $\omega_m$

are the in-band maxima frequencies. In most papers concerning the optimization of the coupling matrix the last term of the cost function is missing. Because of the high order of the filter, the value of the transmission coefficient at the prescribed zeros  $\omega_p$  is comparable to the precision of the computer. This make the optimization process hard to converge at the global minimum of the cost function. The global minimum is the Chebyshev solution for the microwave filter. As the values of the second term of the cost function needs to be weighted, in order to achieve comparable values to the other terms of the cost function. Obviously there will come up a problem with the choice of the weighted constant. For each filter topology and frequency response, a different constant will be necessary. One possible solution for the problem with the weights is to make each term of the cost function in logarithmic scale with no weight coefficient. Another solution is to add to the cost function another term equalizing the reflection coefficient at its maxima to the ripple factor  $\epsilon$ . In this case the cost function contains all possible constraints for the filter response. The zeros for the transmission coefficient  $S_{21}$  are set at the prescribed frequencies. The reflection coefficient must be zero at the frequencies  $\omega_z$ , equal to  $\epsilon/\sqrt{\epsilon^2 + 1}$  at the normalized cut-off frequencies  $\omega_{cut-off} = \pm 1$  and equal to  $\epsilon/\sqrt{\epsilon^2 + 1}$  at the frequencies  $\omega_m$  at the minimum of the cost function. The cost function may be modified with respect to the transmission coefficient at the frequencies  $\omega_m$ . At these frequencies  $S_{21}$  must be equal to  $\epsilon$ , but the cost function will not be changed in its character.

In this way it is possible to formulate the local optimization problem for obtaining the coupling matrix.

The starting point for optimization of the coupling matrix is very important for the reaching of the global minimum of the cost function (7). Having on mind that a local optimizer is used, the starting vector should be close to the target value in order to assure a fast convergence of the method. One of the possible starting coupling matrices is to set all self-couplings to zero ( $M_{ii} = 0$ ) and all direct couplings to 1. The cross-coupling coefficients are all set to zero. The second possible starting coupling matrix is to use classical Chebyshev filter from the same order. All self- and cross-couplings are set to zero.

The investigation of the problem of high order filter design two numerical designs are investigated.

### III. NUMERICAL RESULTS

For verification of the optimization method presented in this paper, it is applied to an asymmetric resonator filters.

#### A. Asymmetric 9 Resonator Passband Filter

The first numerical example is 9-th order CT filter sharing common resonator. This filter is of Chebyshev type and it has return loss more than 20dB in the passband. The transmission coefficient prescribed zeros are placed on normalized

frequencies  $\omega_p = [-1.8, -1.4, 1.3, 1.6]$ . The coupling diagram of the synthesized filter is shown on Fig.2.

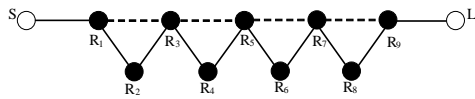


Fig.2. Coupling scheme of an asymmetric 9 pole filter

The reflection and transmission zeroes are calculated and summarized in Table 1.

Table 1. Poles and zeros of asymmetric nine resonator filter

No	Reflection zeros	Transmission zeros	Roots of $V_N$
1	$-j0.9880$	$-j1.8$	$-j0.9514$
2	$-j0.8888$	$-j1.4$	$-j0.7984$
3	$-j0.6795$	$j1.3$	$j0.9566$
4	$j0.9893$	$j1.6$	$j0.8166$
5	$j0.8998$		$j0.5654$
6	$j0.7051$		$-j0.5331$
7	$j0.4006$		$j0.2164$
8	$-j0.3629$		$-j0.1752$
9	$j0.0211$		

The initial point for the coupling matrix elements for the optimization procedure is to set the values of the all pole nine resonator Chebyshev filter  $M_{S1} = M_{9L} = 0.9876$ ,  $M_{12} = M_{89} = 0.9168$ ,  $M_{23} = M_{78} = 0.5870$ ,  $M_{34} = M_{67} = 0.5480$ ,  $M_{45} = M_{56} = 0.5372$ .

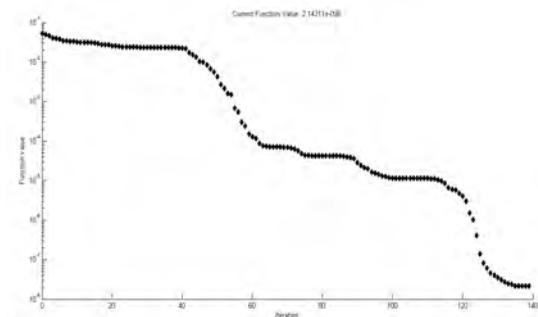


Fig.3 Cost function value during optimization for CT nine resonator filter

All self coupling and cross coupling coefficients are set to zero. The number of the independent values of the coupling matrix is 23.

After 139 iterations for the optimization coefficient, the optimization procedure converges. The values of the cost function vs the number of iterations is shown on Fig.3. The initial value of the cost function is 0.51 and the end value is  $2.143 \cdot 10^{-8}$ . The optimization process stopped because of reaching local minimum of the cost function (7). The final coupling matrix is:

$$M = \begin{bmatrix} 0 & 0.9842 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.0012 & 0.7589 & -0.2868 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.7589 & 0.4095 & 0.5306 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -0.2868 & 0.5306 & -0.0498 & 0.4665 & 0.2788 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.4665 & -0.5535 & 0.4462 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.2788 & 0.4462 & 0.0035 & 0.4844 & 0.2049 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.4844 & -0.4152 & 0.5049 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.2049 & 0.5049 & -0.0656 & 0.4767 & -0.4048 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.4767 & 0.5722 & 0.7031 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.4048 & 0.7031 & 0.0012 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.9842 \end{bmatrix}$$

The frequency response of the designed filter, calculated according to (6) and the coupling matrix derived in the optimization process, is shown on Fig.4. It is clearly seen that the normalized cut off frequency is  $\omega_c = \pm 1$ , while the transmission zero frequencies are exactly at  $\omega_p = -1.8, -1.4, 1.3, 1.6$ . The maximum value of the return loss is with the prescribed value of  $-20dB$ .

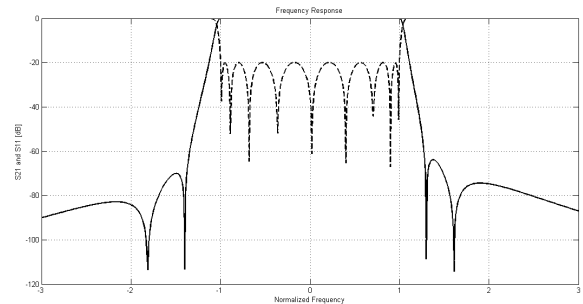


Fig.4 Frequency response of nine resonator filter with asymmetric response. Solid line-  $S_{21}$ , dashed line-  $S_{11}$

### B. Cascaded Quadruplet and Triplet Resonator Passband Filter of 10-th Order

The 10-th order resonator filter is formed by cascade connection of two trisections and one quadruplet section between them (CQT filter). Each trisection realizes one prescribed transmission zero and the quadruplet section realizes two prescribed symmetrical transmission zeros. The filter is of Chebyshev type and it has maximum return loss of  $-20dB$ . The transmission zeroes are placed on frequencies  $\omega_p = [-1.2, \pm 2, 1.6]$ . The coupling scheme of the filter is shown on Fig.5.

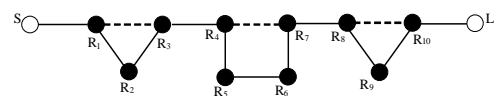


Fig.5. Coupling scheme of CQT filter of 10-th order

The roots of the polynomials in the numerator and denominator in (1) are shown in Table2. The starting point for the optimization process is based on the Chebyshev coupling matrix elements  $M_{S1} = M_{10L} = 0.9854$ ,  $M_{12} = M_{9,10} = 0.8130$ ,  $M_{23} = M_{89} = 0.5839$ ,  $M_{34} = M_{78} = 0.5444$ ,  $M_{45} = M_{67} = 0.5321$ ,  $M_{56} = 0.5321$ .

Table 2. Poles and zeros of CQTfilter

N <sub>z</sub>	Reflection zeros	Transmission zeros	Roots of V <sub>N</sub>
1	j0.9892	-j2	j0.9566
2	j0.9018	-j1.2	j0.8244
3	j0.7246	j1.6	j0.6032
4	-j0.9915	j2	-j0.9655
5	-j0.9201		-j0.8537
6	-j0.7647		-j0.6530
7	-j0.5203		-j0.3696
8	j0.4625		J0.3060
9	-j0.2057		-j0.0342
10	j0.1385		

The number of the independent values of the coupling matrix is 24 The optimization process converges fast in 238 iterations of the optimizer with end cost function value  $1.64519 \cdot 10^{-7}$ . Fig. 6 shows the cost function value with respect to the iterations.

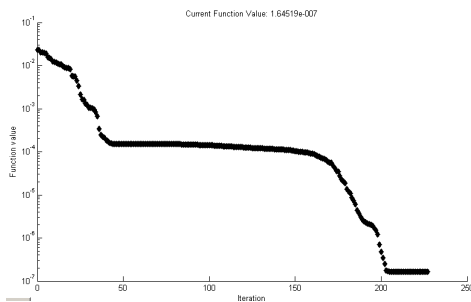


Fig.6 Cost function value for asymmetric five resonator filter

The coupling matrix derived in the optimization process is given by (8). The corresponding frequency response calculated by the coupling matrix and Eq.(6) is shown on Fig.7.

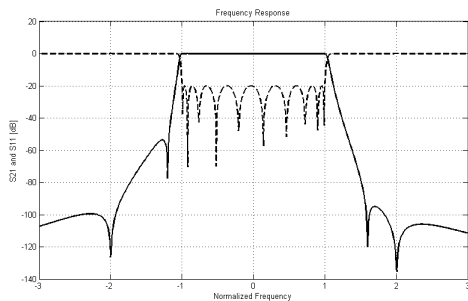


Fig.7 Frequency response of 10-th order CQT filter with asymmetric response. Solid line-S<sub>21</sub>, dashed line- S<sub>11</sub>

As it is clearly seen from Fig.7, the transmission zeros are placed on the prescribed values The maximum value of the reflection coefficient is -20dB. Both presented examples show

fast convergence of the cost function to a local minimum. In both cases this local minimum is found to be a global minimum corresponding to general Chebyshev filter. In both cases the starting point for the optimization process was the coupling matrix of classic Chebyshev filter. Starting from random initial point leads to a local minimum not corresponding to Chebyshev filter.

IV. CONCLUSION

This paper presents optimization method for synthesis of microwave filters with arbitrary topology of high order. The method uses local optimization method for coupling matrix determination. The synthesis procedure converges very fast as for an initial point is used a vector based on the Chebyshev all pole filter for the same degree of the filter. To validate the proposed synthesis method two resonant filters are designed with asymmetrical responses. Both presented examples show fast convergence of the cost function to a local minimum. In both cases this local minimum is found to be a global minimum corresponding to general Chebyshev filter. The frequency responses from the synthesis procedure are within the expectations and found to be consistent with the theoretical responses and given filter specifications.

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, R., Advanced Coupling Matrix Synthesis Techniques for Microwave Filters, IEEE Trans on MTT-50, Jan.2003, pp.1-10
- [3] Cameron, R.J., General Coupling Matrix Synthesis Methods for Chebyshev Filtering Functions, IEEE Trans. On MTT, April 1999, pp.433-442
- [4] Rhodes, J.D., The Design and Synthesis of a Class of Microwave Bandpass Linear Phase Filters, IEEE Trans on MTT 1969 pp.189-204
- [5] Macchiarella, G, An Analytical Technique for the Synthesis of Cascaded N-Tuplets Cross-Coupled Resonators Microwave Filters Using Matrix Rotations, IEEE Trans. On MTT May. 2005 pp.1693-1698
- [6] G. Pfitzenmaier, "Synthesis and realization of narrow-band canonical microwave bandpass filters exhibiting linear phase and transmission zeros," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-30, pp.1300-1311, Sep. 1982.
- [7] Atia W.A., K.A. Zaki and A.E. Atia. "Synthesis of general topology multiple coupled resonator filters by optimization." 1998 MTT-S International Microwave Symposium Digest 98.2 (1998Vol. II MWSYM): 821-824
- [8] Amari, S., Synthesis of Cross-Coupled Resonator Filters Using an Analytical Gradient-Based Optimization Technique, IEEE Trans on MTT Sept. 2000, pp.1559-1564

$$M = \begin{bmatrix} 0 & 0.9832 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.9832 & -0.0019 & 0.6244 & -0.5149 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.6244 & 0.7140 & 0.4020 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -0.5149 & 0.4020 & -0.0802 & 0.5442 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5442 & -0.0293 & 0.5266 & 0 & -0.0427 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5266 & -0.012 & 0.5671 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.5671 & 0.002 & 0.5267 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -0.0427 & 0 & 0.5267 & 0.0157 & 0.5421 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5421 & 0.0563 & 0.5093 & 0.3348 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5093 & -0.4770 & 0.7368 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.3348 & 0.7368 & -0.0019 & 0.9832 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.9832 & 0 \end{bmatrix} \quad (8).$$



# Random High Voltage Impulses Modeling for EMC Testing

Kliment Angelov<sup>1</sup> and Miroslav Gechev<sup>2</sup>

**Abstract** – In this paper an analytic method for random high voltage impulses generating for EMC testing has been described. A statistical assessment of their distribution toward the electrical charge has been made in order to compare with the real physical surges and electric discharges.

**Keywords** – EMC, LEMP, Surge Protection.

## I. INTRODUCTION

Achieving high reliability of modern communication systems imposes the requirement for trouble-free operation in harsh environmental conditions. Much of this environment is dependent on natural factors. For example, a condition is imposed to ensure a device temperature is maintained by air conditioning systems or in certain cases, to take measures against damage from moisture, strong wind, icing, earthquakes etc. For analytical studies of such processes is appropriate to create relevant models and to implement and comply with standards to integrate as well as technical implementation and operation of the apparatus, and in simulating the processes of change resulting from natural factors, especially in construction phase.

When it comes to protection of electronic communication equipment from high voltage disturbances and effects of lightning strikes requirements are also established. Such are, for example, standards of the IEC [1] [2]. Different mathematical models of high disturbing pulses are considered [3] [4], in order to correctly simulate the processes “in case of LEMP” and achieve higher level of protection of electronic communication devices [5].

The purpose of this article is to evaluate the statistical ability to generate random high disturbing pulses with a specific mathematical model. The model could be used to simulate the behavior of communication equipment in situations when taking into account the stochastic nature of the parameters of this type of interference.

## II. DESCRIPTION OF THE PROBLEM

### A. Theoretical basis

High-voltage disturbing pulses are characterized by very

<sup>1</sup>Kliment Angelov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: kna@tu-sofia.bg.

<sup>2</sup>Miroslav Gechev is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: miroslav.gechev@gmail.com

steep climbing forefront and gradually fading rear front of the current in time. Standards have been adopted for different types of effects of lightning strikes, specifying the time duration of these fronts and amplitude value of current. Some of these are:

- 200 kA, 10/350  $\mu$ s – for the first positive and negative direct hit of lightning to the ground;
- 50 kA, 0,25/100  $\mu$ s – for subsequent hits;
- 10 kA, 8/20  $\mu$ s – for secondary lightning strikes and switching surges.

For mathematical models describing the shapes of pulses exponential dependencies can be used. In work [3] the following formula is quoted:

$$I(t) = I_m k (e^{-\alpha t} - e^{-\beta t}) \quad (1)$$

where:

$I(t)$  – current value of the current;

$I_m$  – amplitude value of the current;

$k$  – normalized (scaled in amplitude) coefficient.

The shape of the curve is determined by the coefficients  $\alpha$  and  $\beta$ .

In this study an attempt is made to achieve the modeling of the stochastic nature of high-voltage disturbing pulses, as using a step-exponential function of the type [6]:

$$I(t) = I_m \cdot a \cdot t^b \cdot e^{-c \cdot t} \quad (2)$$

In this case the shape of the curve depends on the coefficients  $b$  and  $c$ , and as using the coefficient  $a$  the amplitude of the pulse can be scaled. By changing these coefficients is possible to obtain the shapes of the curves of the current corresponding to the standard 10/350  $\mu$ s, 0,25/100  $\mu$ s and 8/20  $\mu$ s, which is discussed in other publications of this team [7].

To take into account the random nature of high-voltage phenomena in nature a random number generators can be used. These generators set values to the coefficients of the mathematical models. Changing the form of pulses, on one hand leads to a difference in overall charges of each one, and on the other hand shifts the spectral distribution of the harmonic composition.

B. Test Implementation

Ability to generate random pulses can be used in cases of simulation study of the behavior of the electronic communication equipment under close-to-real conditions. In such case it is necessary pulses to be generated with a maximum close distribution as occurred in nature.

To check whether the model (2) is able to generate trustworthy random pulses a statistical methods applied to the sample of the generated pulses can be used. Preliminary studies performed on model (1) showed that it doesn't give good results in a similar direction.

The generated pulses can be compared in their charge, which can be determined by the following relation:

$$Q = \int_{t=0}^{\infty} I(t).dt, \quad (3)$$

as obtained by (3) charge is in coulombs. For the purpose of comparative study we can work with normalized charge, so the amplitudes of all pulses are equal. It can be assumed that:

$$I(t)_{\max} = 1, \quad (4)$$

which can be achieved by using the coefficient  $a$ . In this case the unit charge will be for 1 coulomb per ampere (C/A), but it can be denoted  $q$ .

A statistical evaluation of the resulting set of normalized charges has to be made. For this purpose we calculate the average value, the dispersion (variation) and the coefficient of variation respectively on dependencies (5) (6) and (7), where  $n$  is the number of random pulses in the sample.

$$\bar{q} = \frac{1}{n} \sum_{i=1}^n q_i \quad (5)$$

$$\sigma_q = \frac{1}{n} \sum_{i=1}^n [(q_i - \bar{q})] \quad (6)$$

$$v_q = \frac{\sigma_q}{\bar{q}} \quad (7)$$

To reflect plausible the actual distribution of occurrence of impulses to their charge, the model should generate random

pulses that are unsymmetrically distributed to the mean value. In such cases the value of the coefficient of variation is significantly greater than 30%, so the normal (Gaussian) distribution is not applicable. Solution can be obtained by using the distribution of Veybul, which can be described as:

$$f(q) = \begin{cases} 0, & q < q_{\min} \\ \left(\frac{k_1}{k_2}\right) \left(\frac{q - q_{\min}}{k_2}\right)^{k_1 - 1} \cdot \exp\left[-\left(\frac{q - q_{\min}}{k_2}\right)^{k_1}\right], & q > q_{\min} \end{cases} \cdot (8)$$

The values of coefficients  $k_1$  and  $k_2$  are determined by the following formulas:

$$k_1 = 0,111186 + 0,835597 \left(\frac{\bar{q}}{\sigma_q}\right) + 0,0759898 \left(\frac{\bar{q}}{\sigma_q}\right)^2; \quad (9)$$

$$k_2 = \frac{1}{n} \left(\sum_{j=0}^n q_j^{k_1}\right)^{1/k_1} \cdot (10)$$

The probability of occurrence of pulses defined or greater amount of charge can be given with the dependency:

$$P(q) = \begin{cases} 1, & q < q_{\min} \\ \exp\left[-\left(\frac{q - q_{\min}}{k_2}\right)^{k_1}\right], & q > q_{\min} \end{cases}; \quad (11)$$

III. RESULTS

Using the software environment MATLAB [8] simulations of the probability distribution obtained with the dependence (2) models of high-voltage disturbing impulses were made. Using a random number generator values are given to the coefficients  $b$  and  $c$ , as Table I indicates the intervals of their variation.

TABLE I  
COEFFICIENT VALUES

Coefficient	10/350 μs	8/20 μs	Random
$b$	0,145	2,78	0,145 ÷ 2,78
$c$	0,00325	0,26	0,00325 ÷ 0,26

In Table I the values of the coefficients  $b$  and  $c$ , in case of pulses corresponding to the standard 10/350  $\mu\text{s}$  and 8/20  $\mu\text{s}$ , are shown. They are the boundary and the distribution of random values in these intervals is given by equal probability law.

Fig. 1 shows a family of randomly generated curves, for clarity of the image their number was reduced to 10. There are variously shaped pulses with different slope of the forefront and rear front respectively integrand area, respectively, different relative charge.

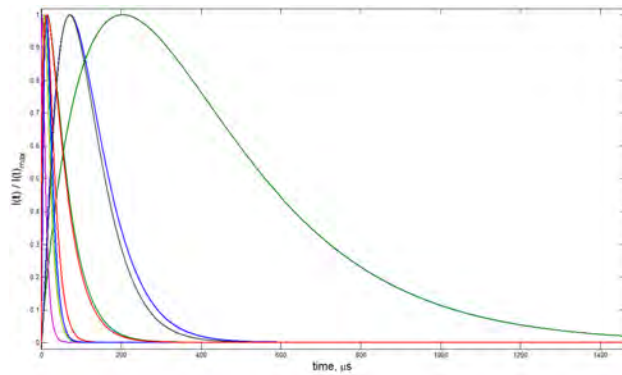


Fig. 1. Random Generated Impulses

To achieve credible data on the statistical analysis is necessary to sample a sufficiently large number of elements. In Table II are given values of average relative charge, dispersion (variation) and coefficient of variation when  $n = 10\,000$ , and minimum and maximum value of the relative charges.

TABLE II  
STATISTICAL PARAMETERS

Parameter	Value	Unit
Mean Value	5,27	$\times 10^{-5}$ , C/A
Standard Deviation	4,73	$\times 10^{-5}$ , C/A
Coefficient of Variation	89,7	%
Number of impulses	10 000	-
Minimum value	5,46	$\times 10^{-6}$ , C/A
Maximum value	1,09	$\times 10^{-3}$ , C/A

Figure 2. shows the density of the probability distribution determined by formula (8) and the values of Table II. There is a minimum threshold of the relative charge below where no pulses are generated. From the perspective of a real physical process, this can be explained by the nature of the occurrence of disturbing impulses. In most cases they are the product of

electrostatic discharge, that breakthrough occurs when they reach a certain critical value of the charge.

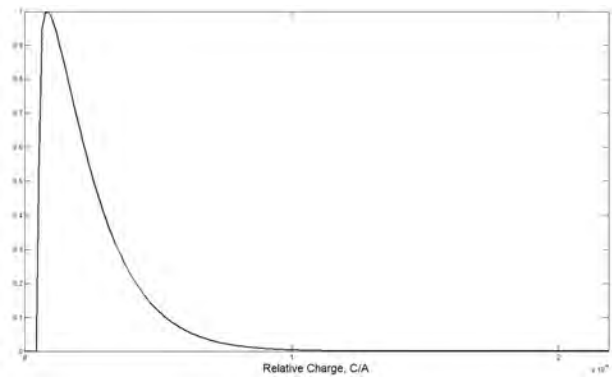


Fig. 2. Probability Density Function

There is a cluster of pulses with relative charges in a specific area, fairly close to the minimum, but there are a few others with significantly higher charges than average. This corresponds to the actual distribution of the disruptive signals in nature - in most cases disturbing impulses are relatively weak, but there also are extremely strong discharges (e.g. direct hit by lightning), under which the charge is significant.

Fig. 3 shows a graph of the probability distribution function of the relative charges. It can determine the probability of occurrence of disturbing impulse with relative charge greater than or equal to a certain value.

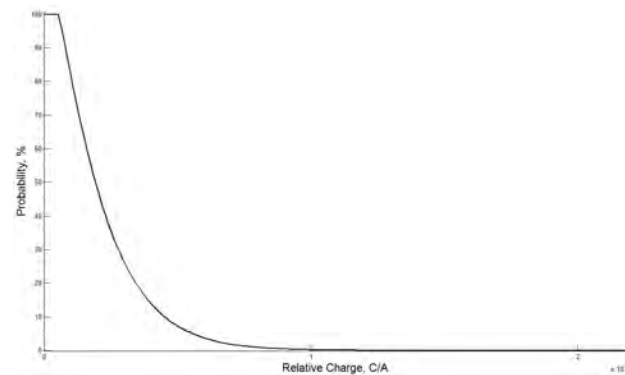


Fig. 3. Distribution Function

By using these results a pattern of disruptive impulses generated by lightning activity in certain geographic areas can be obtained. In Fig. 4. is presented a map showing the density of lightning activity worldwide, which was published in George M. Kauffman's [9] work about similar topic . A similar map is shown in the corporate material of SPINNER GmbH [10]. It shows the frequency of lightning falling, measured in number of lightnings per square kilometer per year. It is seen that the highest density regions are Central Africa, the Himalayas, the Caribbean and Indochina, where the density of lightning hits is in the range  $50 \div 70$  lightnings.km<sup>2</sup>.year<sup>-1</sup>. For Bulgaria, this density is in the range  $6 \div 10$  lightnings.km<sup>2</sup>.year<sup>-1</sup>.

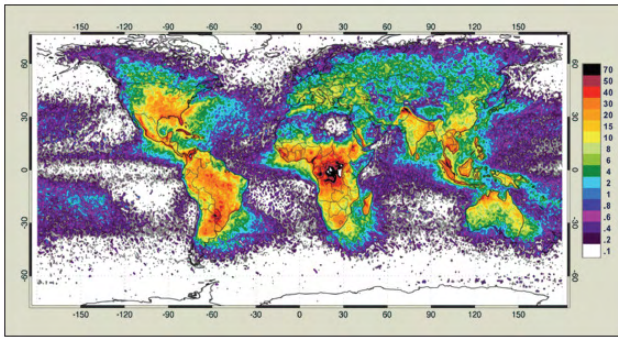


Fig. 4. Worlds Lightning Strike Density

Using the diagram of Fig. 3 the number of pulses with a relative charge greater than or equal to a value per square kilometer within one year can be determined. In Table III are given the relative charges of the three main types of disturbing impulses. The same table gives the average number of impulses exceeding that charge per square kilometer within one year and the average time between the occurrence of two pulses with exceeding relative charges per square kilometer. These values are calculated for Bulgaria.

TABLE III  
YEAR LIGHTNING ACTIVITY

	10/350 $\mu$ s	0,25/100 $\mu$ s	8/20 $\mu$ s
<b>Relative charge, <math>\times 10^{-4}</math> C/A</b>	4,399	1,326	0,166
<b>Number of impulses for 1 km<sup>2</sup> per year</b>	0,002	0,7	5,8
<b>Time between impulses for 1 km<sup>2</sup>, days</b>	182 500	521	62

It is noted that the values for the period of occurrence of the disturbing pulses is large. On one hand, due to the fact that the

probability of strong interference (e.g. direct hit by lightning) is significantly less than that of the weaker ones. On the other hand these Thunderstorms data have very general nature. It strongly depends on topography, soil and other factors and for more accurately determine the probabilities of occurrence of a specific impulse with relative charge more accurate statistics on lightning activity at the site should be used.

#### IV. CONCLUSION

From the foregoing it can be drawn:

- The mathematical model of high-voltage disturbing pulse described by dependence (2) can be used to generate random pulses to take into account stochastic character in nature;
- The probability distribution of occurrence of disturbing pulses to a particular relative charge meets the reality;
- It is appropriate to experimentally confirm the simulation results, and compared with actual statistics on lightning activity and its consequences.

#### REFERENCES

- [1] IEC 60060, High-voltage test techniques.
- [2] IEC 61312, Protection against LEMP.
- [3] Maceika K., Lightning Protection of Electronic Data Processing Systems, Scientific Proceedings of RTU. Series 7. Telecommunications and Electronics, 2003, vol.3.
- [4] Gamlin M., Impulse current testing, Lightning Protection Forum Shanghai, June 2004.
- [5] Hasse P., Overvoltage protection of low voltage systems, The Institution of Electrical Engineers, London, 2000.
- [6] Каменов О., Висша математика, Част I, Издателство "Сиела", София, 2001.
- [7] Angelov K., Model of Highvoltage Interferential Impulses, International Scientific Conference UNITECH 2010, Proceedings vol. 1 (pp 286), 19-20 November 2010, Technical University Of Gabrovo.
- [8] MATLAB IMAGE PROCESSING TOOLBOX. User's Guide, "The Math - Works Inc.", 2000. [www.mathworks.com](http://www.mathworks.com).
- [9] Kauffman G.M., Using commercial lightning protectors in defense applications, RF Design Magazine, 2005, pp.16-21, [www.rfdesign.com](http://www.rfdesign.com).
- [10] Spinner GmbH, Muenchen, Deutschland, [www.spinner.de](http://www.spinner.de).

# Optimum Divergence of the Transmitter Optical Radiation in FSO Systems

Tsvetan Mitsev<sup>1</sup>, Nikolai Kolev<sup>2</sup>, Hristo Ivanov<sup>3</sup> and Kalin Dimitrov<sup>4</sup>

**Abstract** – The determination of the optimum divergence of the transmitter optical beam  $\theta_{t,opt}$  in FSO systems can largely compensate for the negative impact of the change in the direction of propagation of the optical radiation due to various random factors. Depending on the system parameters, the length of the communication channel and the typical weather working conditions, the proper choice of  $\theta_t$  can significantly increase the reliability of information transmission and reduce the probability of outage. In this paper the influence of the optical output power of the transmitter and the length of the communication channel on the value of optimum divergence of the laser beam after the transmitting antenna are shown. When the divergence of the transmitter beam is set, the FSO system of TU-Sofia can work reliably under conditions where the deviations of the beam from its main direction exceed more than twice the deviations in the absence of adjustment.

**Keywords** – Communications, Free Space Optics, Laser Beam, Diverging Angle, Beam Wander

## I. INTRODUCTION

The application of FSO systems is becoming more and more frequent with specific connection conditions in the contemporary communication systems and networks. This is due to their wide bandwidth, tight radiation pattern of antenna, small size and weight, lower price, license free frequency band, that is, no frequency planning is necessary. The increased interest in FSO systems, however, creates new requirements for improvement of their characteristics, as well as for optimizing some of their parameters, in particular those of the divergence of the transmitter optical radiation [1-4].

One of the reasons for decreasing the functioning reliability of FSO systems are the random angle fluctuations of the transmitter laser beam from the direction where the receiver is placed. The main reasons for their existence are the turbulent fluctuations in the atmosphere and the mechanical movements of the bases on which the transmitter/receiver sets are placed (or building sway) [5-7]. The phenomena mentioned have

coherent action in order to decrease the connection channel length or increase the outage in case of poor weather conditions. A typical way of overcoming this problem is the use of redundant power with a perfect optical setting of the system and when there is a possibility that there are only geometrical losses, which we have considered in our paper [1]. In it, we have derived an expression for calculating the maximum radial displacement of the receiver antenna center from the transmitter laser beam axis, depending on the initial Gaussian beam radius.

This paper is a continuation of [1]. We have proven the significance of a transmitter optical antenna with adjustable angle width of the transmitter diagram in order to increase the functioning reliability of the system. We have researched the impact of the optical radiation source power and the connection channel length on the value of the optimum divergence angle of the transmitter optical radiation. We have indicated the basic parameters of the system and the connection channel.

## II. OPTICAL PROPAGATION AND INTENSITY DISTRIBUTION. DEFINITION OF THE PROBLEM

In the selected location of the FSO system and a perfect optical setting, that is a coincidence of the optical antennae axes of the opposite transmitter/receiver sets, angle  $\theta = 0$ . The BER value with a perfect setting usually reaches values lower than  $10^{-20}$ , when the values for normal functioning of the FSO systems are within the range of  $10^{-12}$  to  $10^{-8}$ . This allows, when the source power remains the same, for an increase of the divergence of the transmitter  $\theta_t$ , and in this case there is an increase in the value of the maximum acceptable angle deviations  $\theta_{max}$  of the laser beam from its main direction when the condition is fulfilled that the received power  $\Phi_r$  is bigger than the threshold value  $\Phi_{r,min}$ , respectively the minimal average radiation intensity in the receiver aperture  $I_r$  is bigger than  $I_{r,min}$  (fig.1). With the further increasing of  $\theta_t$  we reach the maximum value of  $\theta_{max}$  when the installation and parameters of the system are fixed, and then  $\theta_{max}$  starts decreasing and we derive  $\Phi_r < \Phi_{r,min}$ , including the case where the angle is  $\theta = 0$ .

As it is evident fig.1, for the derivation of the optimum laser beam divergence of the transmitter  $\theta_{t,opt}$ , where on certain conditions we derive the maximum value of  $\theta_{max}$ , we need an intensity distribution model of the light of the source in the receiver antenna plane. This means that at a distance  $z$  from the transmitter in a plane transverse to the distribution with a assumption for azimuthally beam symmetry, we have to derive the radial distribution of the plane density of the power  $I(\rho, z) \equiv I(\theta, z)$ . This distribution depends mainly on the

<sup>1</sup>Tsvetan Mitsev is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd., Sofia 1000, Bulgaria, E-mail: mitzev@tu-sofia.bg.

<sup>2</sup>Nikolai Kolev is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd., Sofia 1000, Bulgaria.

<sup>3</sup>Hristo Ivanov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd., Sofia 1000, Bulgaria.

<sup>4</sup>Kalin Dimitrov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd., Sofia 1000, Bulgaria, E-mail: kld@tu-sofia.bg.

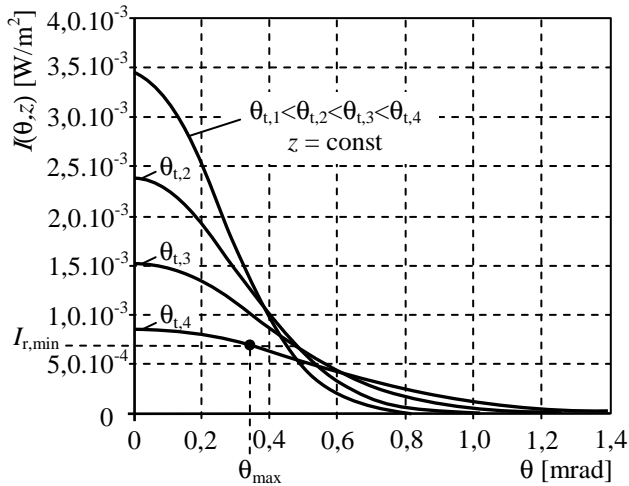


Fig.1. Dependence of the radial distribution of intensity of optical radiation  $I(\theta, z)$  in the plane of the receiver ( $z = \text{const}$ ) at different divergences of the transmitter optical radiation  $\theta_t$  ( $\theta_{\text{max}}$  is the maximum acceptable angular variance of transmitter's beam from its main direction in case of  $\theta_{t,4}$ ).

phase and amplitude distribution of the field in the emitting aperture  $A_t = \pi R_t^2$ .  $R_t$  is the aperture radius of the transmitting antenna. In our model we will use synchronous phase and Gaussian amplitude distribution in the emitting aperture [1], [8]. The maximum intensity value is along the beam axis, respectively in the center of the emitting aperture. When  $\rho = \rho_0$ , the light intensity decreases by  $e^2$  in relation to the maximum and  $\rho_0$  is defined as an initial Gaussian beam radius. In order to keep the Gaussian radial distribution in the Fraunhofer zone

$$z \geq z_{c, \text{exp}} \tag{1}$$

it is necessary to fulfill the condition

$$R_t \geq 2\rho_0. \tag{2}$$

$z_{c, \text{exp}}$  is calculated by the formula

$$z_{c, \text{exp}} = \frac{10\rho_0^2}{K_\theta \lambda_0}, \tag{3}$$

where  $\lambda_0$  is the central wavelength of the light source,

$$\rho_0 = \frac{K_\theta \lambda_0}{\pi\theta_{t, \text{exp}}}, \tag{4}$$

and  $K_\theta$  is the coefficient indicating the random fluctuations of the field phase in  $A_t$ . These fluctuations are due to different stochastic factors in the laser generation, and these factors worsen the radiation coherence level and lead to a difference between the actual divergence and the theoretically defined one  $\theta_{t, \text{teor}}$ ,

$$\theta_{t, \text{exp}} = K_\theta \theta_{t, \text{teor}}, \tag{5}$$

and typically  $K_\theta \geq 10$ .

When the conditions (1) and (2) are fulfilled, the current Gaussian radius is calculated by the formula

$$\rho_z(z) = \sqrt{\rho_0^2 + (\theta_{t, \text{exp}} z)^2} \tag{6}$$

and

$$I(\rho_z, z) = I(0, z) \cdot e^{-2}$$

is fulfilled.

The losses in the light distribution between the transmitter and the receiver when there is a assumption of an uniform volume extinction coefficient  $\alpha_e$ , are calculated by the formula

$$\tau_a = e^{-\alpha_e z}, \tag{7}$$

$$\alpha_e [\text{km}^{-1}] = \frac{3,92}{S_M [\text{km}]} \left( \frac{\lambda [\mu\text{m}]}{0,55} \right)^{-q}.$$

In (7)  $\tau_a$  is the transparence of the connection channel,  $S_M$  is the meteorological visibility of the atmosphere, and for the typical atmospheric conditions the exponent  $q$  is calculated by the formula

$$q = 0,585 \sqrt[3]{S_M [\text{km}]}.$$

With the assumptions made, the optical radiation intensity along the optical axis and its radial distribution are

$$I(0, z) = \frac{2\tau_t \tau_a (\lambda_0, S_M, z) \Phi_L}{\pi \rho_z^2(z)}, \tag{8}$$

$$I(\rho, z) \equiv I(\theta, z) = I(0, z) \cdot e^{-2 \left( \frac{\theta}{\theta_{t, \text{exp}}} \right)^2}.$$

In (8) the losses in the transmitter antenna optics have been expressed by  $\tau_t$ ,  $\Phi_L$  is the power of the source radiation,  $\theta$  is the angle deviation of the transmitter optical beam axis, recognized from the case of perfect alignment, that is coincidence of the optical axes of the transmitter and opposite receiver optical antenna. The power  $\Phi_L$  is the laser power with a assumption for a digital communication system with On/Off modulation (OOK) in the optical code impulse.

With the digital communication systems with OOK modulation, the system functioning quality is guaranteed by the low values of BER. With them, we calculate BER from SNR again using an erfc function, which presupposes a great slope in the changing of BER. A change by one order of SNR leads to a change up to ten orders of BER. Because of that it is more convenient to deal with and to represent graphically the change of SNR from the different parameters of the system. To calculate SNR we need the optical beams at the input of the receiver. The optical beam through the input aperture of the receiver corresponding to the upper level of the optical code impulse, is

$$\Phi_{pd}(\theta, z) = \pi \cdot \tau_r \cdot R_r^2 \cdot I(\theta, z). \quad (9)$$

In (9)  $R_r$  is the aperture radius of the receiver telescope,  $\tau_r$  is the transmission coefficient of the optical receiver system. The above equation is true when the condition  $\rho_z(z) \gg R_r$  is fulfilled.

The second in significance input optical beam, that is the background one, is calculated by the formula

$$\Phi_B = \pi^2 \tau_r L_{\lambda,B}(\lambda_0) R_r^2 \theta_{r,exp}^2 \Delta\lambda_F, \quad (10)$$

where  $L_{\lambda,B}$  is the spectral brightness of the background radiation, and  $\Delta\lambda_F$  is the transmission wavelength bandwidth of the interference filter before the photodetector, placed to restrict the background radiation.

With the indication of the dispersion of the two main types of noise in the optical receivers, the thermal and the quantum one, the expression for SNR calculation is

$$SNR = \frac{R_I(\lambda_0) \Phi_{pd}(\theta, z)}{\sqrt{C_I \left\{ \frac{2k_B T \cdot A}{R_{Fb}} + e R_I(\lambda_0) [\Phi_{pd}(\theta, z) + \Phi_B] \right\}}}. \quad (11)$$

The formula is true for an optical receiver with preamplification and a p-i-n photodiode.

$R_I(\lambda_0) = 8,06 \cdot 10^5 \eta(\lambda_0) \lambda_0$  is the integral sensitivity for current of the photodetector,  $\eta(\lambda_0)$  is the quantum efficiency of the photodetector material,  $k_B$  is the Boltzmann constant,  $e$  is the charge of the electron,  $C_I$  is the information throughput of the digital communication system, and  $R_{Fb}$  is the value of the resistor in the feedback of the preamplifier.

### III. SIMULATION RESULTS AND DISCUSSIONS

For the developed and implemented in TU-Sofia FSO system [9], [10] we will determine the maximum divergence  $\theta_{t,opt}$  of the transmitter optical beam. The system works at a wavelength  $\lambda_0 = 850$  nm with information throughput  $C_I = 100$  Mbps with power in the optical bit impulse  $\Phi_L = 10$  mW. Using a two-lens Kepler collimator, we gradually change the beam divergence within the range of 1 mrad to 5 mrad. The connection channel length is up to 2 km. The other system parameters necessary for the calculation using the method developed in II, are:  $\tau_t = 0,85$ ;  $K_\theta = 10$ ;  $R_r = 5,5$  cm;  $\theta_r = 5$  mrad;  $\tau_r = 0,85$ ;  $\eta(\lambda_0) = 0,7$ ;  $\Delta\lambda_F = 10$  nm;  $R_{Fb} = 1$  k $\Omega$ ;  $A = 5$ . For the calculations we choose values  $S_M = 10$  km,  $L_{\lambda,B} = 10^{-2}$  W/m<sup>2</sup>.sr.Å,  $T = 300$  K, and the constants are  $k_B = 1,38 \cdot 10^{-23}$  J/K,  $e = 1,602 \cdot 10^{-19}$  C.

In fig.2, with an increasing divergence  $\theta_{t,exp}$  of the transmitter beam, we have shown the dependence  $SNR(\theta)$ . It is evident that, when we choose a minimal level for the signal/noise ratio  $SNR_{th} = 11,2$ , which corresponds to  $BER \approx 10^{-8}$ , the maximum possible divergence of the beam  $\theta_{max}$  from the perfect alignment increases with the increasing of  $\theta_t$  in the

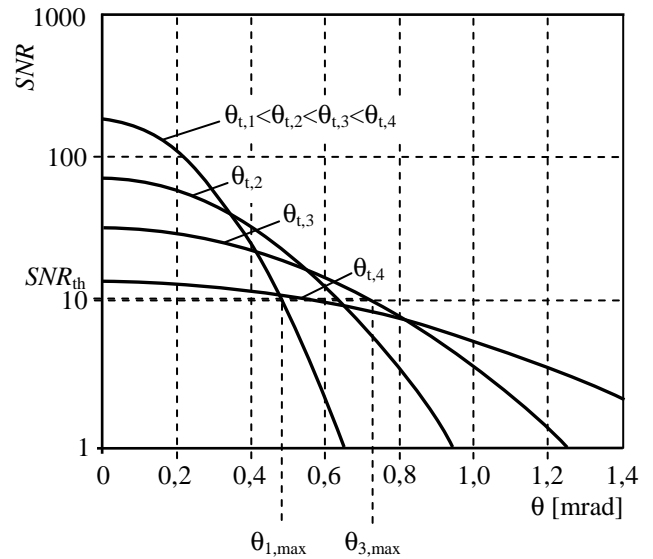


Fig.2. Dependence of the signal to noise ratio from angular deviation  $\theta$  of the beam of the transmitter from its main direction at different divergence of transmitter optical radiation  $\theta_t$ .

beginning, and then it starts decreasing, as we have already predicted.

The dependence  $\theta_{max}(\theta_t)$  for three values of the optical radiation source power  $\Phi_L = [10, 15, 20]$  mW with connection channel length  $z = 2$  km has been shown in fig.3. It is evident

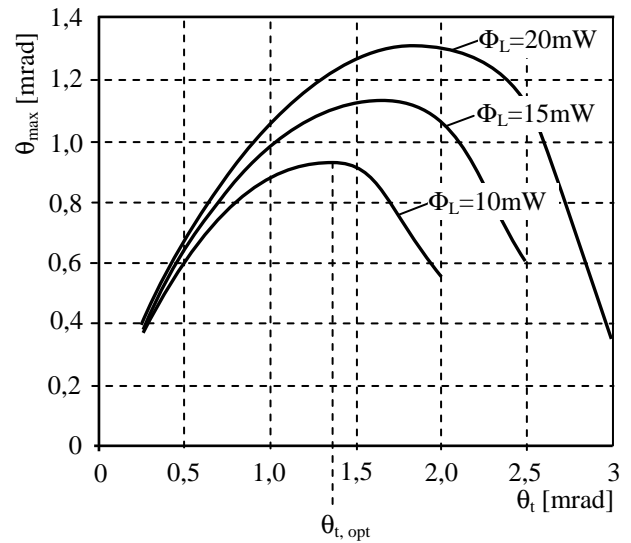


Fig.3. Dependence  $\theta_{max}(\theta_t)$  at  $z = 2$  km for three values of  $\Phi_L$ . Determination of  $\theta_{t,opt}$  ( $\Phi_L = 10$  mW).

that when the power  $\Phi_L$  increases,  $\theta_{max}$  increases, too. From the graphics it is evident that if we want to have the maximum value of  $\theta_{max}$ , it is necessary to change  $\theta_t$  too, this means that its optimum value exists and it is  $\theta_{t,opt}$ . When  $\Phi_L$  increases two times and with an optimum value of the transmitter optical beam divergence, the maximum possible angle beam divergence increases by 37%. It is also evident

from the graphics that  $\theta_{max}$ , depending on  $\Phi_L$ , undergoes more significant changes with the great values of  $\theta_t$ .

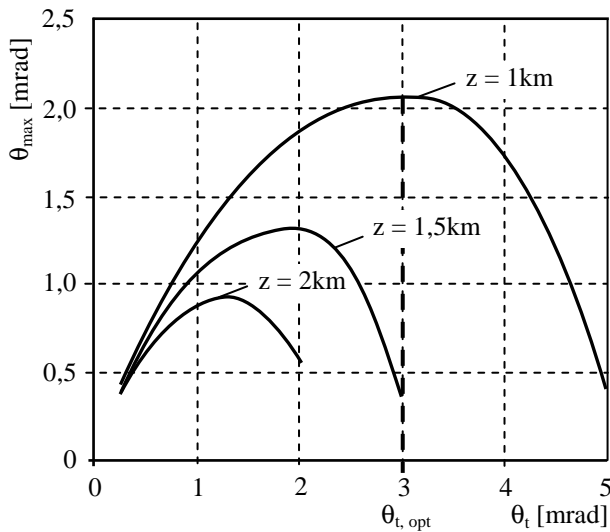


Fig.4. Dependence  $\theta_{max}(\theta_t)$  at  $\Phi_L = 10$  mW for three values of  $z$ . Determination of  $\theta_{t,opt}(z = 1$  km).

In fig.4 is shown the dependence  $\theta_{max}(\theta_t)$  for three connection channel lengths  $z = [1, 1.5, 2]$  km when the optical radiation source power is  $\Phi_L = 10$  mW. With the decreasing of the distance  $z$  is necessary a significant readjustment of the transmitter optical system, but as a result we can achieve a significant improvement of the functioning abilities of the system. When  $z$  is decreased 2 times, it is necessary to increase  $\theta_t$  by almost 3 times in order to maintain the optimum setting of the system. As a result, however, the possibilities of divergence of the beam from the main direction and keeping the functioning of the system, are more than 2,2 times greater.

In the comparisons between fig.3 and fig.4 it is evident that the functioning of the system is more sensitive to the change in the connection channel length that it is to the optical radiation source power. When the values of the optical radiation divergence are  $\theta_t < 1$  mrad, the impact of the changing of  $z$  or of  $\Phi_L$  on  $\theta_{max}$  can be ignored.

#### IV. CONCLUSION

This paper shows the possibility of a significant increase in the functioning and reliability of an FSO system with an optimum optical radiation divergence setting of the transmitter  $\theta_{t,opt}$ . Its value depends on the particular parameters of the system and the communication channel. We have researched the impact of the connection channel length  $z$  and the power in the code impulse of the optical radiation of the source  $\Phi_L$  on the maximum possible divergence  $\theta_{max}$  of the transmitter beam from the perfect direction, that is when there is a location on single optical axis of the opposite transmitter/receiver antennae  $\theta = 0$  (fig.2). We have shown

that the values  $\theta_{max}(\theta_{t,opt})$  increase when  $\Phi_L$  increases and they decrease when  $z$  increases, and the connection channel length  $z$  has a greater impact on them. When there is a constant collimation of the transmitter beam, that is a constant value  $\theta_t$ , the value of  $\theta_{max}$  is influenced to a much greater extent by  $z$  and  $\Phi_L$  when the values of  $\theta_t$  are big than when they are small, for instance when  $\theta_t \leq 1$  mrad. When there is an optimum beam divergence setting, within the limits of the research defined (III) it is possible to have a 121% increase in the acceptable value of the divergence  $\theta_{max}$ .

#### ACKNOWLEDGEMENT

This work has been supported by the Research Programme of the Technical University of Sofia, Bulgaria, Internal Projects 2011, contract Nb 112pd033-07.

#### REFERENCES

- [1] Ts. Mitsev, K. Dimitrov, B. Bonev, "Influence of Laser Beam Divergency on Free Space Optic Systems Functionality", TELECOM'2008, Conf. Proc., Varna, Bulgaria, 2008.
- [2] Zh. Zhao, R. Liao, Y. Zhang, "Impact of Laser Beam Diverging Angle on Free-Space Optical Communications", Aerospace Conference, IEEE, pp. 1-10, 2011.
- [3] A. Farid, S. Hranilovic, "Outage Capacity Optimization for Free-Space Optical Links with Pointing Errors", Journal of Lightwave Technology, Vol. 25, Issue 7, pp. 1702-1710, 2007.
- [4] Y. Ren, A. Dang, B. Luo, H. Guo, "Capacities for Long-Distance Free-Space Optics Links Under Beam Wander Effects", Photonics Technology Letters, IEEE, vol. 22, issue 14, pp. 1069-1071, 2010.
- [5] Shlomi Arnon, "Effects of Atmospheric Turbulence and Building Sway on Optical Wireless Communication Systems", Opt. Lett., vol. 28, No. 2, pp. 129-131, 2003.
- [6] E. Ferdinandov, B. Pachedjieva, B. Bonev, Sl. Saparev, "Joint Influence of Heterogeneous Stochastic Factors on Bit-Error Rate of Ground-to-Ground Free-Space Laser Communication Systems", Optics Communications, vol. 270, issue 2, pp. 121-127, 2007.
- [7] Bonev B., Relative Influence of Some Stochastic Factors on Bit-Error Rate of Ground-to-Ground Free Space Optics, ICEST 2007, Vol. 1, pp. 203-206, Ohrid, Macedonia, 2007.
- [8] E. Ferdinandov, B. Pachedjieva, K. Dimitrov, Optical Communication Systems, Sofia, Technika, 2007.
- [9] Ts. Mitsev, N. Kolev, K. Dimitrov, "Optical Wireless Communication System", XII-th International Scientific Conference SMOLJAN-2010, Smoljan, Bulgaria, 26-27 June 2010.
- [10] N. Kolev, "Selection of optimal settings depending on the FSO system parameters", XIII International PhD Workshop OWD 2011, Conf. Proc., vol. 29, pp. 467-472, Warsaw, 2011.



# Body Bias Influence on Ring Oscillator Performance for IR-UWB Pulse Generator in 0.18 $\mu\text{m}$ CMOS technology

Jelena Radic, Alena Djugova, Laszlo Nagy and Mirjana Videnovic-Misic

**Abstract** – A CMOS standard three-stage ring oscillator is examined in UMC 0.18 $\mu\text{m}$  technology. The ring oscillator performance dependence on the bulk (substrate) resistors introduced into inverter PMOS and NMOS transistors is investigated. Simulation results showed that the ring oscillator frequency is strongly dependent on the substrate resistor value. This fact can be used to increase the ring oscillator frequency. As the ring oscillator is a part of an IR-UWB (Impulse Radio Ultra Wide Band) pulse generator, its oscillating frequency determines the central frequency of the pulse spectrum and have significant effect on spectrum fitting within UWB FCC mask.

**Keywords** – Body biasing, bulk resistor, CMOS process, impulse radio ultra-wideband (IR-UWB), ring oscillator.

## I. INTRODUCTION

Impulse Radio Ultra-Wide-Band (IR-UWB) technology has emerged as a potential solution for very high data rate short-range communication, and low data rate communication related to localization, targeting both low cost and low power consumption [1] – [3]. It transmits extremely short pulses, on the order of a nanosecond or less, which occupy a bandwidth up to several GHz. Additionally, IR-UWB technology offers high fading margin for communication systems in multipath environments [3].

The American Federal Communications Commission (FCC) defines a signal as ultra-wideband if it occupies more than 500 MHz of radio frequency spectrum or exhibits a fractional bandwidth of at least 25% [4]. As the FCC allocated frequency spectrum for UWB technology is 3.1 – 10.6 GHz, the power level from the UWB transmitter should be small enough not to interfere with the already existing communication systems such as WiMax, Bluetooth and GSM. This requirement limits output power level of UWB TXs at  $-41.3$  dBm/MHz [4]. In the GPS band (0.96 – 1.61 GHz), there is even more stringent regulation: less than  $-75.3$  dBm/MHz is needed to avoid interference problems. The PSD (Power Spectral Density) in frequency interval from 1.61 GHz to 3.1 GHz depends on the type of application (indoor, outdoor, GPS, wall & medical imaging, through-wall imaging & surveillance system). In spite of these regulations, there have been many reports of interferences with wireless local area network (WLAN) systems operating in the 5 – 6 GHz band. Due to practical reasons, the UWB

Jelena Radic, Alena Djugova, Laszlo Nagy, and Mirjana Videnovic-Misic are with the Department of Power, Electronic and Telecommunication Engineering, Faculty of Technical Sciences, University of Novi Sad, Trg Dositeja Obradovica 6, Novi Sad 21000, Serbia, E-mail: {jelena\_r, alena\_d, lnadj, mirjam}@uns.ac.rs.

bandwidth is subdivided into two bands: 3 – 5 GHz (lower-band) and 6 – 10.6 GHz (higher-band).

One of the most critical components of an UWB system is the pulse generator, which has to be designed with relatively low-complexity and low power consumption. Moreover, generated pulse train spectrum has to satisfy the FCC spectral mask, making pulse generator design very challenging. There are several typical techniques for designing it which usually follow all-digital [5] – [6], analogue-digital [7] or all-analogue [7] – [8] design approach. Digital solutions offer higher integration, lower consumption and better controllability while all-analogue techniques demonstrate circuit simplicity.

As an essential part of an analogue-digital pulse generator [5], a ring oscillator is studied in this work. To increase ring oscillating frequency, resistors are introduced between the bulk (substrate) transistors terminals and appropriate voltage terminal (ground in case of NMOS transistor and  $V_{dd}$  in case of PMOS transistors). Dependence of the ring oscillator frequency on the value of the bulk resistors inserted in the NMOS, PMOS, and both kind of transistors is examined. The ring oscillator is designed and simulated in 0.18 $\mu\text{m}$  CMOS technology.

## II. STANDARD THREE-STAGE RING OSCILLATOR DESIGN

The pulse generator represents a key block in the impulse UWB communication. As the pulse shape determines the spectrum characteristic of the UWB signal and effectively dictates specific system requirements, its generation is one of the essential considerations in the UWB design. Fig. 1 shows the basic topology of an IR-UWB transmitter based on ring oscillator as a part of the pulse generator.

It consists of a glitch generator, a switched ring oscillator, a buffer stage and a pulse shaping (band-pass) filter [7]. The glitch generator turns the ring oscillator on/off approximately

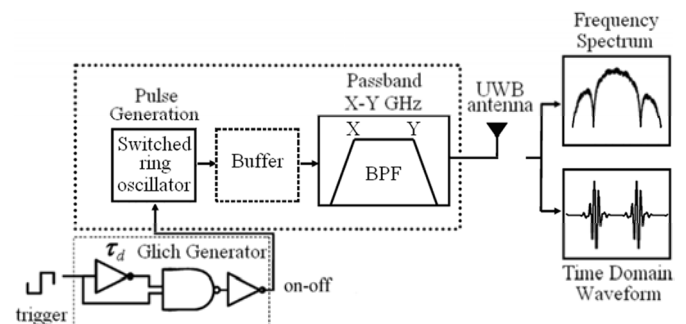


Fig. 1. An IR-UWB transmitter based on ring oscillator as a part of pulse generator.

defining the duration of its oscillation and thus the width of the pulse generated at the oscillator output. The switched ring oscillator frequency defines position of the transmitted pulse spectrum within the FCC mask [8]. Since in the IR-UWB communication the time domain pulse width roughly determines the width of the generated frequency spectrum [8], it is important to design a pulse, which makes optimal usage of the available spectrum, within the limits imposed by the FCC. The buffer isolates the ring oscillator from the pulse shaping filter loading and improves the pulse generator current driving capability. The band-pass filter additionally accommodates the pulse in the allowed FCC spectral mask.

The switched ring oscillator topology is shown in Fig. 2. It is composed of the three-stage ring oscillator ( $M_1 - M_3$  inverters stages) and a pair of oscillation-enabling switches (transistors  $M_4$  and  $M_5$ ). Due to its simplicity and short start-up time, the ring oscillator is the most used architecture in the IR-UWB transmitter applications. It has small resistance at each feedback nod which allows fast transient response.

The oscillation-enabling switches, as their name says, control the oscillation process. When the *on-off* signal (produced by the glitch generator) is high,  $M_4$  is turned on ( $M_5$  is turned off), the inverters stages  $M_1 - M_3$  outputs have initial voltage values determined by the size ratio of the corresponding PMOS and NMOS transistors. Due to the small inverter reactance, the oscillation can start immediately. Transistor  $M_5$  is turned on ( $M_4$  is turned off) at the *on-off* signal low level, connecting the  $M_1$  stage output (the  $M_2$  stage input) to  $V_{dd}$ , and effectively shutting down the oscillations. As the ring oscillator is switched off by  $M_4$  transistor during the inactive period of time, the power consumption is minimized.

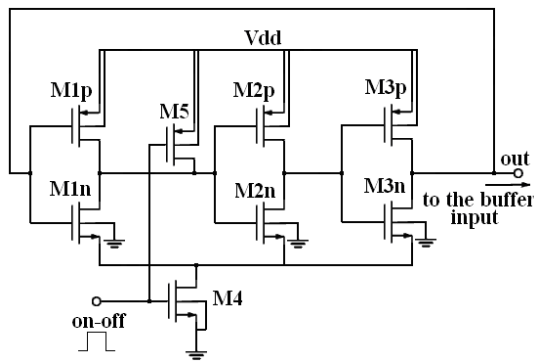


Fig. 2. The three-stage ring oscillator design.

### III. THE RING OSCILLATOR PERFORMANCE

The proposed designs have been simulated in mixed mode/RF 0.18 $\mu$ m CMOS technology using SpectreRF Simulator from Cadence Design System. Supply voltage  $V_{dd}$  of this technology is 1.8 V. The main problem with the standard ring oscillator design was limited set of transistor sizes available in the used technology as the ring oscillator working frequency depends directly on transistors sizes. If the transistors are larger, the period of the oscillation  $T$  rises

proportionally, while the oscillating frequency decreases ( $f_0=1/T$ ), and vice versa. For the smallest NMOS and PMOS transistors (transistor width/length:  $W/L=25\mu\text{m}/0.18\mu\text{m}$ ), the oscillation frequency of 3.95 GHz has been obtained. To utilize the whole UWB band more effectively, the center frequency of at least 6 GHz is required. Higher ring oscillator frequency could be achieved without PMOS transistor  $M_5$ . However, this transistor provides the identical oscillation start from the initial state, defined by connecting  $A'$  ( $B$ ) node to  $V_{dd}$  at the falling edge of the *on-off* signal.

#### A. Influence of the NMOS and the PMOS transistors bulk resistors

The bulk (substrate) terminals of the transistors are floated to improve the performance of CMOS SPST (single pole single throw) and SPDT (single pole double throw) switches, especially the power-handling ability and insertion loss, in [9] – [10], respectively. The series transistor particularly played an important role in the switch's insertion loss, while the shunt transistors enhanced the isolation when the switch was turned off. To improve the ring oscillator performance, the method that uses the resistors placed in the inverter stages transistors bulk terminals is investigated in the paper. First, the bulk resistors connected to ground are inserted in the ring oscillator inverter NMOS transistors, shown in Fig. 3.

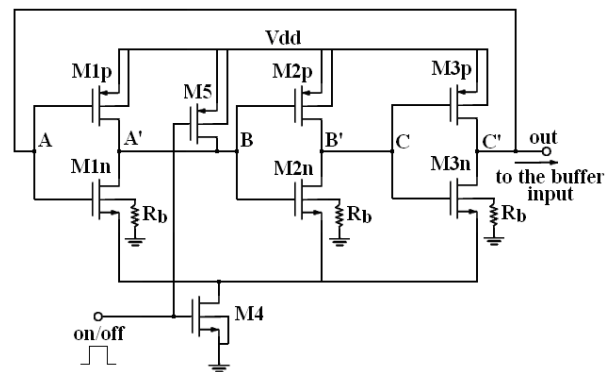


Fig. 3. The ring oscillator with NMOS bulk resistors architecture.

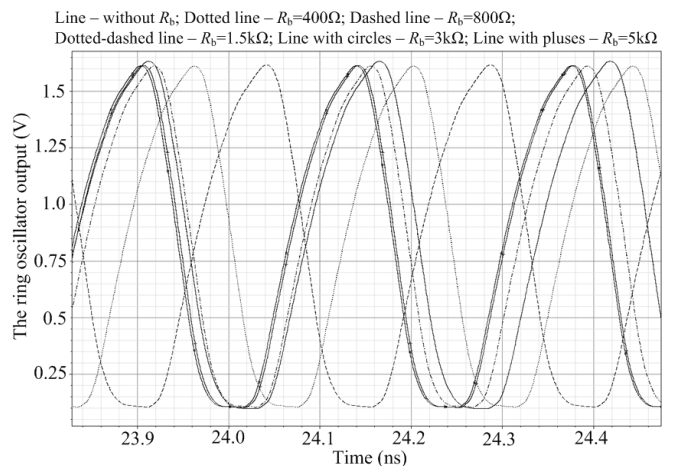


Fig. 4. Dependence of the ring oscillator performance on NMOS transistors substrate resistors values.

Dependence of the ring oscillator performance on the NMOS transistor bulk resistors  $R_b$  value is depicted in Fig. 4. It can be noticed that with resistor  $R_b$  value increase, the ring oscillator frequency  $f_0$  rises, while the peak-to-peak amplitude at the ring oscillator output changes negligibly. This can be explained by the two effects. First cause, describing the changes in the NMOS transistor threshold voltage  $V_{IN}$  by change in the bulk-to-source voltage  $V_{BS}$ , is of less concern. This phenomenon is sometimes called the “back-gate effect”, since the body influences the threshold voltage when it is not tied to the source, being considered as the second or the “back-gate”. The NMOS body effect upon threshold voltage  $V_{IN}$  is given by [11]

$$V_{IN} = V_{t0} + \gamma \left( \sqrt{2\phi_F - V_{BS}} - \sqrt{2\phi_F} \right), \quad (1)$$

where  $V_{t0}$  is threshold voltage for zero  $V_{BS}$  voltage,  $\phi_F$  is the Fermi level deep in the bulk,  $\gamma = \left( t_{ox} / \epsilon_{ox} \right) \sqrt{2q\epsilon_{si}N_A}$  is the body effect parameter,  $t_{ox}$  is oxide thickness,  $\epsilon_{ox}$  is oxide permittivity,  $\epsilon_{si}$  is silicon permittivity,  $N_A$  is a doping concentration, and  $q$  is the charge of an electron. With increase in the NMOS transistor bulk resistors value, due to the small substrate current the NMOS bulk voltage becomes slightly positive with respect to the source (the NMOS body-source junction is forward biased), resulting in a low  $V_{IN}$  voltage drop. This leads to increase in the current discharging the ring oscillator output and internal node capacitances. As a result, the period of reaching the peak values is decreased and the ring oscillator frequency is increased. Dominant effect represents reduction in the bulk parasitic capacitance current caused by increase in the substrate resistors value. With decrease in the bulk parasitic currents, more current is available for the inverter output capacitance discharging and thus the  $f_0$  parameter rises.

In the second case, the resistors tied to  $V_{dd}$  are introduced as the PMOS transistor substrate resistors, Fig. 5. Since the influence on the ring oscillator performance was nearly the same as presented in the simulation above (increase in the PMOS bulk resistors reduces the PMOS threshold voltage  $V_{IP}$  and/or decreases the substrate parasitic capacitance currents leading to increase in the current charging the ring oscillator output and internal node capacitances), the output voltage waveform are not given here. However, simulated  $f_0$  values for both approaches are summarized in Tables I, and presented in Fig. 6. Varying  $R_b$  from 0.4 k $\Omega$  to 5 k $\Omega$ , the

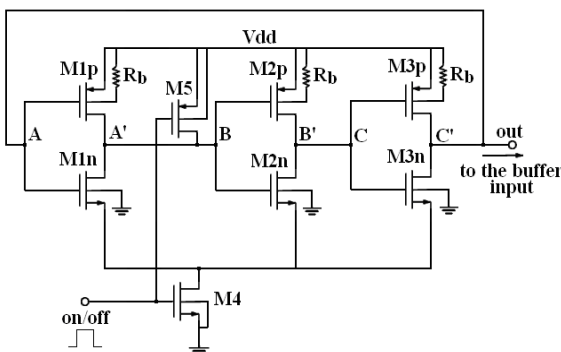


Fig. 5. The ring oscillator with PMOS bulk resistors architecture.

TABLE I  
INFLUENCE OF THE NMOS/PMOS TRANSISTORS BULK RESISTORS VALUE ON THE RING OSCILLATOR FREQUENCY

$R_b$ (k $\Omega$ )	Bulk resistors in NMOS transistors $f_0$ (GHz)	Bulk resistors in PMOS transistors $f_0$ (GHz)
0.4	4.10	4.0
0.8	4.15	4.10
1.5	4.20	4.20
3.0	4.25	4.25
5.0	4.25	4.25

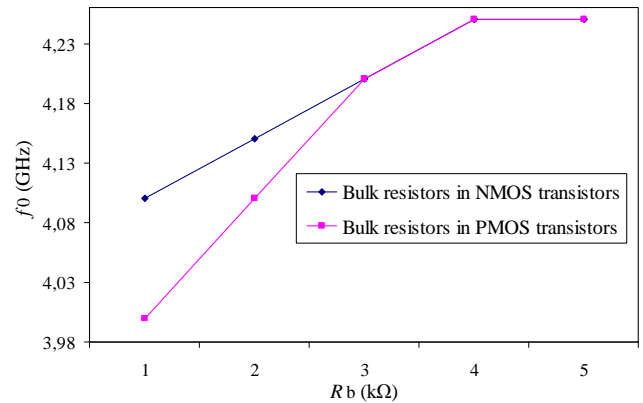


Fig. 6. Dependence of the ring oscillator frequency on the NMOS and PMOS transistors substrate resistors value.

oscillating frequency was increased from 4.1 GHz to 4.25 GHz in case of the NMOS substrate resistors, while  $f_0$  parameter was in the range from 4 GHz to 4.25 GHz for the PMOS bulk resistors method. It can be noticed that for lower substrate resistor values, the higher oscillation frequency was obtained in the former technique. Furthermore, the  $f_0 - R_b$  curves go into saturation for the bulk resistor values higher than 3 k $\Omega$  in both cases.

### B. Influence of the substrate resistors placed in each ring oscillator inverter stage transistor

To further increase the ring oscillator frequency the bulk resistors are introduced in each inverter transistor, shown in Fig. 7. Influence of the  $R_b$  value on the ring oscillator

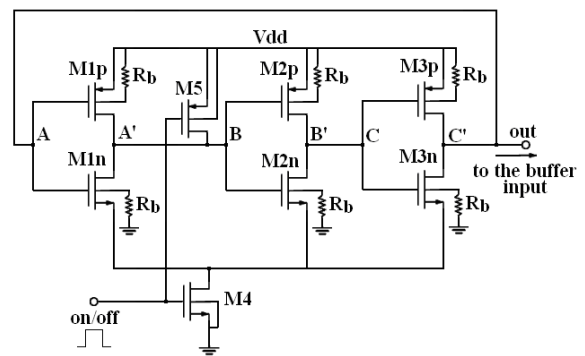


Fig. 7. The ring oscillator with bulk resistor in all inverter's transistors topology.

TABLE II  
INFLUENCE OF THE TRANSISTORS BULK RESISTORS VALUE ON THE RING OSCILLATOR FREQUENCY

$R_b$ (k $\Omega$ )	$f_0$ (GHz)
0.1	4.0
0.2	4.05
0.4	4.15
0.8	4.35
1.5	4.50
3.0	4.55
5.0	4.60

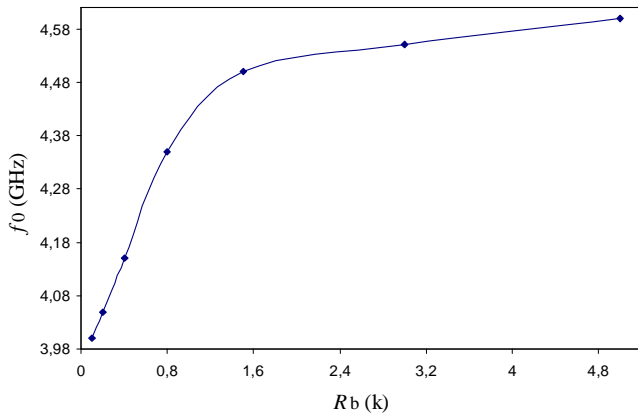


Fig. 8. Dependence of the ring oscillator frequency on the substrate resistors value.

frequency is presented in Table II, and shown in Fig. 8. The same  $f_0$  parameter dependence on the bulk resistor as in the previous simulations can be observed. Nevertheless, it should be emphasized that the achieved ring oscillator frequency is considerably higher comparing to the initial  $f_0$  value (3.95 GHz, obtained in the topology presented in Fig. 2), and the values achieved in previous approaches for the same  $R_b$  values, as expected.

#### IV. CONCLUSION

Standard three-stage ring oscillator topology has been analyzed in 0.18 $\mu$ m CMOS technology. Dependence of the ring oscillator performance on the bulk resistors inserted in the inverter NMOS, PMOS, and both kind of transistors has been investigated. Simulations confirmed strong dependency of the ring oscillator frequency  $f_0$  on the substrate resistor value. The maximum frequency (4.6 GHz) obtained in case the bulk resistors were introduced into each inverter stage transistors is remarkably (16.5 %) higher than value achieved with the standard ring oscillator architecture. The  $f_0$  parameter [12]

changes can be attributed to the two facts: the NMOS/PMOS transistor threshold voltage drop due to minor increase in the bulk voltage, and/or the parasitic currents reduction, with increase in the bulk resistor value. Both effects cause increase in currents available to charge/discharge the load capacitance and thus the ring oscillator frequency. However, the latter effect is much more dominant.

#### ACKNOWLEDGEMENT

This work was supported in part by the Ministry of Education and Science, Republic of Serbia, on the project number TR-32016.

#### REFERENCES

- [1] M. Ghavami, L. B. Michael, and R. Kohno, *Ultra Wideband Signals and Systems in Communications Engineering*, John Wiley&Sons Ltd, 2004.
- [2] K. Siwiak and D. McKeown, *Ultra-Wideband Radio Technology*, John Wiley&Sons Ltd, 2004.
- [3] J. R. Fernandes and D. Wentzloff, "Recent Advances in IR-UWB Transceivers: An Overview", IEEE Int. Conf. on Circuits and Systems, Conference Proceedings, pp. 3284–3287, Paris, France, 2010.
- [4] First Report and Order: Revision of Part 15 of the Commission's Rules Regarding Ultra-Wideband Transmission Systems Federal Communications Commission (FCC), ET Docket 98-153, Adopted February 14, 2002, Released Apr. 22, 2002.
- [5] V. V. Kulkarni, M. Muqsith, K. Niitsu, H. Ishikuro, T. Kuroda, "A 750 Mb/s, 12 pJ/b, 6-to-10 GHz CMOS IR-UWB transmitter with embedded on-chip antenna", IEEE Jour. of Solid. State Circuits, vol. 44, no. 2, pp. 394-403, 2009.
- [6] H. Kim, Y. Joo, S. Jung, "A tunable CMOS UWB pulse generator", IEEE Int. Conf. on Ultra-Wideband, Conference Proceedings, Waltham, MA, pp. 109-112, 2006.
- [7] S. Sim, D.W. Kim, S. Hong, "A CMOS UWB pulse generator for 6–10 GHz applications", IEEE Microwave and wireless components letters, vol. 19, no. 2, pp. 83-85, 2009.
- [8] O. Novak, C. Charles, "Low-power UWB pulse generators for biomedical implants", IEEE Int. Conf. on Ultra-Wideband, Conference Proceedings, pp. 778-782, Vancouver, BC, 2009.
- [9] R. Xu, Y. Jin, C. Nguyen, "Power-Efficient Switching-Based CMOS UWB Transmitters for UWB Communications and Radar Systems", IEEE Transaction on Microwave Theory and Techniques., vol. 54, no. 8, pp. 3271–3277, 2006.
- [10] M.C. Yeh, Z.M. Tsai, R.C. Liu, K.Y. Lin, Y.T. Chang, H. Wang, "Design and analysis for a miniature CMOS SPDT switch using body-floating technique to improve power performance", IEEE Transaction on Microwave Theory and Techniques., vol. 51, no. 1, pp. 31–39, 2005.
- [11] Thomas H. Lee, *The design of CMOS radio-frequency integrated circuits*, Cambridge University press, 2004.

# Comparative Performance Studies of Laboratory WPA IEEE 802.11b,g Point-to-Point Links

José A. R. Pacheco de Carvalho<sup>1</sup>,

Cláudia F. F. P. Ribeiro Pacheco<sup>2</sup>, Hugo Veiga<sup>3</sup>, António D. Reis<sup>4</sup>

**Abstract** –Wireless communications using microwaves are increasingly important, e.g. Wi-Fi. Performance is a very crucial issue, resulting in more reliable and efficient communications. Security is equally very important. Laboratory measurements are made about several performance aspects of Wi-Fi IEEE 802.11 b,g WPA point-to-point links. A contribution is given to performance evaluation of this technology under WPA encryption, using available equipments (DAP-1522 access points from D-Link and WPC600N adapters from Linksys). 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. Comparisons are made to corresponding results obtained for open links. Conclusions are drawn about the comparative performance of the links.

**Keywords** – WLAN, Wi-Fi, WPA Point-to-Point Links, IEEE 802.11b, IEEE 802.11g, Wireless Network Laboratory Performance.

## I. INTRODUCTION

Contactless communication techniques have been developed using mainly electromagnetic waves in several frequency ranges, propagating in the air. Wi-Fi and FSO, whose importance and utilization have been recognized and growing, are representative examples of wireless communications technologies.

Wi-Fi is a microwave based technology providing for versatility, mobility and favorable prices. The importance and utilization of Wi-Fi has been growing as it complements traditional wired networks. It has been used both in ad hoc mode and in infrastructure mode. In this case a WLAN arises

<sup>1</sup>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.

<sup>2</sup>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.

<sup>3</sup>Hugo Veiga is 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-mail: hveiga@ubi.pt.

<sup>4</sup>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.

based on an access point, AP, which permits communications of Wi-Fi electronic devices with a wired based LAN through a switch/router. Wi-Fi has penetrated the personal home, where a WPAN allows personal 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, 802.11g and 802.11n standards [1]. The 2.4 GHz band is intensively used and is having increasing interferences. Therefore considerable attention has been focused on the 5 GHz band where, however, absorption increases and ranges are shorter.

Nominal transfer rates up to 11 (802.11b), 54 Mbps (802.11 a, g) and 600 Mbps (802.11n) are specified. The medium access control is CSMA/CA. There are studies on wireless communications, wave propagation [2,3], practical implementations of WLANs [4], performance analysis of the effective transfer rate for 802.11b point-to-point links [5], 802.11b performance in crowded indoor environments [6].

Performance has been a very important issue, resulting in more reliable and efficient communications. In comparison to traditional applications, new telematic applications are specially sensitive to performances. Requirements have been pointed out [7]. E.g. requirements have been quoted 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 very important. Microwave radio signals can be very easily captured as they travel through the air. Therefore, several security methods have been developed to provide authentication such as, by increasing order of security, WEP, WPA and WPA2. WEP was initially intended to provide confidentiality comparable to that of a traditional wired network. A shared key for data encryption is involved. The communicating devices use the same key to encrypt and decrypt radio signals. The CRC32 checksum used in WEP does not provide a great protection. However, in spite of its weaknesses, WEP is still widely used in Wi-Fi communications for security reasons, mainly in point-to-point links. WPA implements the majority of the IEEE 802.11i standard [1]. It includes a MIC, message integrity check, replacing the CRC used in WEP. Either personal or enterprise modes can be used. In this latter case an 802.1x server is required. Both TKIP and AES cipher types are usable and a group key update time interval is specified.

Several performance measurements have been made for 2.4 and 5 GHz Wi-Fi open [8-9] and WEP links [10], as well as very high speed FSO [11]. It is important to find the effects of WPA encryption on link performance. Therefore, in the

present work new Wi-Fi (IEEE 802.11 b,g) results arise, using personal mode WPA, through OSI levels 4 and 7. Performance is evaluated in laboratory measurements of WPA point-to-point links using new available equipments. Comparisons are made to corresponding results obtained for open links.

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

The measurements used a D-Link DAP-1522 bridge/access point [12], with internal PIFA \*2 antenna, IEEE 802.11 a/b/g/n, firmware version 1.31 and a 100-Base-TX/10-Base-T Allied Telesis AT-8000S/16 level 2 switch [13]. The wireless mode was set to access point mode. The firmware from the manufacturer did not make possible a point-to-point link with a similar equipment. Therefore, a PC was used having a PCMCIA IEEE.802.11 a/b/g/n Linksys WPC600N wireless adapter with three internal antennas [14], to enable a PTP link to the access point. In every type of experiment, interference free communication channels were used (ch 8 for 802.11b,g). This was checked through a portable computer, equipped with a Wi-Fi 802.11 a/b/g/n adapter, running NetStumbler software [15]. WPA personal encryption was activated in the AP and the PC wireless adapter using AES and a shared key with 26 ASCII characters. The experiments were made under far-field conditions. No power levels above 30 mW (15 dBm) were required, as the access points were close.

A laboratory setup has been planned and implemented for the measurements, as shown in Fig. 1. At OSI level 4, measurements were made for TCP connections and UDP communications using Iperf software [16]. 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 determined. Parameterizations of TCP packets, UDP datagrams and window size were as in [10]. 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, representing the smooth mean of differences between consecutive transit times, was continuously computed by the server, as specified by the real time protocol RTP, in RFC 1889 [17]. The scheme of Fig. 1 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 server PC also permitted manual control of the settings in the access point.

The server and client PCs were HP nx9030 and nx9010 portable computers, respectively, running Windows XP. They were configured to optimize the resources allocated to the present work. Batch command files have been 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 official IP University network, via switch.

## III. RESULTS AND DISCUSSION

The access point and the PC wireless network adapter were manually configured, for each standard IEEE 802.11 b, g, with typical fixed transfer rates (1, 2, 5, 11 Mbps for 802.11b; 6, 9, 12, 18, 24, 36, 48, 54 Mbps for 802.11g). For every fixed transfer rate, data were obtained for comparison of the laboratory performance of the WPA and Open links at OSI layers 1 (physical layer), 4 (transport layer) and 7 (application layer) using the setup of Fig. 1. For each standard and 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.

At OSI level 1, noise levels (N, in dBm) and signal to noise ratios (SNR, in dB) were monitored and typical values are shown in Fig. 2 and Fig. 3, for open and WPA links, and 802.11b, g, respectively.

The main average TCP and UDP results are summarized in Table I, both for WPA and open links. In Fig. 4 polynomial fits were made to the 802.11b, g TCP throughput data for WPA links, where  $R^2$  is the coefficient of determination. It was found that the best TCP throughputs are for 802.11 g, for every link type. The 802.11 b, g average data are reasonably close for both link types. The best average 802.11g TCP throughput is for open links. In Figs. 5-7, the data points representing jitter and percentage datagram loss were joined by smoothed lines. Concerning jitter it was found that, on average, the best jitter performances are for 802.11 g for both link types. For each standard, jitter performances agree reasonably well within the experimental errors. However average jitter for 802.11 b is slightly higher for WPA (5.5+-0.2 ms) than for Open links (5.3+-0.3 ms), meaning that in this case increasing security leads to a minor degradation of jitter performance. Fig. 7 shows percentage datagram loss data. Except for 802.11 g, where the highest value is for WPA, no significant sensitivities were found for most data (1.4 % on average), within the experimental errors, either to standard or link type.

At OSI level 7 we measured FTP transfer rates versus nominal transfer rates configured in the access point and the PC wireless network adapter for IEEE 802.11 b, g as in [10]. The results show the same trends found for TCP throughput.

Generally, except for 802.11g TCP throughput, 802.11 b jitter and 802.11g percentage datagram loss, the results measured for WPA links were found to agree, within the experimental errors, with corresponding data obtained for Open links.

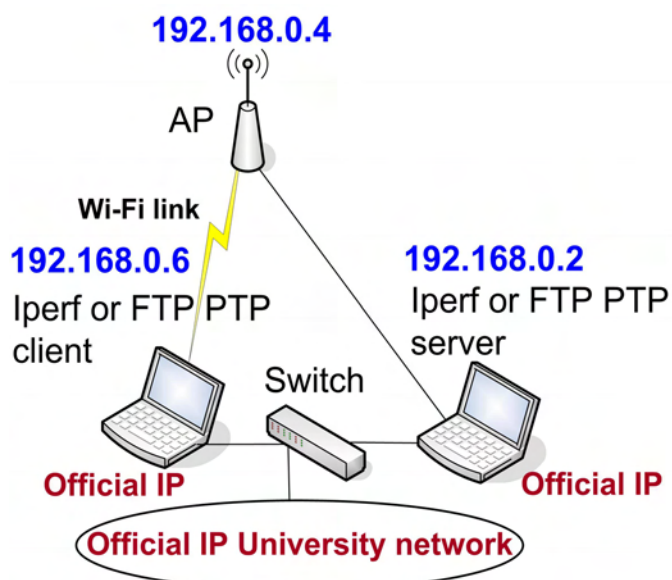


Fig. 1- Laboratory setup scheme.

TABLE I  
Average Wi-Fi (IEEE 802.11 b,g) results; WPA and Open links.

Link type	WPA		Open	
	802.11b	802.11g	802.11b	802.11g
Parameter/ IEEE standard				
TCP throughput (Mbps)	2.9 +/-0.1	13.4 +/-0.4	3.0 +/-0.1	14.5 +/-0.4
UDP-jitter (ms)	5.5 +/-0.2	2.3 +/-0.1	5.3 +/-0.3	2.3 +/-0.1
UDP-% datagram loss	1.2 +/-0.2	1.8 +/-0.2	1.2 +/-0.2	1.2 +/-0.1
FTP transfer rate (kbyte/s)	280.2 +/-11.2	1450.6 +/-58.0	289.9 +/-11.6	1526.9 +/-61.1

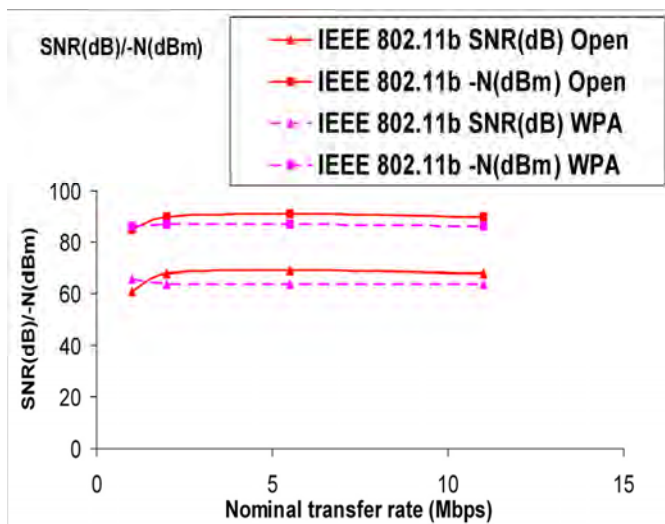


Fig. 2- Typical SNR (dB) and N (dBm); 802.11b; WPA and open links.

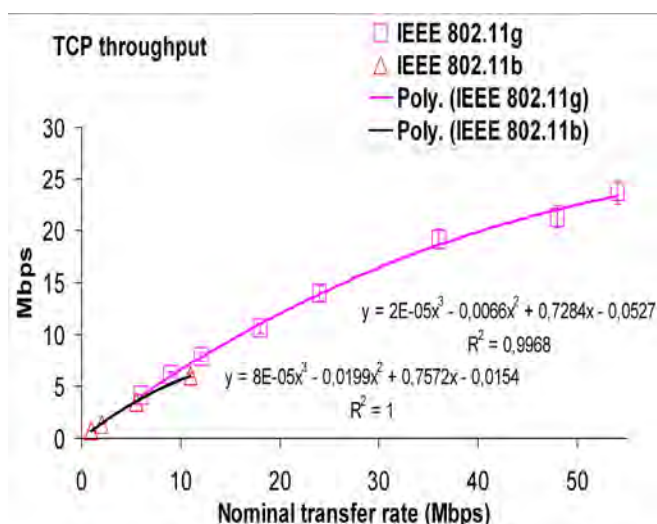


Fig. 4- TCP throughput versus technology and nominal transfer rate; WPA links.

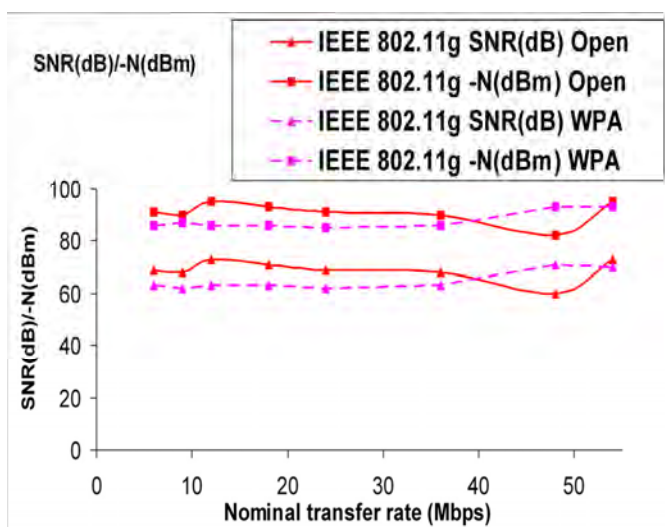


Fig. 3- Typical SNR (dB) and N (dBm); 802.11g; WPA and open links.

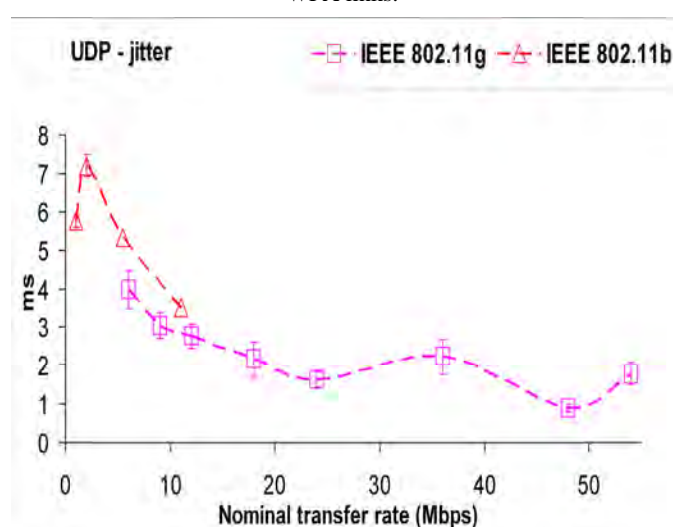


Fig. 5- UDP - jitter results versus technology and nominal transfer rate; WPA links.

ACKNOWLEDGEMENT

Supports from Universidade da Beira Interior and FCT (Fundação para a Ciência e a Tecnologia)/PEst-OE/FIS/UI0524/2011 (Projecto Estratégico-UI524-2011-2012) are acknowledged.

REFERENCES

- [1] Web site <http://standards.ieee.org> Web site; IEEE 802.11a, 802.11b, 802.11g, 802.11n, 802.11i standards.
- [2] J. W. Mark, W. Zhuang, *Wireless Communications and Networking*, Prentice-Hall, Inc., Upper Saddle River, NJ, 2003.
- [3] T. S. Rappaport, *Wireless Communications Principles and Practice*, 2nd ed., Prentice-Hall, Inc., Upper Saddle River, NJ, 2002.
- [4] W. R. Bruce III, R. Gilster, *Wireless LANs End to End*, Hungry Minds, Inc., NY, 2002.
- [5] M. Schwartz, *Mobile Wireless Communications*, Cambridge University Press, 2005.
- [6] N. Sarkar, K. Sowerby, "High Performance Measurements in the Crowded Office Environment: a Case Study", In *Proc. ICCT'06-International Conference on Communication Technology*, pp. 1-4, Guilin, China, 27-30 November 2006.
- [7] E. Monteiro, F. Boavida, *Engineering of Informatics Networks*, 4th ed., FCA-Editor of Informatics Ld., Lisbon, 2002.
- [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, pp. 409-422, 2008.
- [9] J. A. R. Pacheco de Carvalho, H. Veiga, P. A. J. Gomes, C. F. Ribeiro Pacheco, N. Marques, A. D. Reis, "Wi-Fi Point-to-Point Links- Performance Aspects of IEEE 802.11 a,b,g Laboratory Links", in *Electronic Engineering and Computing Technology, Series: Lecture Notes in Electrical Engineering*, Sio-Iong Ao, Len Gelman, Eds. Netherlands: Springer, 2010, Vol. 60, pp. 507-514.
- [10] J. A. R. Pacheco de Carvalho, H. Veiga, N. Marques, C. F. Ribeiro Pacheco, A. D. Reis, "Wi-Fi WEP Point-to-Point Links-Performance Studies of IEEE 802.11 a,b,g Laboratory Links", in *Electronic Engineering and Computing Technology, Series: Lecture Notes in Electrical Engineering*, Sio-Iong Ao, Len Gelman, Eds. Netherlands: Springer, 2011, Vol. 90, pp. 105-114.
- [11] J. A. R. Pacheco de Carvalho, N. Marques, H. Veiga, C. F. Ribeiro Pacheco, A. D. Reis, "Experimental Performance Evaluation of a Gbps FSO Link: a Case Study", *Proc. WINSYS 2010- International Conference on Wireless Information Networks and Systems*, pp. 123-128, Athens, Greece, 26-28 July, 2010.
- [12] Web site <http://www.dlink.com>; DAP-1522 wireless bridge/access point technical manual.
- [13] Web site <http://www.alliedtelesis.com>; AT-8000S/16 level 2 switch technical data.
- [14] Web site <http://www.linksys.com>; WPC600N notebook adapter user guide.
- [15] Web site <http://www.netstumbler.com>; NetStumbler software.
- [16] Web site <http://dast.nlanr.net>; Iperf software.
- [17] Network Working Group. "RFC 1889-RTP: A Transport Protocol for Real Time Applications", <http://www.rfc-archive.org>

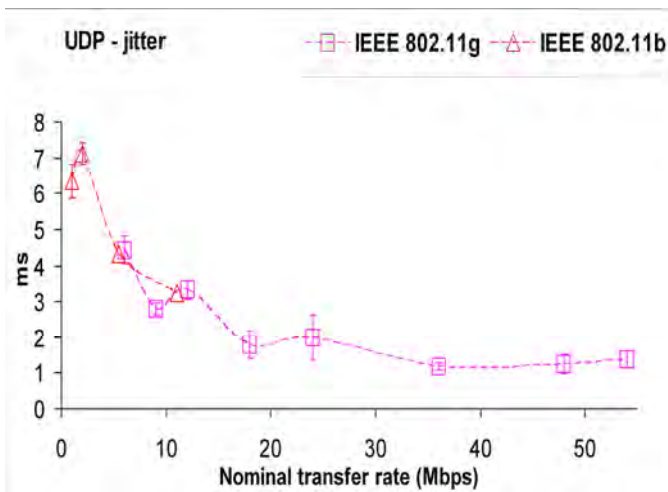


Fig. 6- UDP - jitter results versus technology and nominal transfer rate; open links.

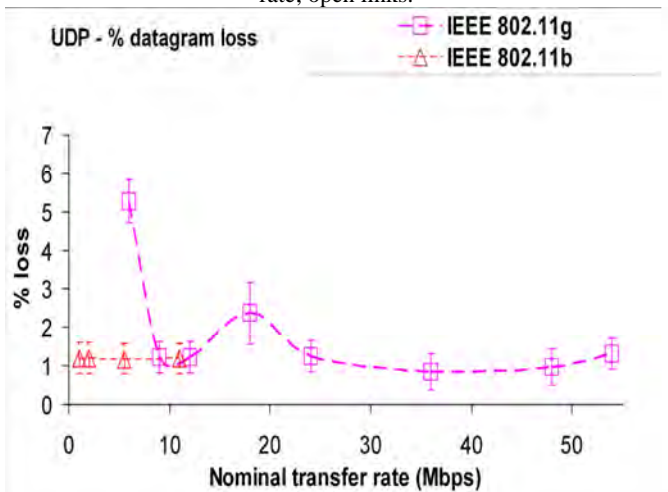


Fig. 7- UDP – percentage datagram loss versus technology and nominal transfer rate; WPA links.

IV. CONCLUSION

A new laboratory setup arrangement has been planned and implemented, that permitted systematic performance measurements of new available wireless equipments (DAP-1522 access points from D-Link and WPC600N adapters from Linksys) for Wi-Fi (IEEE 802.11 b,g) in WPA point-to-point links.

Through OSI layer 4, TCP throughput, jitter and percentage datagram loss were measured and compared for WPA and open links. Generally, except for 802.11g TCP throughput, 802.11 b jitter and 802.11g percentage datagram loss, where increasing security encryption was found to degrade performances, the results measured for WPA links were found to agree, within the experimental errors, with corresponding data obtained for Open links.

At OSI layer 7, FTP performance results have shown the same trends found for TCP throughput.

Additional performance measurements either started or are planned using several equipments, security settings and experimental conditions, not only in laboratory but also in outdoor environments involving, mainly, medium range links.



# Customer Satisfaction based Demand Analysis of Mobile Services

Aleksandar Tsenov<sup>1</sup>

**Abstract** – The rapid technology growth of the mobile networks has led to the situation where the competition for each one customer became a great importance. The customer satisfaction is increasingly attended to the quality, to the performance and to the usefulness of the services. These are in nature different characteristics and their complex evaluation requires implementation of new methods and tools such as fuzzy logic and fuzzy evaluation schemes. The paper proposes an approach for analyzing the customer demand on mobile services according their satisfaction with these services. The demand analysis is object of fuzzy evaluation approach based on the customer value hierarchy.

**Keywords** – Customer Satisfaction, Fuzzy Evaluation, Mobile Services, Significant Attribute Definition.

## I. INTRODUCTION

There are many known attempts for creating a methods and models for meeting the improved customer requirements according the service quality. All of them requires considerable preparation before the evaluation schemes are implemented. The first step of the whole evaluation process is the definition of an appropriate customer experience model in order to find out the customer's goal and purpose, the desired consequences in use situation and the desired products/ services attributes and performances [1]. In the same work an exemplary "mobile customer value hierarchy" was defined. In [2] the author introduces an overall demonstration of the applicability of fuzzy evaluation method for Service Level Management metrics.

There are not many similar researches in the field of the customer centric evaluation. In [3] the authors use the fuzzy similarity approach for clustering the QoS opinions for Web services. In [4] a fuzzy oriented approach for clustering of the services attributes and for definition of the most significant of these attributes is introduced. The research is based on the customer demand in personalized services. In [5] the authors apply fuzzy evaluation of SLA (Service Level Agreement) oriented quality metrics in NGN. All mentioned research works are based on hypothetical values, and not on real data and are only used for proving the applicability of the fuzzy logic in such complex evaluation problems.

This work is an extension of [6]. In the work mentioned, the authors attempt to apply the fuzzy evaluation approach to the customer experience hierarchy in order to evaluate the

customer satisfaction with a defined number of services. The experiment included a wide study over three groups of mobile services users, conducted at the Technical University of Sofia. The three target groups were the university teachers, the administration and students. These groups are being chosen because of the differences in age, in job, in activity of service use etc.

## II. CUSTOMER VALUE HIERARCHY AND ATTRIBUTES DEFINITION

### A. The mobile customer value hierarchy

Based on the complete chain of customer value layers, the first step in the procedure is to shift the layers from the individual perspective to the aggregate perspective of a group of customers. Based on the mobile customer investigation, an model for constructing the mobile customer value hierarchy is introduced [1].

### B. Attributes definition

In order to apply the fuzzy evaluation of the customer satisfaction the following investigation where performed: Three different groups of mobile users where asked about the services mentioned above – teachers, administration staff and students. Each group covers 100 people. The questionnaire includes four questions according each of the services:

1. Do You use the service .....
2. How could You evaluate the service?
3. Could You evaluate service parameters?

The answers on the questions 1 and 3 are "Yes" and "No". The answers on the question 2 include different evaluation levels. The evaluation is defined with 5 evaluation levels – from 1 to 5. Table I shows the results of the answers on the first question and the corresponding coding for each service. The values give the number of the group members which are engaged with the corresponding service.

In [6] these definitions were used for calculating the customer satisfaction score - S. For the three groups being studied it was obtained:

**Satisfaction Score S (Teachers) = 3.0333;**  
**Satisfaction Score S (Administration) = 2,994157;**  
**Satisfaction Score S (Students) = 3,005901.**

That means – the group "Teachers" has the higher satisfaction score with the services have being studied.

<sup>1</sup>Aleksandar Tsenov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: akz@tu-sofia.bg.

TABLE I  
EVALUATION METRICS SERVICE USAGE

No.	Attributes	Votes	Code
1	Voice mail box	Teachers – 57 Administration – 76 Students - 87	S1
2	SMS	Teachers – 78 Administration – 89 Students - 92	S2
3	Voice call	Teachers – 100 Administration – 100 Students - 100	S3
4	Conference call	Teachers – 14 Administration – 2 Students - 34	S4
5	Routine Services	Teachers – 82 Administration – 91 Students - 89	S5
6	E-Bank	Teachers – 8 Administration – 11 Students - 13	S6
7	Data Service	Teachers – 23 Administration – 8 Students - 47	S7
8	Mobile Purchase	Teachers – 26 Administration – 18 Students - 55	S8
9	Internet browsing	Teachers – 42 Administration – 32 Students - 70	S9
10	Mobile movie	Teachers – 4 Administration – 28 Students - 65	S10
11	Mobile games	Teachers – 4 Administration – 12 Students - 75	S11
12	MMS	Teachers – 34 Administration – 52 Students - 78	S12
13	Travel Info	Teachers – 5 Administration – 11 Students - 23	S13
14	Entertainment Info	Teachers – 6 Administration – 14 Students - 19	S14
15	Finance Info	Teachers – 9 Administration – 32 Students - 11	S15

In this work we will follow other way – further decomposition of the attributes defined above into most important quality parameters of each one service, represented as attribute in the value hierarchy. The number of parameters for each service is limited on 5, but it can be higher or respectively lower.

Table II represents 5, randomly chosen services and their quality characteristics. We choose only 5 services in order to make the work more understandable and clear. We suppose that when the approach is proven with smaller number of services, it will become applicable for great amount of services too. Each characteristic is presented with unique code.

TABLE II  
SERVICE QUALITY PARAMETERS

No.	Objectives	Attributes	Code
1	Voice mail box (S1)	Message duration	S11
		Number of messages to be stored	S12
		Number of rings before start	S13
		VoiceMailBox	S14
		Broken messages	S15
2	SMS (S2)	Speech quality	S21
		Message length	S22
		Message validity	S23
		Number of messages to be stored	S24
		Multi-user send message	S25
3	Voice call (S3)	Service coverage	S31
		Speech quality	S32
		Second voice call during conversation	S33
4	Conference call (S4)	Service coverage	S41
		Speech quality	S42
		Number of participants	S43
		Service control	S44
5	Data Service (S7)	Down speed	S71
		Up speed	S72
		Lost data	S73
		Service coverage	S74

Table 2 is produced with applying the proposed in this work Attribute – objective map. This means: uncover the relationship between quality parameters that the customers engaged (attributes layer) and the used services (objective layer). That approach enables the mobile provider to identify the customer’s goal from their consume history. So we reduce the layers in the customer value hierarchy to an attribute – objective map. We discovered attributes that are equal to more than one objective – for example: Speech quality.

So we obtained 21 service quality parameters. The main goal of the following analysis is to find out the most important parameters that are significant not only for the corresponding service but for the overall quality grade of the delivered services too.

C. Significant Attributes Analysis

The significant attributes of customer value hierarchy are the key attribute variables of the attribute layer which distinctly correlate to the objective layer. Because of the large numbers of mobile telecommunication products/services and the relatively small percentage of the mobile services/products engagement, the original data of customer value hierarchy is high dimensional sparse feature data. This paper adopts the fuzzy cluster analysis method [7] to find the significant attribute.

According to the rough set theory, data of the customer value objective layer and attribute layer can be defined as  $S=(U, A, V, f)$ . Here:  $U=\{u_1, u_2, \dots, u_n\}$ : the set of customers where n is the total number of customers.  $A=\{a_1, a_2, \dots, a_m\}$ : the set of variables of the objective layer and of the attribute layer.  $A = C \cup D$ , where C is the characteristics set of the attribute layer, and D is the characteristics set of the objective

layer.  $V$  is the set of the customer attribute parameters. The value of  $f(u_j, a_i)$  indicates the value of  $u_j$  about  $a_i$ .

The significant attributes analysis is solved by fuzzy clustering [8]. The process of the analysis includes the following steps:

Step1. Calculate the similarity matrix for the attributes. The pair-wise comparison method is used to obtain the values of the corresponding element  $a_{ij}$ , where  $(i=1,2, \dots,k$  and  $j=i+1, i+2, \dots,i+(k-1))$ . The values of  $a_{ij}$  are obtained as  $1-a_{ji}$ . Here  $k$  is the number of the attributes for the corresponding objective.

Step2. Calculate the fuzzy similarity matrix  $R$ . As shown in equation (1) the research adopts the cosine distance measure as the method of similarity measurement of the study objects.

$$r_{ij} = \frac{\sum_{k=1}^m (a_{ik} a_{jk})}{\sqrt{\sum_{i=1}^m a_{ik}^2 \sum_{j=1}^m a_{jk}^2}} \quad (1)$$

During the study the calculation of the fuzzy similarity matrix  $R$  using Euclidian Distance measure was performed as well. The results obtained were almost the same and will be not shown here.

Step3. Calculate the fuzzy transitive closure  $t(R)$  of the fuzzy similarity matrix  $R$  with the square method [9]. If the fuzzy similarity matrix can be expressed as  $R = (r_{ij})_{n \times n}$ , then

$$R \circ R = (t_{ij})_{n \times n} \quad \max_{k=1 \text{ to } n} (\min (r_{ik}, r_{kj})) \quad (2)$$

If  $[R]^{2^k} \circ [R]^{2^k} = [R]^{2^k}$ , then the fuzzy transitive closure  $[t(R)] = [R]^{2^k}$ .

Use the cluster method to analyze  $t(R)$  with intercept  $\lambda$  and determine the significant attributes set.

### III. DATA ANALYSIS

The investigation gave 300 questionnaires out to the individual mobile customers. The questionnaire enumerates the quality parameters of the attribute layer corresponding to a given service of the objective layer. For each service on of the objective layer the number of the customers that use the service is given and on the attribute layer - the number of users that have evaluate the corresponding attribute. The results obtained by the investigation of all three groups are shown in the following Table III. The number of answers for each attribute gives the relative importance of the corresponding attribute in the group. This relative importance is used as a weight of the attribute for the calculation of the fuzzy set values.

The calculation procedure is as follows:  
Step 1. For each one group, after partitioning  $A$  into  $C$  and  $D$  the membership degree of each one attribute is calculated. For example: For the service 1 (S1) in group "Teachers" and the corresponding service attributes the following set  $A$  is obtained: There are 57 ( $N = 57$ ) positive answers on the above question 1. This is equal to 0,19 ( $K = N/N_{all}$ ) of all participants in this study. Then the number of the positive answers on question 3, related to  $K$  is calculated. So the weight of each one attribute value is defined.

The elements in set  $A$  are calculated according the pairwise[10] comparison of the attribute value with respect to the

TABLE III  
INVESTIGATION RESULTS – GROUPS "TEACHERS",  
"ADMINISTRATION", "STUDENTS"

N o.	Teachers		Administration		Students	
	Objectives	Attributes	Objectives	Attributes	Objectives	Attributes
1	S1 - 57	S11 - 31 S12 - 11 S13 - 5 S14 - 24 S15 - 55	S1 - 76	S11 - 44 S12 - 21 S13 - 31 S14 - 25 S15 - 48	S1 - 87	S11 - 54 S12 - 43 S13 - 56 S14 - 44 S15 - 48
2	S2 - 78	S21 - 77 S22 - 65 S23 - 71 S24 - 45 S25 - 76	S2 - 89	S21 - 69 S22 - 75 S23 - 77 S24 - 58 S25 - 82	S2 - 92	S21 - 88 S22 - 74 S23 - 70 S24 - 66 S25 - 87
3	S3 - 100	S31 - 99 S32 - 93 S33 - 45	S3 - 100	S31 - 90 S32 - 75 S33 - 80	S3 - 100	S31 - 100 S32 - 98 S33 - 100
4	S4 - 14	S41 - 7 S42 - 11 S43 - 8 S44 - 6	S4 - 2	S41 - 1 S42 - 2 S43 - 0 S44 - 1	S4 - 34	S41 - 13 S42 - 12 S43 - 14 S44 - 12
5	S7 - 23	S71 - 15 S72 - 16 S73 - 21 S74 - 19	S7 - 8	S71 - 5 S72 - 8 S73 - 4 S74 - 3	S7 - 47	S71 - 31 S72 - 30 S73 - 18 S74 - 42

weight of the attribute values:

$$a_{ij} = \frac{e^{\delta_j - \delta_i}}{1 + e^{\delta_j - \delta_i}} = \log it^{-1}(\delta_j - \delta_i), \quad (3)$$

where  $\delta_k$  is the scale location of object  $k$  and  $\log it^{-1}$  is the inverse  $\log it$  function.

Here the calculations for the group "Teachers" are shown:

So the fuzzy set  $A$  is obtained:

$$A = \begin{bmatrix} 1 & 0.56218 & 0.57444 & 0.54157 & 0.37754 \\ 0.43782 & 1 & 0.51250 & 0.47918 & 0.32082 \\ 0.42556 & 0.48750 & 1 & 0.46672 & 0.31003 \\ 0.45843 & 0.52082 & 0.53328 & 1 & 0.33924 \\ 0.62246 & 0.67918 & 0.68997 & 0.66076 & 1 \end{bmatrix} \quad (4)$$

Then, on Step 2, the fuzzy similarity matrix  $R$  is calculated:

$$R = \begin{bmatrix} 1 & 0.95463 & 0.94595 & 0.95094 & 0.92991 \\ 0.95463 & 1 & 0.98009 & 0.96527 & 0.92442 \\ 0.94595 & 0.98009 & 1 & 0.95725 & 0.90856 \\ 0.95094 & 0.96527 & 0.95725 & 1 & 0.92817 \\ 0.92991 & 0.92442 & 0.90856 & 0.92817 & 1 \end{bmatrix} \quad (5)$$

After that the transitive closure  $[t(R)]$  for the set of parameters is calculated according equation (2). In all cases, being studied, the intercept  $\lambda$  is chosen from the values obtained for  $[t(R)]$ .

$$[t(R)] = \begin{bmatrix} 1 & 0.95463 & 0.95463 & 0.95463 & 0.92991 \\ 0.95463 & 1 & 0.98009 & 0.96527 & 0.92991 \\ 0.95463 & 0.98009 & 1 & 0.96527 & 0.92991 \\ 0.95463 & 0.96527 & 0.96527 & 1 & 0.92991 \\ 0.92991 & 0.92991 & 0.92991 & 0.92991 & 1 \end{bmatrix} \quad (6)$$

Here we can have the following values for  $\lambda$ . For each  $\lambda$  we can define the corresponding clusters of parameters:

$$\begin{aligned} \lambda=1 & \quad \{S11\}, \{S12\}, \{S13\}, \{S14\}, \{S15\} \\ \lambda=0.98009 & \quad \{S12, S13\}, \{S11\}, \{S14\}, \{S15\} \\ \lambda=0.96527 & \quad \{S12, S13, S14\}, \{S11\}, \{S15\} \\ \lambda=0.95463 & \quad \{S11, S12, S13, S14\}, \{S15\} \\ \lambda=0.92991 & \quad \{S11, S12, S13, S14, S15\} \end{aligned}$$

Then we build the dynamic cluster diagram, from which the most significant parameters can be obtained. As shown on the figure below, for the group “Teachers”, the most significant parameters of the service S1 – Voice mail box are the parameters coded as S12 and S13. All others parameters are concatenated one after another to the base cluster, built from these two parameters.

This dynamic cluster diagram can be implemented later as a model of deductive database for easier search of significant attributes also in cases of greater amount of parameters or in case of deeper investigation of the significant service parameters. An appropriate method for cost effective search in such database structures is the 2P-Method introduced in [11].

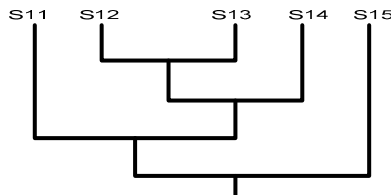


Fig. 1. Dynamic cluster diagram

Following the same steps the following calculation and results are achieved for all other services in correspondence to the services being studied.

TABLE IV  
SIGNIFICANT PARAMETERS

Service	Group	$\lambda$	Parameter clusters	Significant Parameters
S1	Teachers	0.91577	{S11},{S12,S13}, {S14},{S15}	S12, S13
	Administration	0.96807	{S11},{S12,S14}, {S13},{S15}	S12, S14
	Students	0.95929	{S11},{S12,S14}, {S13},{S15}	S12, S14
Final				<b>S12, S14</b>
S2	Teachers	0.96113	{S21},{S22,S24}, {S23},{S25}	S22, S24
	Administration	0.96956	{S21,S24},{S22}, {S23},{S25}	S21, S24
	Students	0.95954	{S21},{S22},{S23,S24}, {S25}	S23, S24
Final				<b>S24</b>
S3	Teachers	0.93725	{S31, S33},{S32}	S31, S33
	Administration	0.85246	{S31}, {S32, S33}	S32, S33
	Students	0.86190	{S31, S33}, {S32}	S31, S33
Final				<b>S31, S33</b>
S4	Teachers	0.95469	{S41, S44}, {S42}, {S43}	S41, S44
	Administration	0.98709	{S41, S43, S44}, {S42}	S41, S43, S44
	Students	0.94694	{S42, S44}, {S41}, {S43}	S42, S44
Final				<b>S41, S44</b>
S7	Teachers	0.95217	{S71, S72},{S73},{S74}	S71, S72
	Administration	0.95968	{S71},{S72},{S73, S74}	S73, S74
	Students	0.86853	{S72},{S71, S73},{S74}	S71, S73
Final				<b>S71, S73</b>

Here, from all other calculations, only the highest grade of  $\lambda$ , the corresponding parameter clustering and the final result–

the most significant parameters for all groups and services will be shown (Table IV).

#### IV. CONCLUSION

In this paper a fuzzy evaluation approach is introduced, used for definition of significant service parameters, that defines the customer satisfaction with the mobile services. The study has to be continued in order to evaluate the already defined significant attributes from the mobile operator point of view.

The correspondence of the evaluation results will be a good starting point for development of appropriate approaches, methods and tools for improving the grade and the effectiveness of the Customer Relationship Management and the Customer Experience Management.

#### REFERENCES

- [1] Al. Tsenov, “Customer Experience Hierarchy Model”, Proceedings of the XLV-th International Scientific Conference ICEST 2010, Ohrid, Macedonia, June 24 – 26, vol.1, pp. 77-80, 2010.
- [2] Al. Tsenov, G. Yoncheva, E. Stoyanova and Al. Pavlov, “Fuzzy Evaluation of Service Level Management Metrics”, XLVI-th International Scientific Conference ICEST 2011, Nish, Serbia, June 29 – July 01, presented paper, 2011
- [3] W. Lin, C. Lo, K. Chao, and N. Godwin, “Fuzzy Similarity Clustering for Consumer-Centric QoS-aware Selection of Web Services”, in Proc. CISIS 2009, pp. 904-909, 2009
- [4] S. Yajing et al., “Customer Value Hierarchy Based Customer Demand Analysis In Personalized Service Recommender System”, International Journal of Simulation, Vol. 7 No 7, ISSN 1473-0831 print, pp. 77-84, 2007
- [5] Y. Danfeng, Y. Fangchun, “Fuzzy evaluation of SLA-oriented QoSM (The Quality of Service Management) in NGN”, Proceedings of IC-BNMT, ISBN 978-1-4244-4591-2, pp. 301-305, IEEE, 2009
- [6] Tsenov Al., I. Ivanov, T. Poparova, S. Neykov, L. Ivanova, M. Gadjeva, “Fuzzy evaluation of customer satisfaction with mobile services”, 10th International Conference TELSIKS 2011, vol. 2, pp. 665 – 668, October 5-8, Nish, Serbia,
- [7] H. Jiang, “Fuzzy Evaluation on ERP System Implementing Risk Based on Membership Degree Transformation New Algorithm”, Second International Symposium on Electronic Commerce and Security, Nanchang, China, May 22-May 24, pp 409-416, 2009
- [8] Jordanova and V. Mladenov, “Fuzzy sets and computational intelligence”, Lecture notes, Sofia University, 2005
- [9] Sun Peide, “Novel Fuzzy Clustering Algorithm to Predict Gas Outburst Zone”. Proceedings of the 2006 International Symposium on Safety Science and Technology, October 24-27, 2006 Changsha, Hu’nan, China
- [10] David, H.A. (1988). The Method of Paired Comparisons. New York: Oxford University Press. B. Milovanovic, Z. Stankovic, S. Ivkovic and V. Stankovic, "Loaded Cylindrical Metallic Cavities Modeling using Neural Networks", TELSIKS'99, Conference Proceedings, pp.214-217, Nis, Yugoslavia, 1999.
- [11] Iltchev V., Cost Evaluation of Methods for Query Processing in Deductive Database Systems, Information Technologies and Control, Vol. 5, No 1, 2007, pp. 10-20, ISSN:1312-2622, Bulgarian Union of Automation and Informatics, Sofia, Bulgaria,

# Investigate common work of software phone systems in virtual environments and real switching systems

Borislav Rozenov Necov<sup>1</sup> Krasen Zhivkov Bankov<sup>2</sup> Mario Petrov Georgiev<sup>3</sup>

**Abstract - This paper discusses the common work of Linux - based IP phone systems Trixbox and Elastix, build in virtual environments. Their interconnection with real switching systems is applied. Advantages of using virtualization software in communications are shown using VoIP software analyzators. Experiments are made, revealing the advantages of using VoIP telephony and virtualization, leading to much easier maintenance and more flexible services.**

**Keywords – Voice-over-IP, Virtualization, Phone systems**

## I. INTRODUCTION

Transmission of data and voice via VoIP is extremely attractive to business users, service operators and for home users because it allows Internet and data networks, already established in offices[1], enterprises and administrative areas, to transmit voice calls, video conferences to support and other real-time applications.

VoIP telephone applications such as Asterisk-based Trixbox can be uploaded to a virtual environment, which means reducing the number of physical machines and leads to less power consumption, simplifying IT infrastructure, a much easier maintenance and much greater service flexibility [2]. Installation of multiple virtual environments (operating systems and applications) in one physical server (homogeneous hardware) is more economically reasonable than provision and maintenance of each physical server for each application [3].

Figure 1.1. shows the qualitative leap made in the period 2007 to 2009, when the total number of installed physical servers stops to increase and even decreases at the expense of the rapid growth of virtual machines. In the U.S. over the next 2-3 years the number of installed physical servers will be reduced by 20%.



Fig.1.1. Increasing use of virtual servers

In practice, only 10% of servers capacity in a continuous mode is used. A large data center could save electricity equivalent to the consumption of 200,000 households under loading up to 50% of its capacity. A serious problem for the ecological balance is large carbon dioxide emissions. According to calculations, covering 50,000 data centers, emissions will exceed 10 million tons by 2013. These figures leave no doubt that information technology bear their share of responsibility for environmental protection [4].

Optimized use and better distribution of work is achieved By application virtualization, which in turn leads to fewer servers and to shorten the prolonged periods in which they operate without load. Integrated approach in the management of information technology is beneficial not only for climate but also for budget of the companies. It enables them to reduce hardware investments and make large cost savings for electricity [5].

Two experiments are made to examine the virtualization. Telephone software systems Trixbox and Elastix are used as well as softphones and VoIP analyzators.

## II. EXPERIMENTAL RESULTS

Virtual TRIXBOX and ELASTIX are loaded in the virtual machine.

Different software phone systems TRIXBOX and ELASTIX are used when conducting the experiments. Some of them are in virtual environment, while the rest are real working (not in virtual environment). For the purposes of the study some other phone systems can be used such as 3CX and different hardware switching systems. Test computers have 1GB RAM operating memory and processor Intel Pentium 4. Emulating virtual programs can be VMWare products (Workstation, VMWare Player), Microsoft products, Oracle (Virtual Box) etc.

### 2.1. Virtual Trixbox - Virtual Elastix

At the start of the call as well as at its termination, SIP signaling protocol is used.

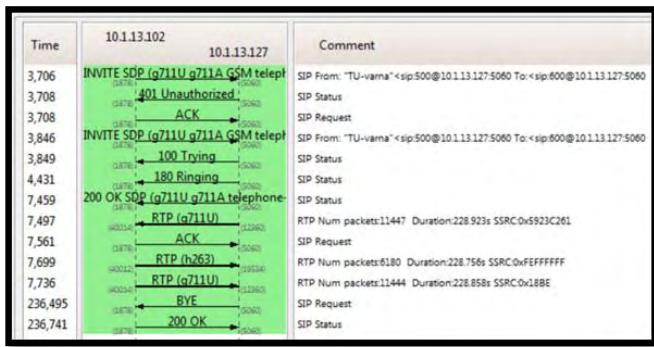


Fig.2.1. Signalling on call

Figure 2.2. shows the number of packets (axis Y) per unit time for 120 seconds. The number of RTP packets vary over time because the length of each packet is not the same and for one unit of time different number of packets can be transmitted.

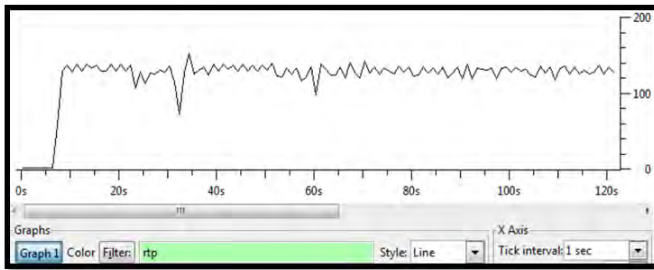


Fig.2.2. Distribution of RTP packets per unit time - IO Graphs

In fig. 2.3. with black the values of jitter are shown, the average is about 3ms. In red delays between packets are marked.

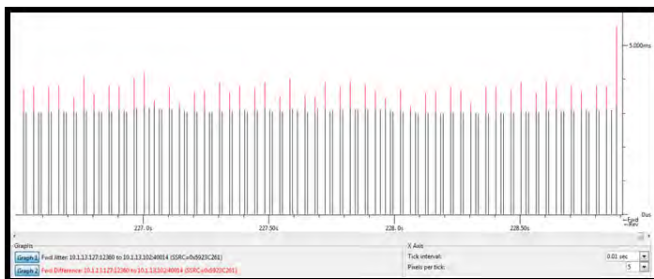


Fig.2.3. Jitter and delay of packets in graphical form

Fig. 2.4 shows service quality of the audio signal (Audio QoS). Reports for 30 seconds back have been noted. R factor (R Factor) is within the allowable (ranging from 0 to 120). Packet loss in unusual circumstances (Burst Packet Loss Rate%) is 0.102% at source and destination at 0.055%. Rejected packs (Discarded Packets) are respectively 2.105 and 26.

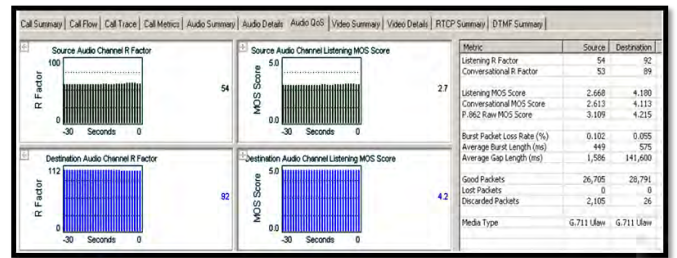


Fig.2.4. Quality of Service audio

## 2.2 Virtual Trixbox - real Trixbox

In this experiment, TRIXBOX is loaded in the virtual machine and real built PBX server TRIXBOX is used.

The graph in Fig. 2.5 shows the number of RTP packets per unit time. Again some hesitation is noticed, because the length of each packet is not the same and therefore requires much time to transmit if more bytes are used.

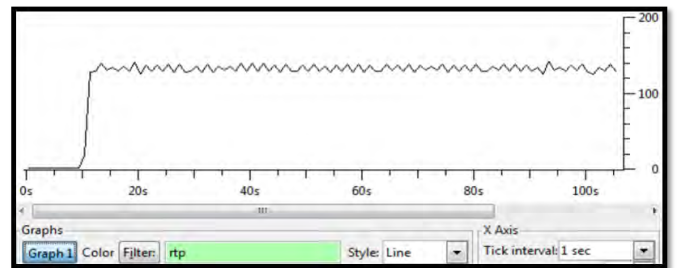


Fig.2.5. Distribution of RTP packets per unit time

The values of jitter (black) and the delay between packets (in red) have improved by about 2ms (Fig. 2.6) compared to the first investigation.

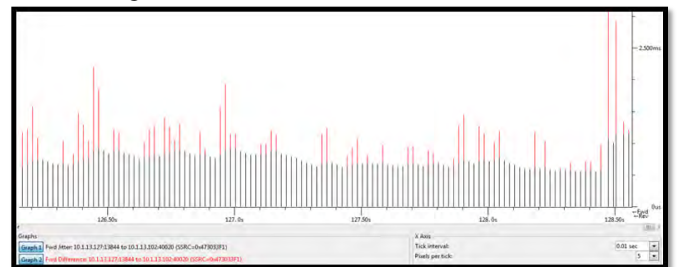


Fig.2.6. Jitter and delay of packets in graphical form

Below is shown the total activity of the network - network activity, VoIP activity, conducted conversations, provided packet activity, protocol activity. Everything is within normal limits.

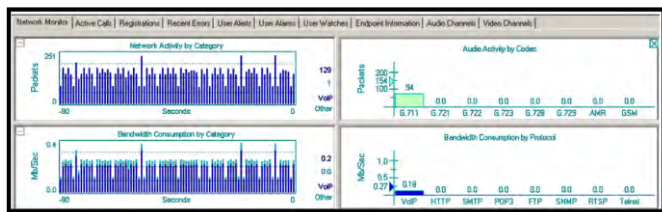


Fig.2.7. Network Monitor

The main parameters that are subject to change in a better direction are shown in fig.2.8 and fig.2.9.

Source R Factor is improved by 10 units and MOS score by 0.5 as the result become almost 3.1. Rejected packets and the loss possibility are less than previous experiment. The average jitter value is significantly reduced in the destination as it becomes 0,415 ms, and in source - 3,5 ms. The average delay between packets is again about 20ms.

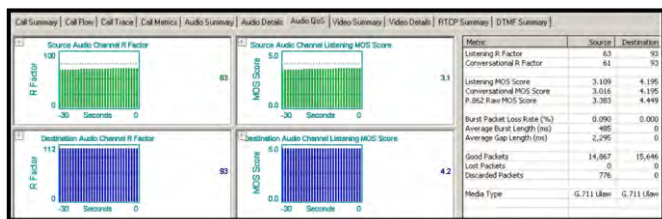


Fig.2.8. Quality of Service audio

Metric	Average	Low	High	Parameter	Source	Destination
Src Jitter (ms)	3.500	0.019	46.316	Address	10.1.13.102	10.1.13.127
Dest Jitter (ms)	0.415	0.005	2.196	Port	40048	11606
Src Packet Interval (ms)	20.000	0.115	335.805	Media Type	G.711 Ulaw	G.711 Ulaw
Dest Packet Interval (ms)	20.004	13.120	35.590	SSRC	00002EA6	3F0EED35
Src Bandwidth (kb/s)	64.028	63.909	67.086	Audio/Package (ms)	20	20
Dest Bandwidth (kb/s)	64.022	63.990	66.916	Frames/Package	20	20
				Total Packets	16,079	16,082
				Packets Lost	0	0
				DTMF Events		
				Current Bandwidth (kb/s)	59.713	59.713
				Longest Packet Loss Bu-	0	0
				Total Payload Bytes	2,572,640	2,573,120

Fig.2.9. Audio Details

### III. CONCLUSION

The advantages of virtualization are undisputed. It goes faster, both in business and in everyday life. In common work with communication systems and applications it has abilities that will undoubtedly make its use more widely.

### REFERENCES

- [1]Свилен Иванов, „Изграждане на виртуални машини с Xen”, [http://0101.nccdn.net/1\\_5/2e6/367/236/Xen.pdf](http://0101.nccdn.net/1_5/2e6/367/236/Xen.pdf)
- [2]Microsoft, —Desktop Virtualization Strategy], White Paper
- [3] Kerry Garrison, —trixbox CE 2.6: Implementing, managing, and maintaining an Asterisk-based telephony system], Packt Publishing, 2009
- [4] Наньо Нанев, „Виртуализацията - що е то?], <http://itanalyses.blogspot.com/2009/01/blog-post.html>
- [5] „Виртуализацията - на прага на версия 3.0], [http://cio.bg/3653\\_virtualizaciya\\_na\\_praga\\_na\\_verseiya\\_30](http://cio.bg/3653_virtualizaciya_na_praga_na_verseiya_30)

### Contacts:

- <sup>2</sup>Krasen Zhivkov Bankov, student in . KTT in FE of TU-Varna „ Studentska” № 1 str, e-mail: banki4@abv.bg
- <sup>1</sup>Borislav Rozenov Necov, student in . KTT in FE of TU-Varna, „ Studentska” № 1 str, e-mail: sharkiller@mail.bg
- <sup>3</sup>Mario Petrov Georgiev, student in . KTT in FE of TU-Varna, „ Studentska” № 1 str, e-mail: markata88@gmail.com

# Analysis of current methods and technologies for encoding, distribution and consumption of IPTV services

Jordan Kanev<sup>1</sup> and Stanimir Sadinov<sup>2</sup>

**Abstract** – Internet Protocol television (IPTV) is a system through which television services are delivered using the Internet protocol suite over a packet-switched network, instead of being delivered through traditional terrestrial, satellite signal, and cable television formats.

**Keywords** – IPTV, VoD, MPEG4 AVC, H.264

## I. INTRODUCTION

A new method of delivering and viewing television channels over IP network and high speed broadband access technology is called IPTV, short for internet protocol television and also known as broadband TV. It is not only a distribution method, but also bring new interactive features and changes the way we watch TV. Ability to pause, resume and fast forward TV shows gives freedom to every user to make individual custom program schedule is only a small part of innovations presented by interactive television. Internet protocol television differs from Internet television as Internet TV is streaming video content over public Internet while IPTV is streaming dedicated video content via private managed network with quality of service (QoS). If service provider is delivering three services: Broadband internet, voice over IP (VoIP) and IPTV the technology is called *Triple play* and if wireless mobility is added then it becomes *Quad play*. World largest IT companies like Microsoft, Cisco and Google are involved in developing end-to-end IPTV solutions: Microsoft TV, Google TV and Cisco Content Delivery System.

## II. IPTV ARCHITECTURE

End-to-end IPTV infrastructure consist of three major components: originate (head-end), distribute (delivery network) and consumption (user-end). An overview of typical IPTV architecture is presented on fig.1

As IPTV is transmitted over IP it needs proper compression techniques to compress the video prior to its transmission depending on the available bandwidth. Researchers from ITU-T have found MPEG compression the best possible solution for this challenge. ITU-T has standardized H.264 and that is equivalent to MPEG-4 (part 10) standardized by Moving

picture expert group MPEG. In addition to this H.264/MPEG-4 is used for HDTV, while older standard MPEG-2 is used for SDTV. Another solution is presented by Society of Motion Picture and Television Engineers (SMPTE) and implemented by Windows Media 9 (WM-9) is named VC-1. It has similar features to MPEG-4, but has better integration with PCs.

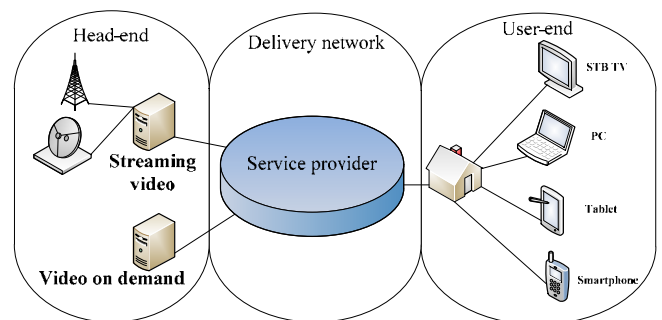


Fig.1. IPTV architecture

Some of the most popular video compression standards are shown in Table I below:

TABLE I  
VIDEO COMPRESSION STANDARDS

Published	Standard	Organization
1993	H.261/MPEG-1	ISO, IEC
1995	H.262/MPEG-2	ISO, IEC, ITU-T
1999	MPEG-4	ISO, IEC
2003	H.264/MPEG-4 AVC	ISO, IEC, ITU-T
2006	VC-1/WM9V	ISO, Microsoft
2008	VC-2 Dirac	ISO, BBC

Originally MPEG-1 is used for Video-CD (VCD) 120 mm optical disc and became the first format for distributing films.

MPEG-2 is internationally accepted standard for digital television and is widely used for digital video broadcast (DVB) systems.

MPEG-4 is used for Internet and mobile video, standard-definition (SDTV) and high-definition television (HDTV). MPEG-4 uses up to 50 % less bandwidth than MPEG-2 with bit rates from 5 Kbit/s to 10 Mbit/s depending on quality and video screen resolution.

To deliver video content over IP network IPTV system needs interactive services. IPTV services can be divided in three main groups: streaming TV, video on demand (VOD) and time-shifted TV. According to ETSI specification [6] NGN integrated IPTV include following IPTV services: Broadcast TV; Content on Demand (nCoD, pCoD); Personal

<sup>1</sup>Jordan Kanev is with the Faculty of Electrical Engineering and Electronics, Technical University of Gabrovo, 4 H. Dimitar St., 5300 Gabrovo, Bulgaria, E-mail: dancho Kanev@gmail.com.

<sup>2</sup>Stanimir Sadinov is with the Faculty of Electrical Engineering and Electronics, Technical University of Gabrovo, 4 H. Dimitar St., 5300 Gabrovo, Bulgaria, E-mail: murry@tugab.bg.



Video Recording (cPVR, nPVR); Pay Per View (PPV); Interactive TV (iTV); User Generated Content (UGC); Profiling and personalization; Content Recommendations (CR); Advertising (Ad) and Targeted Advertising (TAI); Messaging services; Notification services; Personalized channel; Bookmarks or Content Marking (CM).

IPTV services provide users with more control over viewed TV channels and give them ability to make decision what TV content to choose and when to watch it. User can watch IPTV with a number of network-addressable devices like laptop, personal computer, set-top box and TV, smartphone, tablet, gaming consoles, etc.

### III. OBTAINING VIDEO AT HEAD-END

Service providers need to acquire and encode video content. They can use several different video sources and analog, digital and IP technologies to do that. Streaming video can be obtained from following sources:

- Satellite - DVB-S (SDTV) / DVB-S2 (HDTV)
- Cable – DVB-C / DVB-C2
- Terrestrial – DVB-T / DVB-T2
- Analog TV - obsolete

TV signal acquisition needs different equipment like satellite dishes, TV antennas, cables modems, DVB receivers, Set-top boxes (STB) depending on source type and technologies of distribution systems. There are also a variety of modulation methods and schemes, some of them are shown in Table II.

TABLE II  
MODULATION SCHEMES

Standard	Modulation schemes
DVB-S	QPSK,8PSK,16-QAM
DVB-S2	QPSK,8PSK,16APSK,32APSK
DVB-C	QAM : 16- to 256- QAM
DVB-C2	COFDM: 16- to 4096-QAM
DVB-T	COFDM(OFDM) : QPSK, 16QAM, 64QAM
DVB-T2	OFDM: QPSK, 16QAM, 64QAM, 256QAM
Analog	AM(VSB),FM,QAM

Video on demand can be recorded live video stream for later review or uploaded video content from various sources to a dedicated VoD server or media library. Sources can be live TV, Video CDs, DVDs, Blu-ray discs, cameras and various multimedia devices.

After video acquisition it must be processed by video encoder device. Typically live video content is compressed using H.264/MPEG-4 AVC codec. Other codecs can also be used instead. The most common codecs used for VoD are MPEG-2, MPEG-4 and VC-1. In case video is acquired directly in older MPEG-2 or other video format transcoder device can be used to convert data into desired video codec.

VoD content can be encrypted, scrambled and embedded with a tag to avoid content piracy enforcing Digital rights management (DRM).

Then encoded video content is sent in an MPEG transport stream and depending on service it can be forwarded to service provider delivery network in case of live TV or stored on VoD server in case of video on demand service.

### IV. DISTRIBUTION VIA IP NETWORKS

Video originated at Head-end must be preserved and transported across IP network to reach end users. Broadband networks used for Internet (data) and voice services are ideal for adoption of IPTV. Physical layer (Layer 1) can be copper, fiber optic, combination of both types and even broadband wireless (IEEE 802.11n). Technologies used for data link layer (Layer2) are MetroEthernet, xDSL, FTTx, ATM, WiMAX, LTE, DVB-H, etc. Depending on cable types and physical topology IP networks include intermediate devices like routers and switches to forward user data across the network. Additional mechanisms and protocols are used to ensure reliable delivery of IPTV traffic. IPTV traffic is sensitive to losses and delay. Quality of Service (QoS) is responsible for prioritizing video over other traffic on the network. QoS Standards include IEEE 802.1p QoS, IEEE 802.1q VLAN, CEA2007 VLAN Mapping.

Additional control software (or middleware) is required to: gain user access privileges, manage video content, to protect intellectual property via Digital Rights Management (DRM), billing software and offer services like electronic programming guide (EGP), VoD catalog, web services, mobile applications, domain name system (DNS), Network Time Protocol (NTP), Dynamic Host Configuration Protocol (DHCP), nCoD, nPVR, advertising, etc.

IPTV is delivered via IP multicast in case of streaming TV or via IP unicast in case of video on demand. More complex distribution method is available with combination of both unicast and multicast delivery mode depending on underlying IP network topology of service provider delivery network. Internet Group Management Protocol (IGMP) is used to join and leave multicast streams. IP Multicast Standards include IGMPv1, IGMPv2 and IGMP Snooping.

According to ETSI specification [7] the transport streams should be encapsulated with Real-Time Transport Protocol (RTP). Real Time Streaming Protocol (RTSP) is responsible for control over delivery of video content. Real-Time Transport Control Protocol (RTCP) is used to send statistics and control information for QoS.

### V. USER-END

User devices for IPTV reception include: traditional set-top box (STB), IP STB, PC, smartphone, tablet and any internet browsing device able to play video content.

### VI. CONCLUSION

This paper describes current state of IPTV technology. Main goal is to deliver better video quality at the lower bit rate, offering virtually unlimited number of TV channels. Adding new features and services is very easy and straightforward. Services can be adopted with minimal investments and with use of existing IP network infrastructure delivered to users.

#### ACKNOWLEDGEMENT

This paper is supported by Experimental Research Unit (UTzNIT) at the Technical University of Gabrovo.

#### REFERENCES

- [1] O'Driscoll, G., "Next Generation IPTV Services and Technologies", John Wiley & Sons, Inc, 2008.
- [2] Cooper, W., Lovelace, G., "IPTV.Guide", informitv, 2006.
- [3] Microsoft, "Microsoft TV IPTV Edition", available online at <http://www.microsoft.com/TV>.
- [4] Cisco, "Cisco IPTV Solutions", available online at <http://www.cisco.com/go/iptv>.
- [5] IEEE 802-2001: "IEEE Standards for local and metropolitan area networks: overview and architecture", 2002.
- [6] ETSI TS 182 028: TISPAN NGN Release v.3.5.1: "NGN integrated IPTV subsystem Architecture", 2011.
- [7] ETSI TS 102 034 v1.4.1 "Transport of MPEG-2 TS Based DVB Services over IP Based Networks", 2009
- [8] Конов, К., "Цифрово радио и телевизионно разпръскване", Диос, 2011.

# Average SIR Comparison for SC Systems Using Different Decision Algorithms in the Presence of Interference

Aleksandra Panajotović<sup>1</sup>, Dragan Drača<sup>2</sup> and Nikola Sekulović<sup>3</sup>

**Abstract** – Ascertaining the importance of the selection combining (SC) as the most efficient diversity technique, average signal-to-interference ratio (SIR) at the output of dual SC receiver is investigated in the paper as important performance criterion. The diversity system operates in microcell environment and applies desired signal power algorithm. Numerical results are presented to show effects of fading severity and level of correlation. Moreover, they are used to compare performance of dual SC systems using different decision power algorithms.

**Keywords** – Cochannel interference, Fading, Selection combining.

## I. INTRODUCTION

The performance of wireless system is severely affected by fading and cochannel interference (CCI). Fading is result of multipath propagation, and CCI is result of frequency reuse [1]. Space diversity techniques, which combine input signals from multiple receive antennas (diversity branches), are the well known techniques that alleviate the deleterious effects of fading and CCI. The most popular diversity techniques are maximal-ratio combining (MRC), equal-gain combining (EGC) and selection combining (SC). Among of these types of diversity technique, SC has the least implementation complexity since it processes only one of diversity branches [2]. Usually, SC receiver chooses the branch with the highest signal-to-noise ratio (SNR), which corresponds to the strongest signal if equal noise is assumed among the diversity branches. In some systems, where CCI is more significant than noise, SC receiver can employ one of following decision power algorithms: the desired signal power algorithm, the total signal power algorithm and the signal-to-interference power ratio (SIR) algorithm [3].

There are few statistical models used to describe fading in wireless environment. Rician statistical model is typically observed in the line-of-sight (LoS) path of microcellular urban and suburban land mobile environment [4]. In such environment, CCI experiences significantly deeper fading than the desired signal. Therefore, the different fading models for the desired signal and CCI have to be used. Cochannel interference from distant microcell may be modelled by

<sup>1</sup>Aleksandra Panajotović is with the Faculty of Electronic Engineering at University of Niš, Aleksandra Medvedeva 14, Niš 18000, Serbia, E-mail: [aleksandra.panajotovic@elfak.ni.ac.rs](mailto:aleksandra.panajotovic@elfak.ni.ac.rs).

<sup>2</sup>Dragan Drača is with the Faculty of Electronic Engineering at University of Niš, Aleksandra Medvedeva 14, Niš 18000, Serbia, E-mail: [dragan.draca@elfak.ni.ac.rs](mailto:dragan.draca@elfak.ni.ac.rs).

<sup>3</sup>Nikola Sekulović is with the School of Higher Technical Professional Education, Aleksandra Medvedeva 20, Niš 18000, Serbia, E-mail: [sekulani@gmail.com](mailto:sekulani@gmail.com).

Rayleigh statistics [5,6].

There is set of performance criteria that allow the system designer to evaluate the performance of wireless systems and investigate influence of key system parameters. The most popular first order performance criteria are outage probability (OP), average bit error probability (ABEP), channel capacity, average output SNR/SIR, etc [7-11]. In this paper, analytical expression for average SIR at the output of dual SC receiver operating over correlated Rician fading channels in the presence of Rayleigh CCI is derived for the case when receiver applies desired signal power algorithm. Numerical results illustrate the influence of system and channel parameters on the system performance. Moreover, our previous published result [12] is used to compare average output SIR of considered SC system with SC system applying SIR decision algorithm.

## II. SYSTEM AND CHANNEL MODEL

If diversity system is applied in small terminal, correlation arises between diversity branches due to insufficient antenna spacing. In that case, desired signal envelope,  $r_1$  and  $r_2$ , on two diversity branches follow correlated Rician distribution whose probability density function (PDF) is given by [13]

$$p_{r_1 r_2}(r_1, r_2) = \frac{r_1 r_2}{\sigma^4(1-\rho^2)} \exp\left(-\frac{r_1^2 + r_2^2 + 2b^2(1-\rho)}{2\sigma^2(1-\rho^2)}\right) \times \sum_{k=0}^{\infty} \varepsilon_k I_k\left(\frac{r_1 r_2 \rho}{\sigma^2(1-\rho^2)}\right) I_k\left(\frac{b r_1}{\sigma^2(1+\rho)}\right) I_k\left(\frac{b r_2}{\sigma^2(1+\rho)}\right), \quad (1)$$

$$\varepsilon_k = \begin{cases} 1, & k=0 \\ 2, & k \neq 0 \end{cases}$$

where  $\rho$  is branch correlation coefficient and  $I_k(\cdot)$  is modified Bessel function of the first kind and  $k$ -th order. Rice factor and average desired signal power are defined as  $K = b^2/(2\sigma^2)$  and  $\beta = \sigma^2(1+K)$ , respectively.

We assume that there is a single dominant interferer, independent of the desired signal, subjected to Rayleigh fading [14]. PDF of its envelope is expressed by

$$p_a(a) = \frac{a}{\sigma_a^2} \exp\left(-\frac{a^2}{2\sigma_a^2}\right), \quad (2)$$

where  $\sigma_a^2$  is average CCI power.

The considered dual SC receiver uses the desired signal power decision algorithm. Actually, it selects the branch with the largest instantaneous desired signal power, i.e.

$r^2 = \max\{r_1^2, r_2^2\}$ . The instantaneous SIR at its output is given by  $\eta = \max\{r_1^2, r_2^2\} / a^2 = r^2 / a^2$ .

### III. AVERAGE OUTPUT SIR

Average output SIR is important performance criterion for wireless system operating in interference-limited environment. The average SIR envelope,  $\bar{\mu}$ , at the output of SC system can be obtained by averaging the instantaneous SIR envelope,  $\mu = \sqrt{\eta} = r/a$ , over its PDF [15]

$$\bar{\mu} = \int_0^{\infty} \mu p_{\mu}(\mu) d\mu, \quad (3)$$

where  $p_{\mu}(\mu)$  for SC system with desired signal power decision algorithm was derived in [16] as

$$p_{\mu}(\mu) = \exp\left(-\frac{2K}{1+\rho}\right) \sum_{k,p,n,l=0}^{\infty} \varepsilon_k \frac{\rho^{2p+k} K^{n+l+k} (1+K)^{p+k+1}}{n!l!p!\Gamma(p+k+1)\Gamma(l+k+1)} \times \frac{S\mu^{2p+2k+1}}{\Gamma(n+k+1)(1-\rho)^p(1+\rho)^{2k+p+n+l}} \times \left[ \frac{(p+n+k)!(1+K)^l \mu^{2l} (1-\rho)^n}{(1+\rho)^l} \left[ \frac{2(p+k+l+1)!}{\left(S + \frac{(1+K)\mu^2}{1-\rho^2}\right)^{p+k+l+2}} - \sum_{i=0}^{p+n+k} \frac{(p+l+k+i+1)!(1+K)^i \mu^{2i}}{2^{p+k+l+i+1} i! (1-\rho^2)^i \left(\frac{S}{2} + \frac{(1+K)\mu^2}{1-\rho^2}\right)^{p+k+l+i+2}} \right] + \frac{(p+l+k)!(1+K)^n \mu^{2n} (1-\rho)^l}{(1+\rho)^n} \left[ \frac{2(p+k+n+1)!}{\left(S + \frac{(1+K)\mu^2}{1-\rho^2}\right)^{p+k+n+2}} - \sum_{j=0}^{p+l+k} \frac{(p+n+k+j+1)!(1+K)^j \mu^{2j}}{2^{p+k+l+j+1} j! (1-\rho^2)^j \left(\frac{S}{2} + \frac{(1+K)\mu^2}{1-\rho^2}\right)^{p+k+n+j+2}} \right] \right] \quad (4)$$

Substituting Eq. (4) into Eq. (3), the average output SIR envelope is obtained in the analytical form using [17, Eq. (3.194(3))]

$$\bar{\mu} = \exp\left(-\frac{2K}{1+\rho}\right) \sum_{k,p,n,l=0}^{\infty} \varepsilon_k \frac{\rho^{2p+k} K^{n+l+k} \Gamma(0.5)}{n!l!p!\Gamma(p+k+1)\Gamma(l+k+1)} \times \frac{\sqrt{S}(1-\rho)^{k+n+l+1.5}}{\Gamma(n+k+1)\sqrt{1+K}(1+\rho)^{k+n+l-1.5}} \times \left\{ (p+n+k)! \left[ \Gamma(p+k+l+1.5) - \sum_{i=0}^{p+n+k} \frac{\Gamma(p+k+l+i+1.5)}{2^{p+k+l+i+2} i!} \right] + (p+l+k)! \left[ \Gamma(p+k+n+1.5) - \sum_{j=0}^{p+l+k} \frac{\Gamma(p+k+n+j+1.5)}{2^{p+k+n+j+2} j!} \right] \right\}, \quad (5)$$

where  $S$  is average input SIR defined as  $S = \beta / \sigma_a^2$ .

The average SIR envelope at the output of SC system applying SIR decision algorithm can be evaluated using Numerical Integration in program package Mathematics after substitution  $p_{\mu}(\mu)$  [12]

$$p_{\mu}(\mu) = \exp\left(-\frac{2K}{1+r}\right) \sum_{k,p,n,l,m=0}^{+\infty} \frac{2\varepsilon_k K^{p+l+k}}{\beta^{2k+2n+p+l+2}} \cdot \frac{\Gamma(n+l+m+k+2) r_a^{2m} r^{2n+k} \mu^{4n+4k+2p+2l+3}}{\Gamma(m+1)\Gamma(l+k+1)\Gamma(n+k+1)\Gamma(p+k+1)} \cdot \frac{(K+1)^{2k+2n+p+l+2} \Gamma(n+p+m+k+2)}{n!p!m!l!(1-r)^{2n+k+1}(1+r)^{2p+2l+2n+3k+1}} \times \left[ \frac{(1-r_a^2)^{n+l+k+1-m} \sigma_a^{2n+2l+2k-2m}}{\left(\frac{1}{\sigma_a^2(1-r_a^2)} + \frac{\mu^2(K+1)}{\beta(1-r^2)}\right)^{n+p+k+m+2}} \cdot \frac{{}_2F_1[n+l+k+m+2, n+l+k+1, n+l+k+2, -\alpha\mu^2]}{(n+l+k+1)} + \frac{{}_2F_1[n+p+k+m+2, n+p+k+1, n+p+k+2, -\alpha\mu^2]}{(n+p+k+1)} \right] \cdot \left[ \frac{(1-r_a^2)^{n+p+k+1-m} \sigma_a^{2n+2p+2k-2m}}{\left(\frac{1}{\sigma_a^2(1-r_a^2)} + \frac{\mu^2(K+1)}{\beta(1-r^2)}\right)^{n+l+k+m+2}} \right], \quad (6)$$

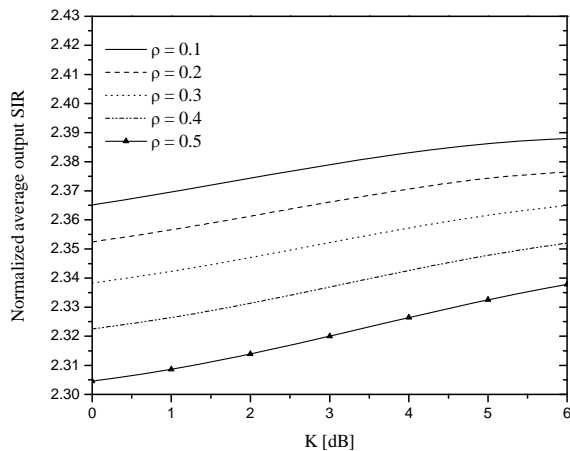
in (3). The parameters  $r_a$  and  $r$  are branch correlation coefficients and  $\alpha = \frac{\sigma_a^2(1-r_a^2)(K+1)}{\beta(1-r^2)}$ .

### IV. NUMERICAL RESULTS

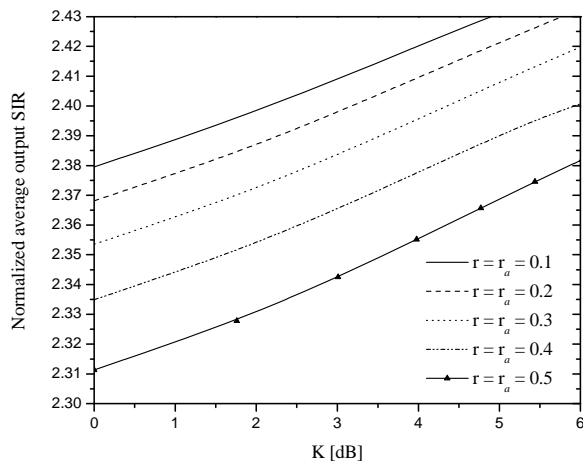
Previous proposed mathematical analysis is complemented in this section through illustration of influence of system and channel parameters on the average output SIR as important system performance criterion.

Figure 1 shows normalized average SIR ( $\bar{\mu}/\sqrt{S}$ ) at the output of dual SC system applying different decision

algorithms in function of Rice factor. The performance curves are evaluated for different values of correlation coefficient in order to show influence of distance between diversity branches on considered performance criterion. Regardless of applied decision algorithm, diversity gain decreases with increase of branch correlation coefficient. Also diversity gain is greater for environment with light fading than for environment with severe fading. Comparison of Fig.1 (a) and Fig. 1 (b) shows advantage of SIR signal power decision algorithm because it provides better diversity gain. That advantage is more noticeable for greater values of Rice factor. In environment with severe fading it is better to use SC system with desired signal power algorithm because it requires less complicate receiver and gives almost the same diversity gain as SC system with SIR decision algorithm.



a)



b)

Fig.1. Average output SIR versus Rice factor for several values of correlation coefficient:

a) desired signal power algorithm; b) SIR power algorithm

The main problem in the infinite series expression of the average output SIR is its convergence. The nested infinite

sums in (5) and (6) converge for any values of branch correlation coefficient and Rice factor. As is shown in Table I, the number of terms that need to be summed to achieve the desired accuracy strongly depends on branch correlation.

TABLE I  
NUMBER OF TERMS SUMMED TO ACHIEVE THREE-SIGNIFICANT-FIGURE ACCURACY OF AVERAGE OUTPUT SIR (DESIRED SIGNAL/SIR ALGORITHM)

$\rho = 0.1$	10/11
$\rho = 0.2$	9/13
$\rho = 0.3$	9/12
$\rho = 0.4$	10/9
$\rho = 0.5$	11/16

## V. CONCLUSION

In this paper the performance of dual SC system operating in interference-limited microcell environment has been studied. Actually, average output SIR as important performance criterion has been derived in infinite series form for the case when SC system using desired signal power decision algorithm. Presented numerical results have described influence of fading severity and correlation coefficient on considered performance criterion. Moreover, evaluated results have been compared with results obtained for SIR decision algorithm. The general conclusion of this paper is that SC diversity system with SIR algorithm provides better diversity gain regardless of working conditions.

## ACKNOWLEDGEMENT

This work has been funded by Serbian Ministry for Education and Science under the projects TR-32052, III-44006, TR-33035.

## REFERENCES

- [1] W. C. Jakes, *Microwave Mobile Communications*, New York, Wiley, 1974.
- [2] M. K. Simon, M. -S. Alouini, *Digital Communications over Fading Channels*, 2<sup>nd</sup> ed., New York, Wiley, 2005.
- [3] L. Yang, M. -S. Alouini, "Average Outage Duration of Wireless Communications Systems", ch. 8, *Wireless Communications Systems and Networks*, Springer, 2004.
- [4] K. A. Stewart, G. P. Labeledz, and K. Sohrabi, "Wideband channel measurements at 900 MHz", VTC'95, Conference Proceedings, pp. 236-240, Chicago, USA, 1995.
- [5] Y. -D. Yao, 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.
- [6] Y. -D. Yao, A. U. H. Sheikh, "Outage probability analysis for microcell mobile radio system with cochannel interferers in Rician/Rayleigh fading environment", *Electron. Lett.*, vol. 26, no. 13, pp. 864-866, 1990.
- [7] I. Petrović, M. Stefanović, P. Spalević, S. Panić, and D. Stefanović, "Outage Analysis of Selection Diversity over Rayleigh Fading Channels with Multiple Co-Channel

- Interferers”, *Telecomm. Sys.*, accepted for publication, DOI: 10.1007/s11235-011-9438z.
- [8] I. Petrović, Z. Nikolić, M. Stefanović, S. Panić, P. Spalević, and Dj. Bandjur, “Multiple Co-Channel Interferers Influence on Selection Combining over Correlated Weibull Fading Channels”, *Frequenz*, accepted for publication.
- [9] Dj. Bandjur, M. Stefanović, and M. Bandjur, “Performance Analysis of SSC Diversity Receiver over Correlated Ricean Fading Channels in the Presence of Co-Channel Interference”, *Electron. Lett.*, vol. 44, no. 9, pp. 587-588, 2008.
- [10] Z. Popović, S. Panić, J. Anastasov, P. Spalević, and M. Stefanović, “Performance Analysis of Selection Diversity over Exponentially Correlated  $\alpha$ - $\mu$  Fading Environment”, *Intern. J. Commun. Syst.*, vol. 24, no. 7, pp. 925-937, 2011.
- [11] S. Jovković, S. Panić, M. Stefanović, P. Spalević, and D. Krstić, “Performance Analysis of SSC Diversity Receiver over Correlated Ricean Fading Channels in the Presence of Co-Channel Interference”, *EURASIP J. Wireless Commun. Net.*, vol. 2010, DOI: 10.1155/2010/583093.
- [12] A. Panajotović, M. Stefanović, and D. Drača, “Performance Analysis of System with Selection Combining over Correlated Ricean Fading Channels in the Presence of Cochannel Interference”, *AEÜ*, pp. 1061-1066, vol. 63, no. 12, 2009.
- [13] M. K. Simon, *Probability Distributions Involving Gaussian Random Variables – a Handbook for Engineers, Scientists and Mathematicians*, New York, Springer, 2002.
- [14] H. Yang, M. –S. Alouini, “Outage Probability of Dual-Branch Diversity Systems in the Presence of Co-Channel Interference”, *IEEE Trans. Wireless Commun.*, vol. 2, no. 2, pp. 310-319, 2003.
- [15] G. K. Karagiannidis, “Performance Analysis of SIR-based dual Selection Diversity over Correlated Nakagami-m Fading Channels”, *IEEE Trans. Veh. Technol.*, vol. 52, no. 2, pp. 1207-1216, 2003.
- [16] A. Panajotović, N. Sekulović, M. Stefanović, and D. Drača, “BEP Comparison for Dual SC Systems Using Different Decision Algorithms in the Presence of Interference”, *TELSIKS’11*, Conference Proceedings, vol. 2, pp. 467-470, Niš, Serbia, 2011.
- [17] I. S. Gradshteyn, I. M. Ryzhik, *Table of Integrals, Series, and Products*, CD version 1.0, Academic Press, 1996.

# Optimization of Traffic Distribution Coefficients in IP Radio-Relay Network with Path Diversity

Dragana Perić<sup>1</sup>, Miroslav Perić<sup>2</sup>, Branislav M. Todorović<sup>3</sup>,  
Milan Šunjevarić<sup>4</sup>, Miroslav Popović<sup>5</sup>

**Abstract** - This paper presents results of optimization of traffic distribution coefficients between primary and backup route in IP radio-relay network. For generation of primary and backup route NOPHY algorithm is used which ensures that they don't have common links. Results of network performance simulation with random link degradation are given.

**Keywords** - load balancing, backup routes, fading, traffic protection

## I. INTRODUCTION

Radio networks can be used for transmission of IP traffic between computer networks. Achievable throughput of individual point-to-point link that is used for router interconnection in these networks is greater than 1 Gbit/s, especially in the case of millimeter wave links [1]. This solution is competitive to fiber optics and free space optics for short distances in urban areas, and its advantage is quick installation and reconfiguration. Main drawback of high capacity radio link is its susceptibility to additional signal attenuation (fading) caused by multipath propagation or rain attenuation, that can cause link outages [2]. In order to reduce influence of fading on an individual IP link various techniques of tradeoff between link capacity and signal level are performed like: adaptive rate, adaptive coding and adaptive modulation [3]. Such radio link could be described as a function of link capacity vs. receiver signal level. Using propagation model described in [4] for each individual link, percentages of time when it operates at full capacity, reduced capacity or when it is unavailable (in link down state) could be calculated. In the procedure of radio link planning, radio link parameters like transmitter output power and antennas gains are chosen to meet link performance and availability objectives. Usually telecom operators accept link performance objectives given in ITU-R recommendations [5], [6]. Depending on link class, unavailability objectives are in range from about 0.01% to 0.05% (about several tens of minutes to several hours at annual level). When link is available, the

majority of time (typically about 99.9% of time) it operates with full capacity. In smaller percentages of time 0.01% to 0.09% link operates with reduced capacity.

When smaller percentage of unavailability is needed, the only solution is to use diversity techniques in radio network, when signal is transmitted over independent links. Switching between paths could be done by simple rerouting [11], when unavailable links are not used for IP packet transmission. However, for rerouting network convergence process might not be fast enough to track network changes caused by fading [7]. In order to overcome this problem, cross-layer routing techniques could be used [8].

One of such methods, described in [9],[10], exploits traffic distribution on primary and backup routes, that can serve both as unavailability protection and network performance improvement method. It is shown that pre-calculated backup routes could be good protection mechanism and algorithms for backup route generation are explained. Network performance improvement is achieved because traffic distribution over primary route and backup route can also serve as load balancing and therefore decrease possibility of link congestion [11].

This paper is focused on traffic distribution between primary route and backup route. It is shown that it could be used both for traffic protection and load balancing.

## II. MODEL OF IP RADIO NETWORK

Network consists of  $N$  nodes with IP routers, nodes are connected with  $E$  links, with capacities  $\mathbf{C}=\{c(e)\}$ ,  $c(e)=1,\dots,E$ . Traffic demands between two nodes are denoted by  $h(i,j)$ ,  $i=1,2,\dots,N$ ,  $j=1,2,\dots,N$ . It is assumed that traffic demand exists for each pair of nodes in the network, and they define traffic demand matrix  $\mathbf{H}=\{h(i,j)\}$ ,  $h(i,j)>0$ ,  $i\neq j$ ,  $h(i,j)=0$ ,  $i=j$ .

$$\mathbf{H} = \begin{bmatrix} 0 & h(1,2) & \dots & h(1,N) \\ h(2,1) & 0 & \dots & h(2,N) \\ \dots & \dots & 0 & \dots \\ h(N,1) & \dots & h(N-1,N) & 0 \end{bmatrix} \quad (1)$$

Each traffic demand  $h_{ij}$  can be served by different flows  $x(i,j,k)$ ,  $k=1,\dots,N_k(i,j)$ . Number of flows  $N_k(i,j)$ , can be different for every pair of nodes  $(i,j)$ , with  $N_k(i,j)\geq 1$ .

Routing algorithm assigns route  $r(i,j,k)$  to each flow  $x(i,j,k)$ . Routes are then memorized in the routing table  $\mathbf{R}$ . For path  $r(i,j,k)$ , vector  $Dr(i,j,k)$  indicates if link belongs to path is:

<sup>1</sup>Dragana Perić is with "IMTEL Komunikacije a.d.", Institute of Microwave Techniques and Electronics, 165b M. Pupina Blvd., 11070 Belgrade, Serbia, E-mail: dragana@insimtel.com

<sup>2</sup>Miroslav Perić is with "VLATACOM d.o.o.", Research and Development Center, 5 M. Milankovića Blvd., 11070 Belgrade, Serbia, E-mail: miroslav.peric@vlatacom.com.

<sup>3</sup>Branislav M. Todorović is with RT-RK, Institute for Computer Based Systems, Narodnog Fronta 23A, 21000 Novi Sad, Serbia, E-mail: Branislav.Todorovic@rt-rk.com

<sup>4</sup>Milan Šunjevarić is with RT-RK, Institute for Computer Based Systems, Narodnog Fronta 23A, 21000 Novi Sad, Serbia, E-mail: micosun@eunet.rs

<sup>5</sup>Miroslav Popović is with Faculty of Technical Sciences, Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia, E-mail: Miroslav.Popovic@rt-rk.com

$$Dr(i, j, k) = \{\delta(i, j, k, 1), \dots, \delta(i, j, k, e), \dots, \delta(i, j, k, E)\} \quad (2)$$

$i = 1, 2, \dots, N, j = 1, 2, \dots, N, k = 1, 2, \dots, N_k(i, j)$

where:

$$\delta(i, j, k, e) = \begin{cases} 1, & \text{if link } e \text{ belongs to route } r(i, j, k) \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

for each link  $e=1, \dots, E$ .

Using this notation, maximum number of flows using link  $e$  is defined by capacity of that link  $c_e$  [11]:

$$\sum_{i=1}^N \sum_{\substack{j=1 \\ i \neq j}}^N \sum_{k=1}^{N_k(i, j)} \delta(i, j, k, e) \cdot x(i, j, k) \leq c_e, \quad e = 1, \dots, E \quad (4)$$

It is assumed that each individual link is realized by point-to-point microwave link, which capacity varies with receiver signal level:

$$C = C(n_R) \quad (5)$$

In absence of fading the link operates with full capacity  $C_0$ , while in fading condition it could be reduced by capacity reduction factor  $c_R$ , or be in link down state, when its capacity is equal to zero.

### III. PERFORMANCE CRITERIA BASED ON LINK LOAD

Two performance criteria are defined, both based on link load, with purpose to monitor the highest link load in the network. Traffic demand  $h(i, j)$ ,  $i=1, 2, \dots, N$ ,  $j=1, 2, \dots, N$ ,  $i \neq j$  is served by load balancing among flows  $x(i, j, k)$ ,  $k=1, \dots, N_k(i, j)$ . Balancing factor  $b(i, j, k)$  is calculated as ratio of the flow  $x(i, j, k)$  and the demand  $h(i, j)$  and it is the element of traffic load balancing matrix **B**.

$$b(i, j, k) = \frac{x(i, j, k)}{h(i, j)}, \quad k = 1, \dots, N_k(i, j) \quad (7)$$

It is clear that sum of balancing factors for all flows equals one:

$$\sum_{k=1}^{N_k(i, j)} b(i, j, k) = 1, \quad b(i, j, k) \leq 1 \quad (8)$$

Using elements of the matrix **B**, the vector of link loads is formed  $L_l$  with elements equal number of paths that include link  $e$ .

$$L_l(e) = \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^{N_k(i, j)} b(i, j, k) \cdot \delta(i, j, k, e), \quad e = 1, \dots, E \quad (9)$$

In special case, when  $N_k(i, j)=1$ ,  $i=1, \dots, N$ ,  $j=1, \dots, N$ , each traffic demand is served by only one path,  $L_l(e)$  is an integer that represents number of traffic demands served by given link. If traffic load balancing is applied,  $L_l(e)$  is a rational number that represents equivalent number of flows that serve given link.

Since the link load increases when the number of flows that this link serves increases, and decreases when this link capacity increases, by simple division of  $L_l(e)$  with link capacity we obtain the vector of normalized link loads:

$$L(e) = \frac{L_l(e)}{c(e)}, \quad e = 1, \dots, E \quad (10)$$

A bottleneck in the network is the link with maximum load, so criteria for network performance measure is obtained by maximization of normalized link loads.

$$L_{\max} = \max_{e \in \{1, \dots, E\}} L(e) \quad (11)$$

Links that have  $L_l(e)$  equal to  $L_{\max}$  are actually bottleneck of the network because they are most susceptible to congestion. Apart from maximum load, we can monitor average load:

$$L_{\text{ave}} = \frac{1}{E} \sum_{e=1}^E L(e) \quad (12)$$

In both cases, smaller values of  $L_{\max}$ ,  $L_{\text{ave}}$  indicate that network has better performances, i.e. it has small possibility of congestion in real traffic conditions. In cases when traffic demand cannot be served because it is directed to zero capacity link, criteria become infinity.

### IV. TRAFFIC DISTRIBUTION ALGORITHMS

Even load balancing scheme implies that traffic is equally distributed among available paths. This scheme acts as load balancing and is implemented in many commercially available routers [12]. According to it, service of traffic demand  $h(i, j)$  is equally split between flows  $x(i, j, k)$ . Load balancing coefficients  $b(i, j, k)|_{\text{equal}}$  between nodes  $i$  and  $j$  by  $x(i, j, k)$  are defined by expression:

$$b(i, j, k)|_{\text{equal}} = 1 / N_k(i, j), \quad i = 1, \dots, N, \quad j = 1, \dots, N, \quad k = 1, \dots, N_k(i, j) \quad (13)$$

where  $N_k(i, j)$  is number of flows serving traffic demand  $h(i, j)$ .

Simplicity of implementation is a very strong advantage of this scheme, but it has two drawbacks.

The first drawback is that coefficients calculation is based only on number of flows that serve traffic demand, without consideration of other traffic demands in network. That is why traffic is unequally distributed in network.

The second drawback of this method is that if one route contains radio links that are temporarily unavailable due to fading, traffic distributed to those links would be lost.

In order to overcome these two drawbacks, we propose traffic distribution optimization by varying balancing factor according to link load  $b(i, j, k)$  to gain either minimum value of  $L_{\max}$  as optimization goal, which we denote as **Opt-L<sub>max</sub>**, or to gain minimum value of  $L_{\text{ave}}$  as optimization goal, which we denote as **Opt-L<sub>ave</sub>**. Due to nonlinearity of **Opt-L<sub>max</sub>**, for its implementation NMinimize function for constrained numerical global optimization is used. NMinimize function implemented in software package Mathematica [13] and it is state of the art combination of several optimization methods like: Nelder-Mead, differential evolution, simulated annealing and random search. Due to very intensive calculation caused by a large number of optimization variables, this algorithm does not have practical value, and it is used just to obtain lower limit value of  $L_{\max}$ . Unlike the first one, the second **Opt-L<sub>ave</sub>** is linear and could be treated as a linear programming problem.



V. PERFORMANCE ANALYSIS OF TRAFFIC DISTRIBUTION ALGORITHMS

A. Network topology

Analysis and performance calculation of proposed methods are performed on test network with topology given in Fig. 1 Network consists of  $N=7$  nodes, connected with  $E=18$  unidirectional links (9 bidirectional links). It is assumed that every node communicates with all other nodes in the network, therefore 42 traffic demands should be served.

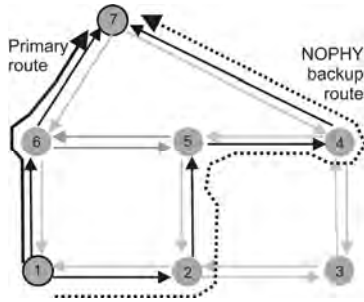


Fig. 1. Network topology and example of NOPHY routes

B. Traffic routing

Primary route that serves specific traffic demand in network is chosen by Dijkstra - shortest path algorithm [14] with link costs inversely proportional to maximum link capacity. It is assumed that full link capacity is  $C_0=1\text{Gbit/s}$  for all links in the network.

In addition to primary route one backup route is defined, for certain number of traffic demands. For performance analysis in this paper, backup routes are calculated using two algorithms: ECMP2 and NOPHY.

For some traffic demands there are several shortest path routes. In this network 18 from 42 traffic demands have multiple shortest path. For each of them two equal cost routes are chosen and traffic is balanced between them. This algorithm is denoted as ECMP2 - Equal Cost Multi Path with two routes.

The second algorithm for backup routes calculation is denoted as NOPHY and it is described in [9], [10]. The key idea of this algorithm is traffic protection in case of link outage, so the backup route is chosen as a shortest path route that has no common links with primary route. Example of NOPHY backup route choice for one traffic demand is shown in Fig. 1. Where multiple shortest path route exists, the one of them is chosen to be primary route, and according to it NOPHY backup route is calculated. In this case each of 42 traffic demands is served by two routes.

C. Network performance in absence of fading

Values of network performance criteria  $L_{max}$  and  $L_{ave}$  in absence of fading, when all links in network operate with maximum capacity of 1Gbit/s are shown in Table 1.

As expected results have shown that, in majority of cases, some gain in load balancing could be achieved by varying traffic distribution coefficients. Also it is shown that Opt- $L_{max}$  give good values both for  $L_{max}$  and  $L_{ave}$ , while Opt- $L_{ave}$  in this case gives result for  $L_{max}$  even worse than equal distribution.

TABLE I. NETWORK PERFORMANCE IN ABSENCE OF FADING

Perform. criterion	Balancing algorithm	BUR type	
		ECMP2	NOPHY
$L_{max}$ [paths / Mbit/s]	Equal	0.0050	0.0070
	Opt- $L_{max}$	0.0040	0.0040
	Opt- $L_{ave}$	0.0080	0.0080
$L_{ave}$ [paths / Mbit/s]	Equal	0.0040	0.0053
	Opt- $L_{max}$	0.0040	0.0040
	Opt- $L_{ave}$	0.0040	0.0040

D. Network performance in presence of fading

For performance analysis in presence of fading, it is assumed that fading cause link capacity reduction of ten fold  $c_R=10$ , that gives new link capacity values of 100Mbit/s. In the case when one link operates with lower capacity, due to network redundancy, performance is not noticeably degraded. Therefore, performance analysis considers the case when two links simultaneously have lower capacity. Other cases when three or more links have degraded capacity, for properly tailored individual link fading margins have very low probability, and therefore are not considered in this analysis.

Two links in network are randomly chosen to have degraded capacity. Results are represented in form of cumulative distribution function of performance criteria in such cases.

E. Performance Analysis Results

In Fig.2. and Fig.3. values of CDF for performance criteria  $L_{max}$  and  $L_{ave}$  respectively are given in case of ECMP2 when two links have ten fold capacity reduction.

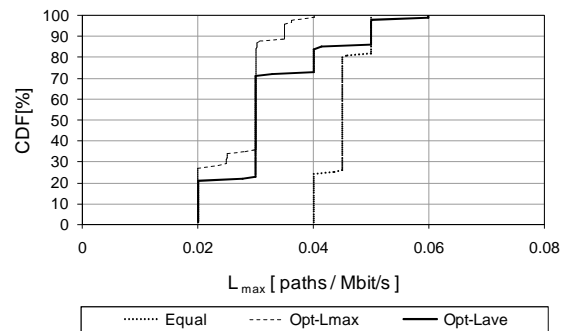


Fig. 2.  $L_{max}$  in case ECMP2 two links have ten fold capacity reduction

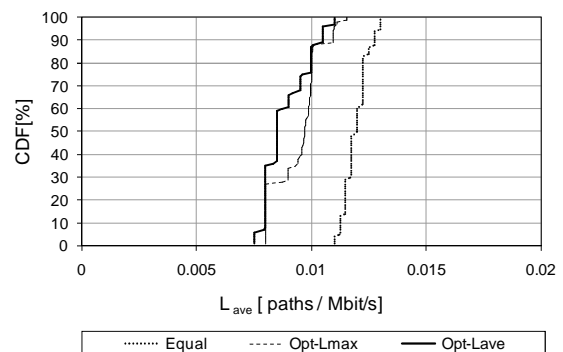


Fig. 3.  $L_{ave}$  in case ECMP2 two links have ten fold capacity reduction

Fig.2. shows that according to  $L_{max}$  criterion Opt- $L_{max}$  has in 42% of cases better results than Opt- $L_{ave}$ , while both

algorithms have considerably better results than equal traffic distribution. Similarly, according to  $L_{ave}$  criterion Opt- $L_{ave}$  has the best results (Fig 3.). Note that differences between CDF curves for Opt- $L_{max}$  and Opt- $L_{ave}$  algorithms are much higher for  $L_{max}$  criterion than for the  $L_{ave}$  criterion.

For NOPHY backup route choice algorithm, compared to ECMP2, the difference between Opt- $L_{max}$  and Opt- $L_{ave}$  CDF curves is greater according to  $L_{max}$  criterion (Fig.4.). According to this criterion Opt- $L_{max}$  has from 25% to 40% better performance according to  $L_{max}$  criterion in 75% of cases, while in 25% of cases the performance is the same. On the other hand, difference between Opt- $L_{max}$  and Opt- $L_{ave}$  CDF curves according to  $L_{ave}$  criterion is only about 10% but in all cases (Fig.5). For NOPHY algorithm both optimization algorithms have about three times better results than equal traffic distribution (Fig.4., Fig.5.).

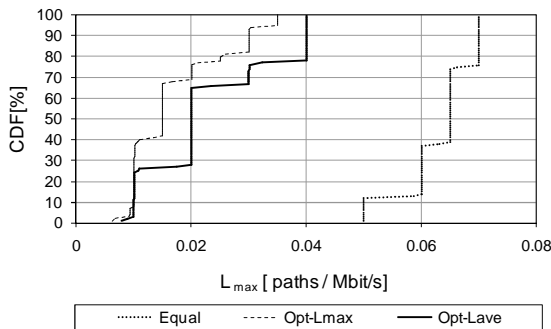


Fig. 4.  $L_{max}$  in case NO PHY two links have ten fold capacity reduction

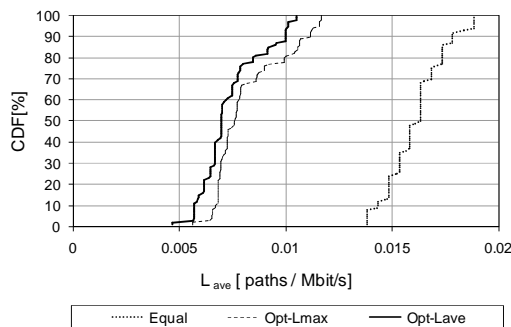


Fig. 5.  $L_{ave}$  in case NO PHY two links have ten fold capacity reduction

VI. REMARKS FOR PRACTICAL IMPLEMENTATION

For practical implementation of traffic distribution coefficient optimization, it is important to have information about current link capacity, which is adjusted to actual fading condition. Typical multipath fading event, known also as fast fading has noticeable signal level changes at about 10ms, while reaction time of adaptive modulation techniques is about 40ms [1][3][4]. For rain attenuation fading [4] signal level changes are much slower about 10s. According to this, for frequency bands above about 18GHz, where rain attenuation is predominant effect, information about link capacities, could be obtained from the link hardware by traffic monitoring protocol such as SNMP [15]. For lower frequency bands more agile technology should be used.

When information about link capacities are obtainable in entire network, since routing tables are known to all routers,

calculation of traffic distribution coefficients could be performed in each router. The straightforward implementation of optimization algorithm, especially for Opt- $L_{max}$  which is nonlinear, could be difficult and time consuming, and therefore it could be subject for future work.

VII. CONCLUSION

Usage of unequal traffic balancing coefficients between main and backup route in IP radio-relay network with path diversity can serve both as traffic protection and load balancing method. Simulation on test network confirmed that unequal traffic distribution gives up to three times better results for link loads than equal traffic distribution. It is also showed that optimizing the maximum of link load, besides from giving the best values of the maximum link load, also gives good performance according average link load criterion. The main drawback of minimizing maximal link load in optimization process is its nonlinear character, which requires much more processing time for its calculation than minimizing average link load.

ACKNOWLEDGEMENT

This work was partially supported by the Ministry of Education and Science of the Republic of Serbia under Grant TR-32024.

REFERENCES

- [1] Wells J., "Multigigabit Wireless Technology at 70 GHz, 80 GHz and 90 GHz", *RF Design*, pp.50-58, May 2006.
- [2] Sklar B. *Digital Communications: Fundamentals and Applications*, Prentice Hall, 2001.
- [3] Xiong F. *Digital Modulation Techniques*, Artech House, 2000.
- [4] *ITU-R Rec. P.530-13*, "Propagation Data and Prediction Methods Required for Design of Terrestrial Line of Sight Systems", 2009.
- [5] *ITU-R F.1703*, "Availability objectives for real digital fixed wireless links used in 27 500 km hypothetical reference paths and connections", 2005.
- [6] *ITU R F.1668*, "Error performance objectives for real digital fixed wireless links used in 27 500 km hypothetical reference paths and connections", 2004.
- [7] Perić D., Perić M., Petrović G., "Redundant Topology in Computer Network Realized with Millimeter Wave Radio Links", *14th IST Mobile and Wireless Summit*, Dresden, June 2005.
- [8] Srivastava V, Motani M., "Cross-Layer Design:A Survey and the Road Ahead", *IEEE Communications Magazine*, Vol.43, pp.113-119, December 2005.
- [9] Perić D., Perić M., Todorović B., "Traffic Protection Method in IP Radio Networks above 70 GHz", *IEEE Communications Letters*, Vol. 14, No. 10, pp. 981-983, October 2010.
- [10] Perić D., *Traffic Protection Method in IP Radio Network above 70 GHz*, PhD Thesis, Faculty of Technical Sciences, University of Novi Sad, June 2011.
- [11] Pioro M., Medhi D., *Routing, Flow, Capacity Design in Communication and Computer Networks*, Morgan Kaufmann, 2004.
- [12] Retana A., Slice D., White R., *Advanced IP network design (CCIE) professional development*, CISCO Press, 1999.
- [13] Wolfram S., *The Mathematica Book*, 5th ed., Wolfram Media, 2003.
- [14] Dijkstra, E.W. "A Note on Two Problems in Connection with Graphs", *Numerische Mathematik* 1, 269-271., 1959.
- [15] *RFC 1157*, "Simple Network Management Protocol (SNMP)", IETF, May 1990.

# Optical Receiver Sensitivity Evaluation in Presence of Noise in Digital Communication System

Krasen Angelov<sup>1</sup>, Stanimir Sadinov<sup>2</sup>, Nataliya Varbanova<sup>3</sup>

**Abstract** – The performance of an optical receiver in a digital optical communication link is studied. In the design of an optical receiver, it is vital that the module is capable of converting and shaping the optical signal while meeting or surpassing the maximum BER. Ultimately, the noise influence on the signal will determine the system sensitivity. The challenge is to find a way to determine the QoS of an optical transmission channel independent of data format and bit rate within a short time frame. The  $Q$ -factor itself significantly reduces measurement time and is thus more cost-effective. The analysis is based, assuming an input signal with impairment from factors like inter-symbol interference, jitter, and transmitter relative intensity noise.

**Keywords** – optical receiver sensitivity, bit error rate, inter-symbol interference, transimpedance and limiting amplifier.

## I. INTRODUCTION

Most applications of optical signals in digital communications require the detection and subsequent conversion of the light to an electrical signal. In this process, the useful signal will be corrupted by noise and the ultimate sensitivity and performance of the system is limited by the noise characteristics. Optical receiver adds noise; usually thermal noise and shot noise. In communication systems, where electrical, radio or optical signals are transmitted; noise can be viewed as an impairment resulting in the degradation of the information contained in the signal [1,7]. Optical amplifiers can be used to improve the effective receiver sensitivity in optical systems. The optical amplifier works on the principle of stimulated emission [5]. The optical amplifiers add noise to the amplified signal, and at some point, this noise becomes the dominant noise source. The basic manifestation of noise in optical amplifier is in the form of amplified spontaneous emission (ASE). So, the bit error probability (BER) is also affected by the ASE noise added by the optical amplifier [2,3,7].

A typical optical receiver is composed of an optical photo detector, a transimpedance amplifier (TIA) [9], a limiting amplifier (LA), and a clock-data recovery (CDR) block. Fig. 1

shows a simple block diagram of the front end of an optical receiver module.

The received optical signal is first converted into photocurrent and amplified by the TIA. The limiting amplifier (LA) acts as a “decision” circuit, where the sampled voltage  $v(t)$  is compared with the decision threshold  $v_{TH}$ . At this data decision point, the signal is significantly degraded by the accumulation of random noise and inter-symbol interference (ISI), resulting in erroneous decisions due to eye closure [3,4].

## II. CHARACTERISTICS OF PERFORMANCE ANALYSIS

### A. Eye diagram

The eye diagram represents a superposition of all bits in the signal on top of each other [8]. Fig. 2 shows the eye diagram of a NRZ signal. There are two basic types of adverse effects visible in the eye diagram: First, the effect of intersymbol interference (ISI) and, second, the effect of jitter. The ISI is caused by overlap of individual modulation pulses and it leads to the amplitude errors at the sampling instances. The jitter is defined as short-time deviations of a digital signal from its ideal position in time [8]. A larger “eye opening” signifies less noise or distortion and therefore a higher quality of signal.

### B. Bit Error Ratio (BER)

In digital communication systems, the decision when to sample and whether the sampled value represents a binary 1 or 0 is affected by noise and signal distortion in the real system and there is nonzero probability of an erroneous decision. Therefore, the received signal quality is directly related to the bit error rate (BER), which is a major indicator of the quality of the overall system [8]. Eq. (1) explains a calculation bit error rate if the  $Q$ -value is known [4,6]:

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

### C. $Q$ -factor

The  $Q$ -factor expresses the quality of an optical signal with respect to its signal-to-noise ratio (SNR). It includes all physical impairments of the signal, such as noise, non-linear effects, dispersion (chromatic and polarization). These impairments degrade the signal and cause bit errors. Consequently, a higher value of the  $Q$ -factor means a better SNR and therefore

<sup>1</sup>Krasen Angelov is with the Faculty of Electrical Engineering and Electronics, Technical University of Gabrovo, 4 H. Dimitar St., 5300 Gabrovo, Bulgaria, E-mail: kkangelov@tugab.bg

<sup>2</sup>Stanimir Sadinov is with the Faculty of Electrical Engineering and Electronics, Technical University of Gabrovo, 4 H. Dimitar St., 5300 Gabrovo, Bulgaria, E-mail: murry@tugab.bg

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

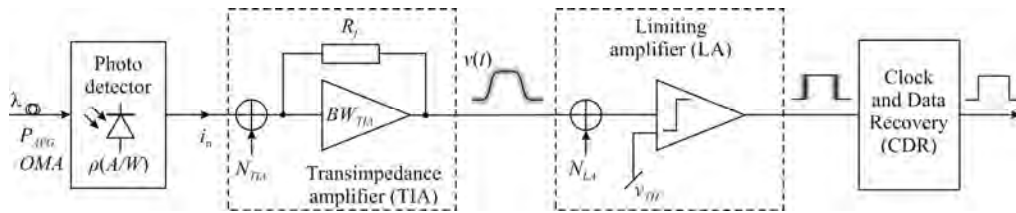


Fig. 1. Simplified block diagram of the optical receiver module

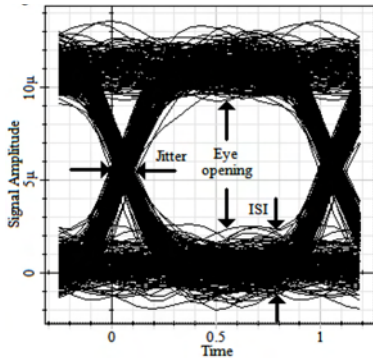


Fig. 2. An example of an eye-diagram and its interpretation

$$OMA_{min} = \frac{i_n SNR}{\rho} \quad (4)$$

In communication systems noise can be viewed as an impairment resulting in the degradation of the information contained in the signal [7]. The optical receiver adds two types of noise namely thermal noise and shot noise. Since optical amplifiers are based on the principle of stimulated emission, its main contribution to noise is ASE noise.

a lower BER. Eq. (2) gives the Q-factor of an optical signal [4,6]:

$$Q = \frac{V_1 - V_0}{\sigma_1 + \sigma_0} \quad (2)$$

where  $V_1$  is the value of the binary 1,  $V_0$  is the value of the binary 0,  $\sigma_1$  is the standard deviation of the binary 1 and  $\sigma_0$  is the standard deviation of the binary 0.

### III. OPTICAL RECEIVER SENSITIVITY EVALUATION

In optical communication systems, sensitivity is a measure of how weak an input signal can get before the bit-error ratio (BER) exceeds some specified number.

Sensitivity can be expressed as average power ( $P_{AVG}$ ) in dBm or as optical modulation amplitude (OMA) in  $W_{p,p}$  (peak-to-peak). Each gives a figure of merit for the receiver [3,4,6].

To achieve the best optical sensitivity, it is important to maximize the signal Q-factor before data decision. The equation for calculating sensitivity is as follows [4,6]:

$$P_{AVG} = 10 \log \left( \frac{i_n SNR (r_c + 1)}{2\rho(r_c - 1)} 1000 \right), \text{ dBm}, \quad (3)$$

where  $i_n$  is the noise of TIA;  $\rho$  – responsivity flux (conversion efficiency) of the photodetector, in A/W;  $r_c$  – the ratio of a logic-one power level ( $P_1$ ) relative to a logic-zero power level ( $P_0$ ) [4,7].

The process in estimating the minimum peak-to-peak swing of the optical signal begins with the choice of the maximum BER. This determines the signal-to-noise ratio (SNR). Next, the RMS input referred noise ( $i_n$ ) of the TIA and the responsivity ( $\rho$ ) of the photodetector must be found from the vendor’s data sheets. These are related by:

#### A. Thermal noise

The thermal noise of a receiver arises from the fact that electrons in a receiver circuit have some probability of generating a current even in the absence of an optical signal. This noise (referred to as Johnson noise), can be represented by the variance of thermal current per unit frequency [5]:

$$\sigma_{th}^2 = 4kT / R, \quad (5)$$

where  $T$  is the absolute temperature,  $k$  is Boltzmann’s constant and  $R$  is the detector load resistance.

#### B. Shot noise

The shot noise arises from the Poisson distribution of the electron-hole generation by the photon stream. The latter is a stochastic process having random arrival times. On average, the number of electron-hole pairs created will be proportional to the number of photons, with a given constant of proportionality.

During a given time interval, with a certain number of photons incident upon the detector, the number of electron-hole pairs generated will have fluctuations as determined by Poisson statistics [5]. A dc photocurrent of  $I_{pd}$  will generate a shot noise power density of:

$$\sigma_{sh}^2 = 2eI_{pd} \quad (6)$$

#### D. ASE Noise

ASE is produced by spontaneous emission that has been optically amplified by the process of stimulated emission in gain medium. Noise associated with ASE is the limiting factor in determining the ultimate signal-to-noise ratio in any system using optical amplifiers [3,4,5]. The output ASE power can be calculated using classical derivation in:

$$P_{ASE} = n_{sp} (G - 1) h\nu B_o, \quad (7)$$

where  $h$  is Plank's constant,  $\nu$  is the optical frequency of transition,  $B_o$  is the optical bandwidth and  $n_{sp}$  is the inversion parameter, given by

$$n_{sp} = \frac{\sigma_e(\lambda)N_2}{\sigma_e(\lambda)N_2 - \sigma_a(\lambda)N_1}, \quad (8)$$

where  $\sigma_a(\lambda)$  and  $\sigma_e(\lambda)$  are the absorption and emission cross sections, respectively;  $N_1$  and  $N_2$  are the population density in lower and upper states;  $G$  is the overall gain of the amplifier.

Eq. (7) gives the ASE power for one polarization mode. So, for single mode fiber, the right hand side of Eq. (7) must be multiplied by a factor of 2.

When an amplified optical signal and accompanying spontaneous emission are detected in a photodetector, the noise is transformed into the electrical domain and appears along with the induced photocurrent as a noise current. Photodetection is a nonlinear square-law process [5]. The photocurrent is therefore composed of a number of beat signals between the signal and noise optical fields  $E_S$  and  $E_N$ , respectively, in addition to the squares of the signal field and spontaneous emission field. The photocurrent  $I_{pd}$  is found as:

$$I_{pd} \propto (\vec{E}_{tot})^2 = (\vec{E}_S + \vec{E}_N)^2 = E_S^2 + E_N^2 + 2\vec{E}_S \cdot \vec{E}_N. \quad (9)$$

In Eq. (9), one can identify the first term as pure signal, the second term as pure noise and it is referred to as spontaneous-spontaneous (sp-sp) beat noise, and the third term as mixing component between signal and noise and it is referred to as signal-spontaneous (s-sp) beat noise [3,6].

The power spectrum of current corresponding to s-sp beat noise is uniform in the frequency interval  $(-B_o/2)$  to  $(B_o/2)$  and has an equivalent one-sided power density of

$$N_{s-sp} = \frac{4e^2}{h\nu} P_{in} n_{sp} (G-1)G. \quad (10)$$

The power spectrum of current corresponding to spontaneous-spontaneous beat noise extends from 0 to  $B_o$  with a triangular shape and a single-sided power density near dc of

$$N_{sp-sp} = 2n_{sp}^2 (G-1)^2 e^2 B_o. \quad (11)$$

### E. SNR calculation

The  $Q$ -factor can be also expressed in terms of the optical signal-to-noise ratio ( $OSNR$ ) as:

$$Q = \sqrt{\frac{B_o}{B_e} \frac{2OSNR}{4\sqrt{OSNR+1}+1}}, \quad (12)$$

with  $B_o$  and  $B_e$  the optical and electrical bandwidths, respectively. Eq. (12) seems nonlinear and can be used to derive the  $OSNR$  needed to obtain a given  $BER$ , for an ideal system with only amplifier noise and without nonlinearities or inter-symbol interference.

In case of amplifier operating with moderately large optical input signals, the SNR at the amplifier output is dominated by signal spontaneous beat noise. In this limit, the equivalent electrical SNR at the amplifier output is given by [5]

$$SNR = \frac{GP_{in}}{4h\nu n_{sp} (G-1)B_e}. \quad (13)$$

Provided that  $G$  is reasonably high, the  $SNR$  is determined only by the input power and the inversion parameter  $n_{sp}$ . More specifically, the  $SNR$  is independent of the gain. This is an important result that governs the system performance and determining of optical receiver sensitivity using Eq. (3):

### F. CDR jitter-tolerance penalty

As the signal goes through the receiver amplifier chain to the limiting stage, the amplitude noise is converted into timing jitter at the data midpoint crossing. Random and deterministic jitter is generated due to the existence of random noise, limited bandwidth, passband ripple, group-delay variation, AC-coupling, and nonsymmetrical rise/fall times. The combination of these jitter components decreases the eye opening available for error-free data recovery. Consequently, CDR jitter-tolerance capability is another critical factor for determining optical sensitivity. CDR jitter tolerance is a measure of how much peak-to-peak jitter can be added to the incoming data before errors occur due to misalignment of the data and recovered clock. For a PLL-based CDR design, a minimum data-eye opening is required, which is determined by the clock-to-data sampling position, the retiming flip-flop setup/hold times, and the phase detector characteristics. Assuming that the random jitter is  $RJ_{RMS}$ , the total deterministic jitter is  $DJ_{P-P}$ , and the CDR minimum required eye opening is  $T_{OPEN}$  at a specified  $BER$ , then the timing  $Q$ -factor is defined as:

$$Q = \frac{T_b - T_{OPEN} - DJ_{P-P}}{2RJ_{RMS}}, \quad (14)$$

$$RJ_{P-P} = 2Q_{BER} RJ_{RMS}. \quad (15)$$

When the jitter frequency at the CDR input is higher than the PLL bandwidth, the CDR jitter tolerance (noted as  $JT_{P-P}$ ) is related to  $T_{OPEN}$  as:

$$JT_{P-P} = T_b - T_{OPEN}. \quad (16)$$

To avoid degrading the optical sensitivity, the CDR high frequency jitter tolerance should satisfy:

$$JT_{P-P} \geq 2Q_{BER} RJ_{RMS} + DJ_{P-P}. \quad (17)$$

At the optical receiver input, it is assumed that the TIA is linear before the limiting amplifier. Therefore, random jitter can be expressed as a function of the peak-to-peak current to the total RMS noise ratio at TIA input:

$$RJ_{RMS} = \frac{t_r}{I_{P-P} / N_{TOTAL} 0,6}, \quad (18)$$

where  $t_r$  is dependent on the overall receiver small-signal bandwidth  $BW_{TOTAL}$ . Assuming a first-order lowpass filter:

$$t_r \approx 0,22 / BW_{TOTAL}. \quad (19)$$

The CDR jitter-tolerance penalty on optical sensitivity can be estimated by combining Eqs. 17 and 18, then solving for  $I_{p-p}$  as:

$$I_{p-p} = \frac{2Q_{BER} \cdot t_r \cdot N_{TOTAL}}{(JT_{p-p} - DJ_{p-p}) \cdot 0,6} \quad (20)$$

Substituting the Eq. (20) in Eq. (3) the optical receiver sensitivity in terms of  $P_{AVG}$  can be obtained.

#### IV. RESULTS

Examples are given for optical receiver using MAXIM devices MAX3277 TIA and MAX3272 LA. The datasheet parameters are as follows:  $N_{TIA} = 0,35\mu A$ ,  $R_f = 3,3k\Omega$ ,  $N_{LA} = 0,22mV$ , the total receiver small-signal bandwidth is  $7.0GHz$ , and  $ISI = 0$ . Assuming  $r_e = 6,6$  and  $\rho = 0,85A/W$ , the calculated optical sensitivity ( $P_{AVG}$ ) versus optical signal-to-noise ratio (OSNR) is shown in Fig. 3. The results on Fig. 3 are based on Eqs. (3), (13) and (20). It is shown the minimum required optical sensitivity  $P_{AVG}$  for a given SNR. For example, when the  $SNR = 14,1dB$  (which is equivalent to  $BER = 10^{-12}$  or  $Q_{BER} = 7,1$ ) the optical sensitivity is  $-21,78dBm$  in ideal case,  $-14,98dB$  considering optical link with amplifier noise, and  $-14,59dB$  considering the jitter-tolerance penalty.

Another useful representation of minimum required optical sensitivity  $P_{AVG}$  is the dependence of  $BER$  needed. The results are shown in Fig. 4.

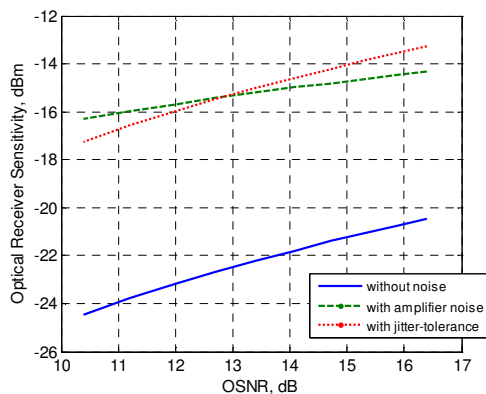


Fig. 3. Optical receiver sensitivity versus optical signal-to-noise ratio

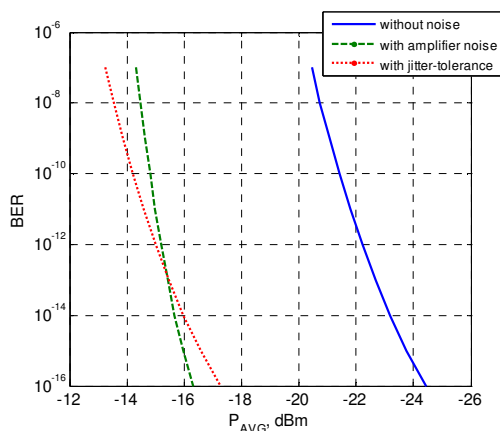


Fig. 4. Bit Error Rate versus minimum required optical receiver sensitivity ( $P_{AVG}$ )

From Figs.3 and 4 it is evident that for OSNR lower than  $12,75dB$  a dominant factor is the amplifier noise when determining the optical receiver sensitivity. For higher OSNR dominant is the jitter-tolerance penalty. In general, to achieve a specified  $BER$ , the minimum TIA input current should satisfy the  $Q_{BER}$  in both amplitude and timing.

#### V. CONCLUSION

By applying the technique presented in this paper, it is easy to estimate and predict more realistic optical receiver sensitivity. It is necessary to consider error sources in both amplitude and timing. It has been shown how the amplitude and timing-error sources separately affect the overall receiver  $BER$ . A general expression of  $Q$ -factor can be predicted when shot noise and thermal noise are also considered. Using these guidelines, optical receiver performance can be accurately predicted. In reality, the optical input is not an ideal signal, because it suffers random noise from the transmitter as well as ISI from fiber dispersion. The approach presented in this article can be used for estimating the signal  $Q$ -factor and, therefore, determining the  $BER$ .

#### ACKNOWLEDGEMENT

This paper has been sponsored under the auspices of the "Increasing Efficiency and Quality of Service in PBX (TU – Gabrovo)" project – a part of the University Center for Research and Technology (UTzNIT) at the Technical University of Gabrovo, contract E1102/2011.

#### REFERENCES

- [1] O. Panagiev, Adaptive Compensation of the Nonlinear Distortions in Optical Transmitters Using Predistortion. Radioengineering, vol. 17, № 4, Dec. 2008, pp. 55-58.
- [2] J. Barry, E. Lee, "Performance of Coherent Optical Receivers", Proceedings of IEEE, Vol. 78, No. 8, August 1990.
- [3] K. Angelov, K. Koitchev, S. Sadinov, An Investigation of Noise Influences in Optical Transmitters and Receivers in Cable TV Networks, ICEST 2006, Proceedings of Papers, pp.102-105, Sofia, Bulgaria, 2006.
- [4] K. Koitchev, K. Angelov, S. Sadinov, Determining Bit Error Rate in Digital Optical Transmission Network Using the Q-Factor, ICEST 2010, Proc. of Papers, Vol. 1, pp.53-56, ISBN: 978-9989-786-57-0, Ohrid, Macedonia, 2010.
- [5] P. Becker, N. Olsson, Erbium-Doped Fiber Amplifiers: Fundamentals and Technology, Acad. Press, New York, 1999.
- [6] R. Freeman, Fiber-Optic Systems for Telecommunications, John Wiley & Sons, New York, 2002.
- [7] S. Derevyanko, S. Turitsyn, "Bit Error Probability for Direct Detection of Optical RZ Signal Degraded by ASE Noise and Timing Jitter," IEEE Lightwave Technol., vol.25, pp.638-643, 2007.
- [8] V. Tejkal, M. Filka, J. Šporik, P. Reichert, Possibilities of Increasing Power Budget in Optical Networks, ElektroRevue, vol.1, №4, December 2010, ISSN 1213-1539.
- [9] B. Karapenev, Analogue Circuits and Systems. Methodological Handbook for Course Design. Publishing house „M-PRES”, ISBN 978-954-8455-47-3, pp. 28-36, 2012.

# New Teletraffic Loss System - Polya/G/n/0

Seferin Mirtchev<sup>1</sup>, Rossitza Goleva<sup>1</sup>, Georgi Balabanov<sup>1</sup> and Velko Alexiev<sup>1</sup>

**Abstract** – In this paper we have proposed to use Polya and Pareto distribution to describe peaked arrival processes. With this arrival processes it is possible to analyze the real loss systems in telecommunication networks. We study a version of the “classical” Erlang’s loss model M/G/n/0 called Polya/G/n/0. This is a full accessible loss system with Polya input flow (negative binomial distributed number of arrivals in a fixed time interval), general distributed service time and “n” servers. This model is evaluated by simulation with Pareto distributed inter-arrival time. An algorithm for calculation of the state probabilities and blocking are presented. It is shown that the variance of the input stream changes significantly the characteristics of the loss systems.

**Keywords** – Polya distribution, Pareto distribution, loss system, peaked flow.

## I. INTRODUCTION

The most common choice for telecommunication network design is based on the exponential assumption. Usual choice is the Poisson arrival of the calls or sessions and exponential service times. However, networks and applications of today generate a traffic that is bursty over a wide range of time scales. A number of empirical studies have shown that the network traffic is self-similar or fractal in nature.

There is no single traffic model that can efficiently capture the traffic characteristics of all types of networks. The study of traffic models for a specific environment has become a crucial and important task. Good traffic modelling is also a basic requirement for accurate capacity planning.

In [1] a careful overview of some of the widely used network traffic models is made, highlighting the core features of the model and traffic characteristics they capture. For heavy-tailed traffic it can be shown that the Poisson model underestimates the traffic. In case of high-speed networks with unexpected demand on packet transfers, Pareto distribution is a good candidate since the model takes into consideration the long-term correlation of packet arrival times.

In this work, we propose to use the Polya and Pareto distribution to describe the input flows. We study a version of the “classical” Erlang’s loss model M/G/n/0 called Polya/G/n/0. This is a full accessible loss system with Polya input flow (negative binomial distributed number of arrivals in a fixed time interval), general distributed service time and “n” servers.

The Polya arrival process is peaked process. It is defined

<sup>1</sup>Seferin Mirtchev, Rossitza Goleva, Georgi Balabanov and Velko Alexiev are with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mails: stm@tu-sofia.bg, rig@tu-sofia.bg, gbalabanov@ieec.org, velko.alexiev@gmail.com

by two parameters – the mean value and the variance [2]. This way we present a basic study of the influence of the offered traffic’s peakedness on the blocking probability.

The MMPP (Markov Modulated Poisson Process) traffic model is proposed in [3]. It accurately approximates the long range dependence characteristics of Internet traffic. The hidden Markov models (D-BMAP/D/1/k) is presented in [4] with an arrival process of frames and service process in the wireless channel. The method for analysis of the M/D/n is given in [5]. A simple accurate model for a multi-server central-queue system M/G/K when service requirements have heavy tails is considered in [6]. The bursty Internet traffic stream is studied in [7,8] as an ON/OFF model.

The main reason for studying the Polya/G/n/0 teletraffic system is that it can be used to analyze the real loss systems in telecommunication networks. We consider that the network analysis requires a technique that can represent any kind of traffic, and especially peaked.

## II. POLYA ARRIVAL PROCESS

The Polya arrival process is a pure birth process with two parameters [9]. The probability  $P_i(t)$  of  $i$  arrivals in an interval with duration of  $t$  seconds is given by

$$P_o(t) = (1 + \beta\lambda t)^{\frac{1}{\beta}}$$

$$P_i(t) = \binom{1/\beta + i - 1}{i} \left( \frac{\beta\lambda t}{1 + \beta\lambda t} \right)^i P_o(t), \quad (1)$$

where  $\lambda > 0$  and  $\beta > 0$ .

The Polya distribution is a variant of the negative binomial distribution. Its mean value (the average number of arrivals in an interval of length  $t$ ) is

$$M(t) = \sum_{i=1}^{\infty} iP_i(t) = \lambda t. \quad (2)$$

This means that  $\lambda$  is an arrival rate.

The variance of the number of arrivals in an interval of length  $t$  is

$$V(t) = \sum_{i=0}^{\infty} [i - M(t)]^2 P_i(t) = \lambda t(1 + \beta\lambda t). \quad (3)$$

The peakedness of the Polya input flow is

$$z(t) = \frac{V(t)}{M(t)} = 1 + \beta\lambda t > 1. \quad (4)$$

When  $\beta = 0$ ,  $M(t) = V(t) = \lambda t$  i.e. it is a regular Poisson process. When  $\beta = 1$  the Polya distribution is a geometric distribution.

III. MODEL DESCRIPTION

Let us consider a multi servers loss system Polya/D/∞ with a Polya input stream with arrival rate λ, constant service time τ and infinite number of servers. This queueing system is a non-Markovian model or renewal process (Fig. 1).

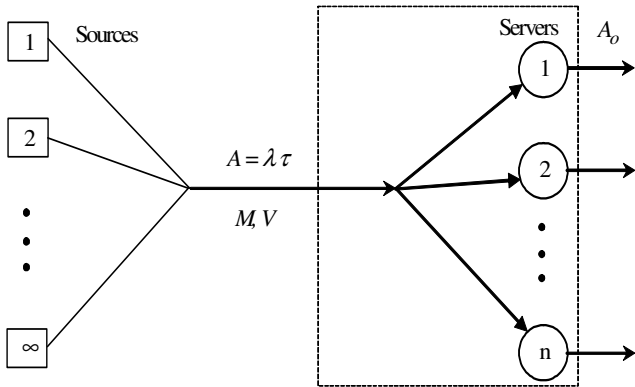


Fig. 1. Generalised queueing model with peaked input flow.

To study this system, we consider two epochs (points of time)  $t$  and  $t + \tau$  at a distance of  $\tau$ . Every customer being served at epoch  $t$  has left the server at epoch  $t + \tau$ . The numbers of the customers arriving during the interval  $(t; t + \tau)$  are still in the system at epoch  $t + \tau$  (being served) and they are described whit Polya distribution as the arrival process.

For this model with Polya arrival process the offered traffic is equivalent to the average number of call attempts in service time:

$$A = \lambda \tau, \text{ erl.} \tag{5}$$

The offered traffic is equal to the carried traffic because there are no losses and delays and it is called intended traffic.

The states probability of the system under the assumption of statistical equilibrium can be calculated by means of Polya distribution because we choose arbitrary epoch to observe the system:

$$P_o = (1 + \beta A)^{-\frac{1}{\beta}}$$

$$P_i = \binom{1/\beta + i - 1}{i} \left( \frac{\beta A}{1 + \beta A} \right)^i P_o \tag{6}$$

The number of busy channels at a random point of time is thus Polya distributed with both mean value equal to the offered traffic  $A$  and different variance which depends from the chosen parameter  $\beta$ .

We still assume the same arrival process. The number of channels is now limited so that  $n$  is finite. The number of states becomes  $n+1$ . In this case we get the truncated Polya distribution:

$$P_i = \frac{\binom{1/\beta + i - 1}{i} \left( \frac{\beta A}{1 + \beta A} \right)^i}{\sum_{k=0}^n \binom{1/\beta + k - 1}{k} \left( \frac{\beta A}{1 + \beta A} \right)^k} \tag{7}$$

The name truncated is due to the fact that the solution may be interpreted as a conditional Polya distribution  $Pi(i \leq n)$ . The fact that we are allowed to truncate the Polya distribution means that the relative ratios between the state probabilities are unchanged. We prove this fact by simulation.

IV. TRAFFIC CHARACTERISTICS OF POLYA/D/N/O SYSTEM

Knowing the state probabilities, we are able to find all performance measures defined by state probabilities.

Time Congestion

The probability that all  $n$  channels are busy at a random point of time is obtained from system equations when  $i = n$ :

$$B_n = \frac{\binom{1/\beta + n - 1}{n} \left( \frac{\beta A}{1 + \beta A} \right)^n}{\sum_{k=0}^n \binom{1/\beta + k - 1}{k} \left( \frac{\beta A}{1 + \beta A} \right)^k} \tag{8}$$

Carried traffic

By definition the carried traffic is:

$$A_n = \sum_{i=1}^n i P_i, \text{ erl.} \tag{9}$$

Utilization

Traffic carried by one channel:

$$\rho = \frac{A_n}{n}, \text{ \%}. \tag{10}$$

Traffic congestion

The ratio of traffic lost and intended traffic:

$$B_a = \frac{A - A_n}{A} \tag{11}$$

There are different time and traffic congestion because the arrival process is peaked.

Insensitivity

A system is insensitive to the holding time distribution if the state probabilities of the system only depend on the mean value of the holding time.

It can be shown with Cox distribution that loss formula (11), which above is derived under the assumption of constant service time, is valid for arbitrary service time distributions. The state probabilities for both the Polya distribution and the truncated Polya distribution only depend on the service time



distribution through the mean value which is included in the offered traffic A. We prove this property by simulation.

It can be shown that all classical loss systems with full accessibility are insensitive to the holding time distribution [5].

### V. SIMULATION WITH PARETO DISTRIBUTION

The fundamental relationship between the number and interval representations is given by Feller-Jensen's identity:

$$P\{N_t < n\} = P\{T_n \geq t\}, \quad n = 1, 2, 3, \dots \quad (12)$$

where:  $N_t$  is the random variable for the number of calls arrived in time interval  $t$

$T_n$  is the random variable for the time interval until there has been  $n$  arrivals.

Simulations are the main tools for studying the performance of telecommunication networks and we will evaluate the Polya/D/ $\infty$  and Polya/G/n/0 queue using simulation.

The family of Generalized Pareto Distributions (GPD) has three parameters: the location parameter  $\mu$ , the scale parameter  $\sigma$  and the shape parameter  $\xi$ . If we choose the location parameter  $\mu = 0$  then the cumulative distribution function of the GPD is:

$$F(x) = 1 - \left[ 1 + \frac{\xi x}{\sigma} \right]^{-\frac{1}{\xi}}, \quad (13)$$

where  $\sigma > 0$  and  $\xi > 0$ .

The mean value of the generalized-Pareto distribution is:

$$M_p = \frac{\sigma}{1 - \xi}. \quad (14)$$

When  $\xi < 1$  the mean value is finite.

The variance of the Generalized Pareto Distribution is:

$$D_p = \frac{\sigma^2}{(1 - \xi)^2 (1 - 2\xi)}. \quad (15)$$

When  $\xi < 0.5$  the variance is finite.

The probability density function of the GPD is:

$$f(t) = \frac{1}{\sigma} (1 + \lambda) \left( 1 + \frac{\xi x}{\sigma} \right)^{-\frac{1}{\xi} - \xi} \quad (16)$$

We choose these substitutions

$$\sigma = \frac{1 - \beta}{\lambda(1 + 2\beta)}; \quad \xi = \frac{\beta}{1 + 2\beta}. \quad (17)$$

Therefore, we receive another form of the mean value of the generalized-Pareto distribution:

$$M_p = \frac{1}{\lambda} \quad (16)$$

The mean value is the average inter-arrival time for our study. The parameter  $\lambda$  is the average call arrival intensity.

The variance of the Generalized Pareto Distribution is:

$$D_p = \frac{1 + 2\beta}{\lambda^2} \quad (17)$$

With these substitutions for every positive value of  $\lambda$  and  $\beta$  the mean and variance of the GOD will be finite.

#### Random number generation

Many programming languages do not yet recognize the Pareto distribution. In the field of telecommunications, the Pareto distribution is widely used to estimate the inter-arrival and service times.

One can easily generate a random sample from Pareto distribution by using inverse distribution function. Given a random variable  $U$  with uniform distribution on the unit interval (0;1), the random variable  $x$  is Pareto-distributed.

$$x = \frac{\sigma(U^{-\xi} - 1)}{\xi} \quad (18)$$

We have developed a real time trace simulation algorithm for evaluating the state probabilities of the Polya/D/ $\infty$  system and make a comparison with M/D/ $\infty$ . The simulation results have shown that with Pareto distribution inter-arrival time the state probabilities of the Polya/D/ $\infty$  system have the same mean and variance as the Polya distribution. When there are truncations (Polya/D/n/0) or we change the service time distribution (exponential and Pareto) the state distribution by simulation is identical with Poly's formula (7) and (6).

### VI. NUMERICAL RESULT

In this section, we give numerical results obtained by a Pascal program on a personal computer. The described methods are tested over a wide range of arguments.

Figure 2 illustrates the stationary probability distribution in a loss system Polya/G/n/0 with a Polya input stream with different peakedness  $z$ , 15 erl offered traffic and 50 servers. It is seen that when the peakedness increases the probability that all servers are busy increase vastly. Figure 3 shows the traffic congestion as a function of the intended traffic for various values of the peakedness when the number of the servers is 30. The peaked input flow increases the congestion with several orders. Figure 4 presents the utilisation of the servers (the carried traffic for one server) as function of the intended traffic intensity with different peakedness of the Polya input flow when the number of the servers is 20. The peakedness of the arrival stream decreases the utilisation and in the same time increases the congestion.

It is shown that the influence of the variance of the input stream over the performance measures is significant and can be easily evaluate by Polya and Pareto distribution.

VII. CONCLUSION

In this paper the Polya distribution is used to describe the peaked arrival processes in telecommunication networks. A basic generalised multiple loss teletraffic model Polya/G/n/0 with peaked input flow, generalise service time and full accessibility is investigated. All performance measures of interest are estimated. The idea is based on the Polya distribution as an analytical continuation of the Poisson distribution and the classical M/G/n/0 system.

The proposed approach provides a unified framework to model peaked teletraffic in real telecommunication systems. Numerical results and subsequent experience have shown that this approach is accurate and useful in analyses of traffic systems and especially in Quality of Service and performance parameters estimation.

The importance of this multiple loss teletraffic system with peaked input stream comes from its ability to describe behaviour that is to be found in complex real time queueing systems. It is a general traffic system which is important in designing telecommunication networks.

ACKNOWLEDGEMENTS

This paper is sponsored by project "DVU\_10\_0271 "Public and Private Multimedia Network Throughput Increase by Creating Methods for Assessment, Control, and Traffic Optimization" with National Science Fund, Bulgarian Ministry of Education and Science, 2010 – 2013 led by Prof. Ph.D. Vladimir Poulkov.

REFERENCES

- [1] Chandrasekaran B. "Survey of Network Traffic Models". This report is available on-line at [http://www.cse.wustl.edu/~jain/cse567-06/traffic\\_models3.htm](http://www.cse.wustl.edu/~jain/cse567-06/traffic_models3.htm)
- [2] Ramos H., D. Almorza and J. Garcia-Ramos. "On Characterizing the Polya Distribution". *ESAIM: Probability and Statistics* June 2002, Vol. 6, pp.105-112, URL: <http://www.emath.fr/ps/>
- [3] Muscariello L., M.Mellia, M.Meo, M.Ajmone Marsan, R. Lo Cigno, "An MMPP-Based Hierarchical Model of Internet Traffic", *IEEE ICC 2004*, Vol. 27, no. 1, June 2004, pp. 2143-2147.
- [4] Moltchanov D., Y. Koucheryavy, J. Harju. "Non-preemptive  $iD$ -BMAP $iD/1/K$  queueing system modelling the frame transmission process over wireless channels", *ITC 19*, Beijing, China, Vol. 6a, 2005, pp. 1335-1344.
- [5] Iversen, V. B. "Teletraffic Engineering Handbook", *ITU-D & ITC*. 312 pp. Edition spring 2004, <http://www.com.dtu.dk/education/34340/>.
- [6] Psounisa K., P. Molinero-Fernandez, B. Prabhakar, F. Papadopoulos. "Systems with Multiple Servers Under Heavy-Tailed Workloads". *Performance Evaluation* 62, 2005, pp. 456-474.
- [7] Yang X. "Designing Traffic Profiles for Bursty Internet Traffic". *Proceedings of IEEE Global Internet*, 2002.
- [8] Siriwong K., L. Lipsky, R. Ammar, "Study of Bursty Internet Traffic", *Sixth IEEE International Symposium on Network Computing and Applications (NCA 2007)*, 2007, pp.53-60.

[9] Korn G. and T. Korn. "Mathematical Handbook for Scientists and Engineering", *McGraw Hill*, 1968, pp. 575.

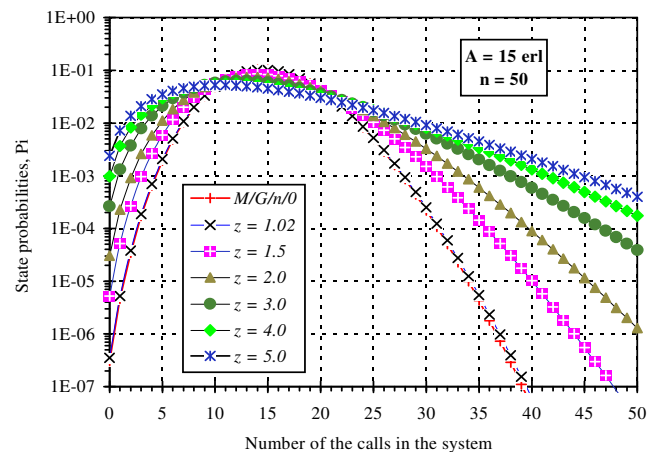


Fig. 2. Stationary probability distribution

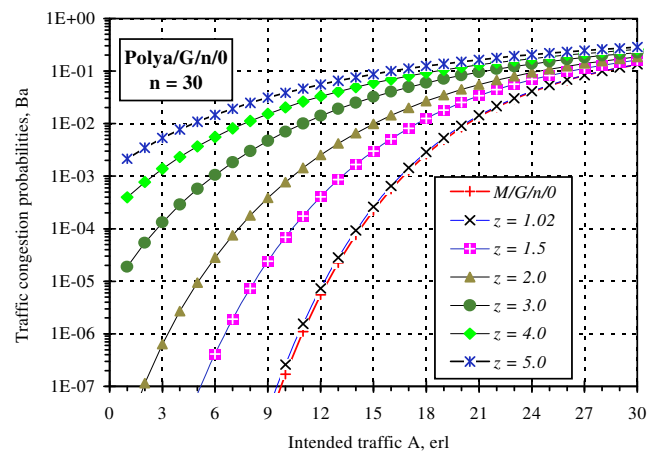


Fig. 3. Traffic congestion as a function of the intended traffic for various values of the peakedness.

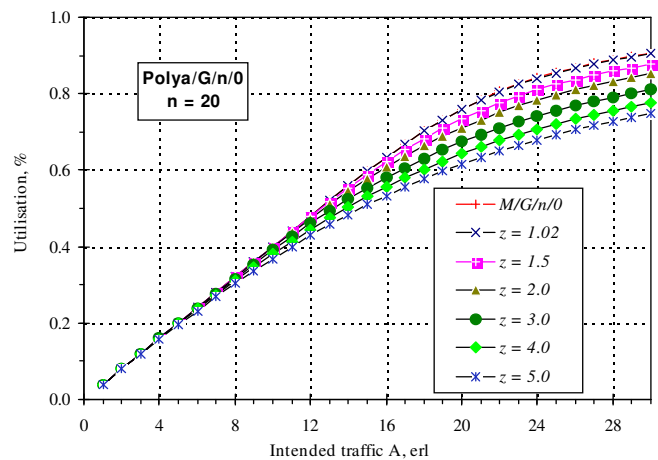


Fig. 4. Utilization of servers

# An Evaluation of an UMTS/WLAN Interworking Architecture using IEEE 802.21

Alexandru Vulpe<sup>1</sup> and Octavian Fratu<sup>2</sup>

**Abstract** – This paper proposes an interworking architecture for mobility between UMTS/HSDPA and WLAN access networks based on the IEEE 802.21 (Media Independent Handover - MIH) standard. This architecture uses MIH signalling and blocks to provide information flows for a Resource Manager which controls the handover. The architecture is flexible and can be used for other types of access networks, but the focus is on UMTS and WLAN interworking because of the common presence of these two technologies in today's wireless networks. The architecture is evaluated using the EXata Network Emulator, and results are presented as handover time between UMTS and WLAN networks.

**Keywords** – access network selection, vertical handover, heterogeneous wireless networks, IEEE 802.21.

## I. INTRODUCTION

The 3rd Generation of mobile communication systems is not focused on a “golden service”, responsible with the most revenues. It considers a diversity of services, with different requirements and, consequently, with different solutions for an optimal access technology. For this reason, a heterogeneous access network was a continuous dream both for operators and for the network equipment providers and only the technical difficulties and the estimations of higher costs for such access networks made that the interest for such solutions to be moderate. Fortunately, the technical progress in this matter is now important (the approval of the IEEE 802.21 standard [1] focused on Media Independent Handover – MIH is only an example) so the doubts on the technical difficulties related to the implementation of such heterogeneous access networks for 3G mobile communications systems, as well on the corresponding CAPEX and OPEX must be now reconsidered.

There are different candidates for the access network technologies in order to be considered as suitable in a heterogeneous environment for the 3G mobile communications systems:

- The regular 3G access network (the two well-known FDD (WCDMA) and TDD (TC-CDMA) access networks promoted by 3GPP for UMTS-like systems or the one proposed by 3GPP2 for CDMA2000 systems).

These access networks represent an optimal answer for covering the regular functionalities of the access networks in a high mobility environment, especially for interactive or conversational services, but not at very low operational costs;

- The WLAN/WiFi-type of access networks, including here the latest versions of this family of wireless access network standards, very simple for deployment and suitable in the case of stationary or nomadic wireless access, with low costs both for investments and for operation;
- The WiMAX type of access networks, very convenient for deployment in rural areas or in emerging regions, brings the benefit of reasonable quality in NLOS propagation conditions in a low mobility environment, but the quality of service (QoS) in high mobility conditions is questionable;
- The digital broadcasting technologies (DVB-H, but also DVB-T or DVB-T2) can bring a benefit for services or applications including a large broadcast or multicast downlink distribution of the required information;
- The WPAN technologies can represent a very attractive access solution for low or very low range, with a potential good management of the battery and of the radio spectrum.

It is clear that the diversity of the available access network technologies represents both a technical challenge and a potentiality in optimizing the access network performances.

Due to the popularity of the UMTS 3G networks and of the WLAN technologies, we consider that both represent a must as starting point in any discussion related to the performances and the optimization of the implied heterogeneous network and, for this reason, these technologies will be discussed in the present paper.

The rest of the paper is organized as follows: Section II presents some related work in this field, Section III presents the proposed algorithm and Section IV presents its implementation in the EXata Network Emulator. Simulation approach and results are presented in Section V while conclusions are drawn and further work is outlined in Section VI.

## II. RELATED WORK

Many papers have studied and are still studying ways to achieve interworking between heterogeneous wireless networks.

<sup>1</sup> Alexandru Vulpe is with the Faculty of Electronics, Telecommunications and IT at POLITEHNICA University of Bucharest, 1-3 Iuliu Maniu Blvd, Bucharest 061071, Romania, E-mail: alex.vulpe@radio.pub.ro.

<sup>2</sup> Octavian Fratu is with the Faculty of Electronics, Telecommunications and IT at POLITEHNICA University of Bucharest, 1-3 Iuliu Maniu Blvd, Bucharest 061071, Romania.

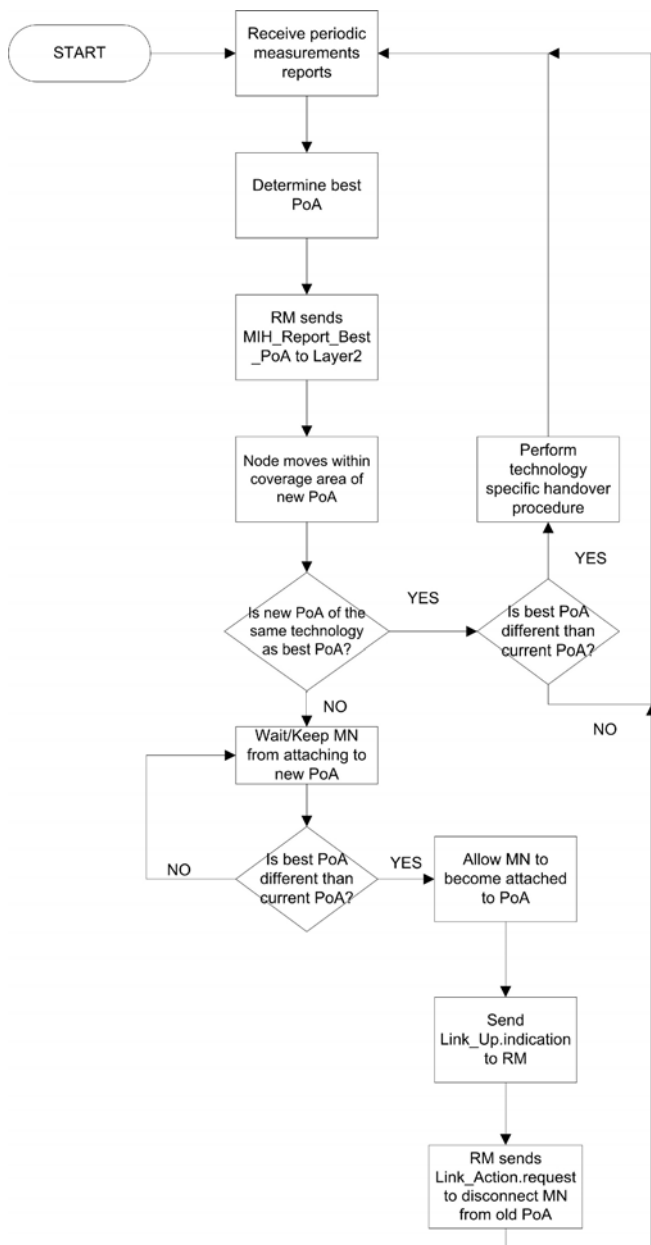


Fig. 1. Proposed framework flow

The current solutions considered in standardization bodies are discussed in [2]. These include I-WLAN (Interworked-WLAN), GAN (Generic Access Network) and IEEE802.21. Of these, I-WLAN is currently adopted by 3GPP, the GAN model, while not being restricted to WLAN networks is more commonly implemented as unlicensed mobile access (UMA)-enabled terminals, and IEEE802.21, although offering a flexible framework to facilitate different handover scenarios, has not seen adoption in 3GPP, but only in IEEE 802 standards.

In [3], the authors propose a WLAN/HSDPA Interworking Architecture for Linux-based mobile terminals. It is implemented on a Linux laptop with commercially available HSDPA modems and WLAN cards. However, the testing is done through a movement simulator, which could provide inaccurate results.

WLAN discovery schemes based on MIH Services are presented in [4]. They are either channel information based, location information based, or channel *and* location information based. The obtained results are good, but the one thing assumed and lacking at the moment is an implementation of a Media Independent Information Server (MIIS).

The authors in [5] propose an interoperability mechanism for WLAN and UMTS-HSDPA based on the output of a cost function as handover metric. However the study is carried out for HIPERLAN/2, which hasn't been widely implemented in commercial networks.

Another paper [6] proposes a load balancing scheme using Admission Control and Vertical Handoff in a cellular/WLAN integrated network. It measures its performance by means of data call throughput per voice call arrival rate, which is a questionable evaluation metric.

### III. FRAMEWORK DESCRIPTION

The flow of the proposed framework is described in Fig. 1. A Resource Manager (RM) that is running on the mobile node collects the Received Signal Strength (RSS) (or Received Signal Code Power (RSCP) for UMTS) and Carrier-to-Interference-plus-Noise-Ratio (CINR) (or  $E_c/N_o$  for UMTS) values from periodic measurement reports done by the multi-radio mobile node (MN). These are fed into a decision-making engine, where it is possible to implement any handover decision-making algorithm. It should be noted that the RM could collect any other handover metrics (such as available bandwidth, user preference or cost), and that these are not currently reported by measurements done by technology-specific link layers. In the end, the RM should output the best point of attachment (PoA) to which the mobile node should be connected. The PoA is termed "the best" in respect to it fulfilling the handover decision criteria as defined by an access network selection algorithm.

The RM then issues a primitive called "Report\_Best\_PoA" which carries the PoA identifier and the radio access technology that is enabled on that specific node. This is broadcast to the lower layers of each radio access technology protocol stack of the Mobile Node where it is stored. When the mobile node roams inside the coverage area of a new PoA, two situations may occur, depending whether it is of the same technology or not. If it is of the same technology, then, if it becomes the best PoA (determined by the RM), then normal technology-specific (horizontal) handover procedures (cell selection/reselection - in UMTS or association and authentication - in WLAN) are carried out. Otherwise, nothing is done, and the RM keeps receiving measurement reports. If the newly discovered PoA is of another technology, then, naturally, the radio interface of the corresponding technology belonging to the mobile node will want to carry out an attachment procedure to the PoA (the terms used are *PS attach* for UMTS and *association* for IEEE 802.11). However, this is not wanted if there is no need for a handover (the current PoA has better SNR, or greater bandwidth etc., i.e. it is still the best one), so the attachment procedures will not be executed until that PoA becomes the best PoA.

TABLE I  
802.21 PRIMITIVES IMPLEMENTED IN EXATA

MIH Service	Primitive	Description
Event	Link_Up	Layer 2 connectivity established
Event	Link_Down	Layer 2 connectivity loss
Event	Link_Parameters_Report	Link parameters periodic reporting
Event	Link_Going_Down	Layer 2 connectivity loss imminent
Command	Link_Get_Parameters	Request parameters from link layer
Command	Link_Action	Request an action on a link layer connection

When the MN becomes attached to a PoA, it sends a Link\_Up.indication as per IEEE 802.21 specifications [1]. When the RM receives it, it will send a Link\_Action.request message to the Link Layer of the radio access technology corresponding to the old PoA. It contains the command to disconnect from the old PoA. Therefore, the control of the handover procedure is partially taken over by the Resource Manager.

#### IV. IMPLEMENTATION

The framework described in Section III has been implemented in the EXata Network Simulator/Emulator. It uses a layered architecture which is similar to the one used in the TCP/IP protocol stack. In order to implement new protocols, or modify existing protocols, the source code (written in C/C++) of EXata has to be modified [7]. The main components of the IEEE 802.21 standard that have been implemented are:

- **The MIH Function (MIHF):** It receives MIH messages from the local link layers or a peer MIH node and forwards them to the MIH User. It is implemented as a protocol included in the Network/IP layer of the EXata protocol stack.
- **The MIH Protocol:** It provides a way for conveying MIH primitives between remote MIH nodes and includes an acknowledgement service in the case that the transport protocol used is unreliable
- **The Link SAP:** It is a collection of primitives used for interfacing between the media dependent link-layer and the MIH Function. Only a selected set of primitives has been implemented, and they are represented in Table I. The SAP is implemented as a separate library and included in the EXata source code.
- **Link layer extensions:** The media dependent link layers have been modified in order to support the above-described algorithm. These are done by modifying the source code of the MAC Layer of the EXata protocol stack associated with a specific link-layer technology (in this case, UMTS and WLAN).
- **MIH User:** The MIH User, which is termed here "Resource Manager", has been implemented as a simple RSS-based handover decision, in order to prove the concept. The IEEE 802.21 standard does not restrict

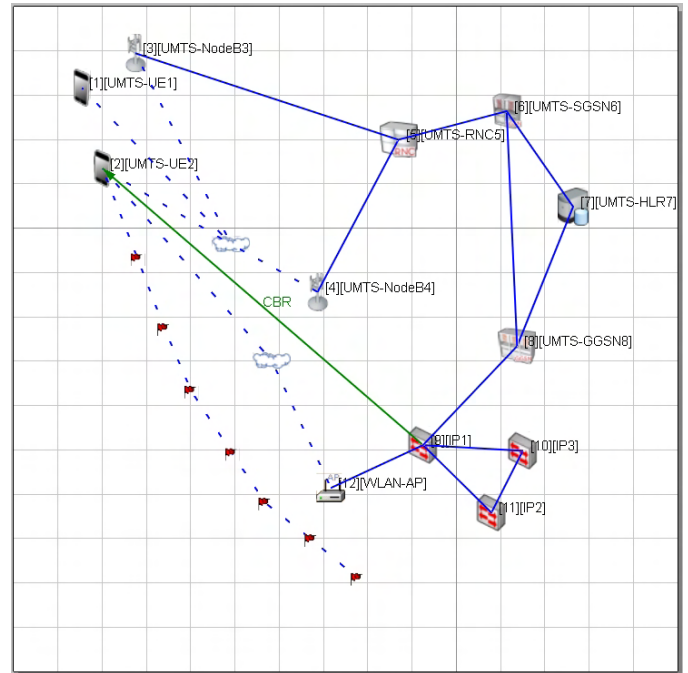


Fig. 2. Scenario Topology

the implementation of the MIH User to be above the MIHF, so, for ease of coding, it is implemented at the same layer as the MIH Function.

Simulations regarding the accuracy and performance of the implementation of the protocol have been published in previous papers [8-10]. However, there we had not implemented a UMTS link extension, but only dealt with WLAN and WiMAX.

#### V. SIMULATIONS AND RESULTS

The simulated network topology consists of a UMTS Radio Access Network (RAN) and a WLAN network connected to each other via an IP network. The topology is depicted in Fig. 2. Node 2 is a mobile node with both UMTS and WLAN radio interfaces and is connected to the UMTS RAN (Nodes 3-8) and the WLAN Access Point (Node 12). Nodes 9, 10 and 11 form an IP network backbone which connects the UMTS RAN to the WLAN network. The MN moves through the UMTS network, following the path outlined by the red waypoints in Fig. 2, and comes into the coverage area of the WLAN Access Point. At this point, the MN would have the tendency to become associated with the AP, but, instead, the procedure described in Section III takes place: only when the RSS of the WLAN access point becomes better than the one from the UMTS network, by a margin of 2 dB, the Resource Manager, by processing measurement reports, determines that the best PoA is the WLAN Access Point, and, therefore, the normal procedure of association and authentication takes place. A Link\_Up.indication primitive, originating from the WLAN link layer of the multi-radio mobile node is sent to the RM via the MIH Function. The RM then sends a Link\_Action.request message to the Mobile Node's UMTS

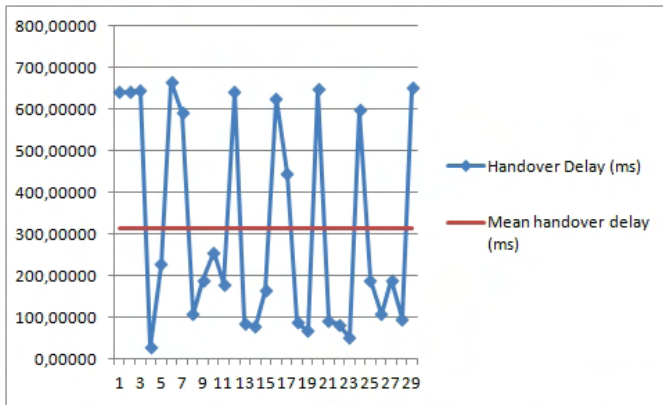


Fig. 3. UMTS to WLAN handover delay

interface, requesting it to disconnect from the network. In this way, a make-before-break handover takes place.

Fig. 3 shows the handover delay measured for 30 simulations, done with different seed values. We can conclude that the handover delay from UMTS to WLAN varies around a mean value of 313 ms, with the lowest value reaching about 27 ms, and the highest value situated at 667 ms.

## VI. CONCLUSION AND FURTHER WORK

This paper proposed an interworking architecture for mobility between heterogeneous networks based on the IEEE 802.21 (MIH) standard and applied to handover from UMTS to WLAN networks. It defines a framework for providing handover-related information flows to a Resource Manager in charge of taking the handover decision.

Results have shown that the handover delay has relatively small but different values as they depend also on network conditions. Further studies have to be carried out. In particular, the handover delay between WLAN and UMTS networks needs to be measured and evaluated in different network conditions (also the corresponding scenarios implementation in EXata can be improved) in order to evaluate it in a general manner. Also, different handover metrics can be introduced. The framework also needs to be enhanced to support more technologies, especially LTE, as it is the leading emerging cellular technology.

## ACKNOWLEDGEMENT

This research activity was supported by the Ministry of Communications and Information Society of Romania under

the grant no. 106/2011 “Evolution, implementation and transition methods of DVB radiobroadcasting using efficiently the radio frequencies spectrum” and by the Sectorial Operational Programme Human Resources Development 2007-2013 of the Romanian Ministry of Labour, Family and Social Protection through the Financial Agreement POSDRU/107/1.5/S/76903.

## REFERENCES

- [1] IEEE 802.21-2008, Standard for Local and Metropolitan Area Networks - Part 21: Media Independent Handover Services, IEEE Std., Jan. 2009.
- [2] R. Ferrus, O. Sallent, and R. Agusti, “Interworking in heterogeneous wireless networks: comprehensive framework and future trends,” *Wireless Communications, IEEE*, vol. 17, no. 2, pp. 22–31, 2010. [Online]. Available: [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=5450657](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5450657)
- [3] D.-H. Kim, J.-Y. Gwak, W.-T. Kim, S.-J. Kim, and C.-H. Lee, “Design and Implementation of WLAN/HSDPA Interworking Architecture,” *2008 Third International Conference on Convergence and Hybrid Information Technology*, no. Cm, pp. 741–744, Nov. 2008. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4682116>
- [4] W. Lim, D. Kim, and Y. Suh, “Efficient WLAN Discovery Schemes Based on IEEE 802.21 MIH Services in Heterogeneous Wireless Networks,” *Conference, 2008. IEEE*, 2008. [Online]. Available: [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=4698822](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4698822)
- [5] K. Peppas, F. Lazarakis, A. Alexandridis, K. Dangakis, and D. Axiotis, “NGL03-4: An Interoperability Mechanism for Seamless Interworking between WLAN and UMTS-HSDPA Networks,” in *Global Telecommunications Conference, 2006. GLOBECOM'06. IEEE*, 2006, pp. 1–5. [Online]. Available: [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=4150888](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4150888)
- [6] W. Song, W. Zhuang, and Y. Cheng, “Load Balancing for Cellular/WLAN Integrated Networks,” *IEEE Network*, no. February, pp. 27–33, 2007
- [7] *EXata 2.0.1 Programmer's Guide*, Scalable Network Technologies, Inc., Jun. 2009
- [8] A. Vulpe, S. Obreja, O.-E. Barbu, C.-D. Penciu, M. Buga, and B. Ciobotaru, “Qualnet implementation for mobility management in a MIH enabled system,” in *8th International Conference on Communications (COMM) 2010*, Jun. 2010, pp. 523–526.
- [9] A. Vulpe, S. Obreja, and O. Fratu, “A study of mobility management using IEEE 802.21,” in *9th International Symposium on Electronics and Telecommunications (ISETC) 2010*, Nov. 2010, pp. 205–208.
- [10] —, “Interoperability procedures between access technologies using IEEE 802.21,” in *2nd International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace Electronic Systems Technology (Wireless VITAE) 2011*, Feb./Mar. 2011, pp. 1–5.

# Simulation of Rare Events in Teletraffic Systems with Single Queue

Elena Ivanova<sup>1</sup>, Rostislav Raev<sup>2</sup> and Dimitar Radev<sup>3</sup>

*Abstract* – The paper is devoted on computer simulation of rare event probability of estimation Quality of Service (QoS) parameters. Under consideration is a basic reference model with single server queuing system M/M/1/N where service disciplines are FIFO and LIFO. Simulation results for the queuing systems [X]/M/1/N M/[X]/1/N and [X]/[X]/1/N are shown.

*Keywords* – Rare event, Teletraffic systems, Quality of Service, CLR, .net.

## I. INTRODUCTION

In modern communications very important is to estimate such parameters of Quality of Service (QoS) parameter as Cell Loss Ratio and packet loss. The values of these parameters are smaller than  $1 \cdot 10^{-9}$ , and they belong to rare events.

Such problems can't be solved with standard stochastic simulation. The aim of this paper is to solve with rare event simulation queuing network model with single queues.

Under consideration is importance sampling for estimation rare event probability as one of the two basic techniques of rare event simulation - importance sampling and splitting.

The program realization is made on the .NET Framework, which includes a large library and provides language interoperability across several programming languages. The .NET Framework's Base Class Library provides user interface, data access, database connectivity, cryptography, web application development, numeric algorithms, and network communications. The .NET Framework is intended to be used by most new applications created for the Windows platform.

## II. REFERENCE QUEUING MODELS

The investigated queuing models are different. The type of investigated queuing systems are with different arrival and service distributions such as discrete and continues: Pareto, Geometrical, Poisson, Erlang. The basic queuing models for rare event simulation, investigated here are classified in three types [X]/M/1/N, M/[X]/1/N and [X]/[X]/1/N. The program realization is made for discipline of service FIFO or LIFO with a finite buffer size N. These types of single queue are necessary, because new telecommunication networks could be investigated with small amount of resources.

<sup>1</sup> Elena Ivanova is with Department of Telecommunications, University of Rousse, Studentska Str 8, 7017, Bulgaria

<sup>2</sup> Rostislav Raev is with Department of Telecommunications, University of Rousse, Studentska Str 8, 7017, Bulgaria

<sup>3</sup> Dimitar Radev is with Department of Telecommunications, University of Rousse, Studentska Str 8, 7017, Bulgaria

A basic reference model for rare event simulation is the single server queuing system M/M/1/N- FIFO with a finite buffer size N. The arrival rate is  $\lambda$  and the service rate is  $\mu$ . We are interested in the probability that the buffer content reaches a high level k during one busy period (i.e., the time interval between two successive periods in which the buffer is empty). The maximum occupancy is  $B=N+1$ , and the traffic load is  $\rho=\lambda/\mu$ . The discrete random variable is describe with stationary complementary distribution function (c.d.f.)  $G(x)=1-F(x)$ , for the loss probability  $P_B(1)$ .

### A. [X]/M/1/N

The investigated reference model for rare event simulation is the single server queuing system G/M/1/N - FIFO with a finite buffer size N.

The inter-arrival times are independent and uniformly distributed with arbitrary integral and  $f_A(\cdot)$  differential distribution laws.

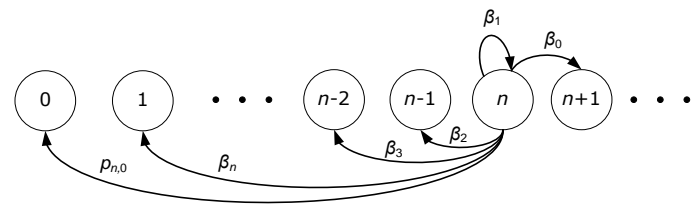


Fig. 1. Queuing system G/M/1

The service rate is exponential  $1/\mu$ . We are interested in the probability that the buffer content reaches a high level k during one busy period (i.e., the time interval between two successive periods in which the buffer is empty). The maximum occupancy is  $B=N+1$ , and the traffic load is  $\rho = \frac{\lambda(i,x)}{\mu} < 1$ . The discrete random variable is

describe with stationary complementary distribution function (c.d.f.)  $G(x)=1-F(x)$  the loss probability  $P_B(1)$  and the local correlation coefficient  $Cor_i(2)$  for the interval  $i-1 \leq x < i, i=1, \dots, k$ .

### B. M/[X]/1/N

The other investigated reference model for rare event simulation is the single server queuing system M/[X]/1/N with a finite buffer size N.

The service times are exponential distributed, and the service rates are uniformly distributed with arbitrary integral and differential distribution laws.

Queuing system M/Erl/1/k is the right opposite Erl/M/1/N, the arrival rate is exponential distributed, and the service rate

is Erlang distributed. The new client arrives in the left side of the system, then he is serviced in  $k$ -stages with  $k\lambda$ , the function of the arrival rate is defined with (1).

$$f(x) = \lambda e^{-\lambda x} \quad x \geq 0 \quad (1)$$

The service rate is (2).

$$F(x) = \frac{k\mu(k\mu x)^{k-1} e^{-r\mu x}}{(r-1)!} \quad x \geq 0 \quad (2)$$

We consider the Erlang distribution like a part of Geometrical distribution.

C.  $[X]/[X]/1/N$

Teletraffic systems with single queue with different distribution as arrival and service rates, represents client behavior in different networks like fast Internet network, optical backbones and multiplex, or even broadband convergence networks. Here, the model of interest for rare event simulation is the single server queuing system  $Erl/D/1/N$  with a finite buffer size  $N$ , the correlation coefficient is smaller than 1,  $\beta$  is non negative  $\beta=t$ , the inter-arrival distribution is  $ERL(\alpha, \beta)$ , with deterministic process of service  $D$ . The arrival rate is determined with (3).

$$F(t) = 1 - e^{-k\mu t} \sum_{i=0}^{k-1} \frac{(k\mu t)^i}{i!} \quad (3)$$

The service rate is defined with  $k$  stages of service,  $Cor_V$  with (4).

$$\mu = \frac{1}{Cor_V^2 k X} e^{-k\mu t}, k = \frac{1}{Cor_V^2} \quad (4)$$

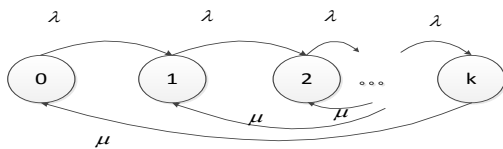


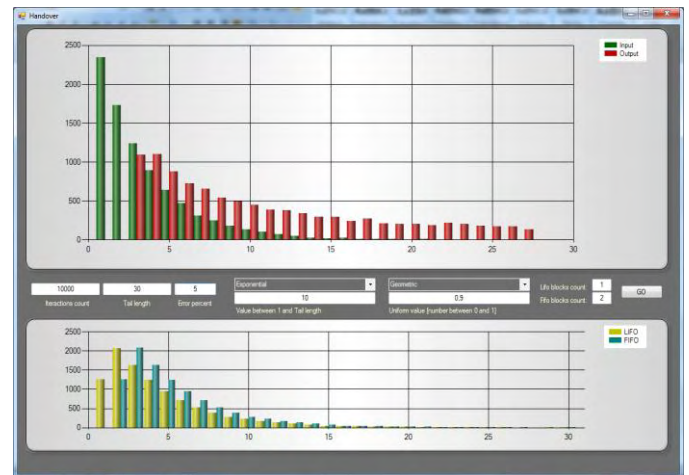
Fig. 2 Queuing system Erl/D/1/k

Again, the assumptions about erlang and geometric distributions are made again.

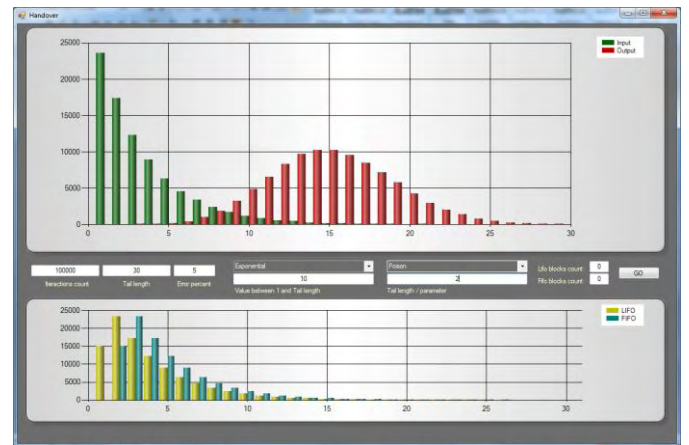
III. PROGRAM REALIZATION

The program for Simulation of Rare events in teletraffic systems with single queue is made with the .NET Framework, because it is the way that programming will be done on a Microsoft machine from now. And not just on a Microsoft machine. The ADO.NET is used for creating web site, and for manipulating databases. This framework create applications for mobile phones and PDA's with .NET. There is even a project in the making that allow to write a programme on a Windows machine that will then work on a computer NOT

running Windows. All this is made possible with the .NET Framework.



(a)



(b)

Fig.3 Program realization: simulation of rare events in teletraffic systems

The .NET Framework is a whole lot of Classes (called Namespaces) and the technology to get those Classes to work. The main component is called the Common Language Runtime.. With .NET, more than 15 different programming languages can use the Common Language Runtime. One of these languages is, of course Visual Basic .NET. Another is C#. They can all use the Common Language Runtime because of something called the Intermediate Language.

One of the most important goal of the program is simulation of different disciplines of service.

The conceptual model for rare event simulation of single queue system starts with Random Number Generator. The RNGs are produced different lend of distributions, such as Uniform, Geometric, Pareto, Poisson and Exponential. The distributions are used as arrival rate and service rate at single queue.

The simulation experiments are realized with different queues, with service discipline First In First Out (FIFO) and



Last In First Out (LIFO). Parameters for blocking probability and packet loss are investigated. With the help of Iterations Generator are made simulations for each of queues using different distributions. The received results are shown as diagrams.

The simulation of the three main types of queuing systems is made of the AMD Athlon(tm) II X2 250 Processor 3.00 GHz and 3 GB RAM, running 32-bits operation system is Windows 7 Professional.

#### IV. SIMULATION AND RESULTS

The simulation of the queuing systems were executed for single server queuing system M/M/1/ N, Geo/M/1/N, Pareto/M/1/ N and M/Geo/1/ N FIFO and LIFO as a part of simulation system., error 5% or 10%. The purpose of relative error is to investigate occurred rare event in different systems.

Consider system M/M/1/N with N=20, 30 and 40 place for service in the queue, discipline of service is FIFO and LIFO, where the arrival and service rate is exponential and Poisson distributed.

The simulation of the buffer content is made for 100 000 observation. The increasing of arrival rate leads to insignificant increasing of rare events. We choose arrival and service rate, to receive appropriate result. As is seen from Table I, rare events in system with LIFO are less than the same system with the same parameters FIFO.

TABLE I  
SIMULATION OF RARE EVENTS IN M/M/1/N SYSTEM

N	RE, %	Discipline of service	Arrival rate	Service rate	Trials	Rare events
20	5	LIFO	15	5	100000	2
		FIFO	15	5	100000	5
	10	LIFO	15	5	100000	5
		FIFO	15	5	100000	6
30	5	LIFO	15	5	100000	0
		FIFO	15	5	100000	2
	10	LIFO	15	5	100000	3
		FIFO	15	5	100000	4
40	5	LIFO	15	5	100000	1
		FIFO	15	5	100000	2
	10	LIFO	15	5	100000	3
		FIFO	15	5	100000	3

The system Geo/M/1/N with p=0,1 is shown on Table II . The simulation was used to receive representative and stable results with sequences of 100 000 observations.

TABLE II  
SIMULATION OF RARE EVENTS IN GEO/M/1/N SYSTEM

N	RE, %	Discipline of service	Arrival rate	Service rate	Trials	Rare events
20	5	LIFO	0,1	5	100000	0
		FIFO	0,1	5	100000	3
	10	LIFO	0,1	5	100000	2
		FIFO	0,1	5	100000	4
30	5	LIFO	0,1	5	100000	0
		FIFO	0,1	5	100000	0
	10	LIFO	0,1	5	100000	1
		FIFO	0,1	5	100000	1
40	5	LIFO	0,1	5	100000	0
		FIFO	0,1	5	100000	0
	10	LIFO	0,1	5	100000	1
		FIFO	0,1	5	100000	2

TABLE III  
SIMULATION OF RARE EVENTS IN M/GEO/1/N SYSTEM

N	RE, %	Discipline of service	Arrival rate	Service rate	Trials	Rare events
20	5	LIFO	10	0,9	100000	6
		FIFO	10	0,9	100000	8
	10	LIFO	10	0,9	100000	10
		FIFO	10	0,9	100000	16
30	5	LIFO	10	0,9	100000	5
		FIFO	10	0,9	100000	7
	10	LIFO	10	0,9	100000	5
		FIFO	10	0,9	100000	7
40	5	LIFO	10	0,9	100000	1
		FIFO	10	0,9	100000	3
	10	LIFO	10	0,9	100000	6
		FIFO	10	0,9	100000	9

TABLE IV  
SIMULATION OF RARE EVENTS IN GEO/GEO/1/N SYSTEM

N	RE, %	Discipline of service	Arrival rate	Service rate	Trials	Rare events
20	5	LIFO	0,3	0,8	100000	0
		FIFO	0,3	0,8	100000	1
	10	LIFO	0,3	0,8	100000	1
		FIFO	0,3	0,8	100000	1
30	5	LIFO	0,3	0,8	100000	1
		FIFO	0,3	0,8	100000	1
	10	LIFO	0,3	0,8	100000	0
		FIFO	0,3	0,8	100000	0
40	5	LIFO	0,3	0,8	100000	0
		FIFO	0,3	0,8	100000	0
	10	LIFO	0,3	0,8	100000	1
		FIFO	0,3	0,8	100000	1

The same simulation is made for M/Geo/1/N queuing systems with exponential characteristics of arrival rate and geometric service rate, the received results for buffer size 20, 30 and 40, are presented on Table III. The prescribed error is

in the range of 5% and 10%. The number of simulation samples is  $n=100\,000$ .

The same simulation is made for Geo/Geo/1/N queuing systems with geometric characteristics of arrival rate and geometric service rate, the received results for buffer size 20, 30 and 40, are presented on Table IV. The prescribed error is in the range of 5% and 10%. The number of simulation samples is  $n=100\,000$ . The main difference here is the absence of rare events for 100 000 trials, for some of the chosen parameters.

## V. CONCLUSIONS AND FUTURE WORK

Simulation of rare events in different teletraffic systems is the target of program realization. The number of rare events defines the simulation probability of blocking different network systems.

The simulation results for the basic reference model, single server queuing system M/M/1/N with FIFO and LIFO service disciplines, M/[X]/1/N, [X]/M/1/N, and [X]/[X]/1/N systems with a finite buffer size  $N$ , are very good. The results for the M/[X]/1/N, [X]/M/1/N and [X]/[X]/1/N systems shows that the prescribed error is in the range of 5% and 10%, where the number of simulation samples is  $n=100\,000$ .

## REFERENCES

- [1] Awan I.U., D.D.Kouvatsos, Arbitrary queueing network models with service priorities and blocking, Proc. Of 12th UK Workshop of Perf. Eng. Of Computer and Telecommunication Systems, ed. D.D. Kouvatsos, ilkley, UK, 1997.
- [2] Boer. P. Analysis and Efficient Simulation of Queuing models of Telecommunication Systems. PhD thesis, Univ. of Twente, 2000.
- [3] Bucklew, J. An Introduction to Rare Event Simulation. *Springer Series in Statistics*, XI. Springer-Verlag, Berlin, 2004.
- [4] Cerou, F., Guyader, A. Adaptive multilevel splitting for rare event analysis. Rapport de Recherché 5710, INRIA, October 2005, pp.32-40.
- [5] L'Ecuyer P., Demers V., Tuffin B., Splitting for rare-event simulation, Proceedings of the 37th conference on Winter simulation, Monterey, California, 2006.
- [6] L'Ecuyer P., Demers V., Tuffin B., Rare events, splitting, and quasi-Monte Carlo, ACM Transactions on Modeling and Computer Simulation (TOMACS), v.17 n.2, p.9-es, 2007.
- [7] Garvels M., D.P. Kroese. A comparison of RESTART implementations. In Proceedings of the 1998 Winter Simulation Conference, Washington, DC, 1998 p. 601–609.
- [8] Garvels M. The splitting method in rare event simulation. PhD thesis, University of Twente, The Netherlands, 2000.
- [9] Garvels, M., D. Kroese, A Comparison of RESTART Implementations, University of Twente, Centre for Telematics and Information Technology, 2005.
- [10] Glasserman P., Heidelberger P., Shahabuddin P., Zajic T., Splitting for rare event simulation: analysis of simple cases, Proceedings of the 28th conference on Winter simulation, Coronado, California, United States, 1996, p.302-308.
- [11] Glasserman P., Heidelberger P., Shahabuddin P., Zajic T., A large deviations perspective on the efficiency of multilevel splitting, IEEE Transactions on Automatic Control AC-43 (12), 1998, p.1666-1679.
- [12] Görg, C. Schreiber, F, The RESTART/LRE method for rare event simulation, Proceedings of the 28th conference on Winter simulation, p.390-397, December 08-11, 1996, Coronado, California, United States.
- [13] Görg, C. Schreiber, F. The RESTART/LRE method for rare event simulation. Winter Simulation Conference, California, USA, 1999, pp. 390-397.

# VoIP over a Cognitive Network with Limited Availability

Yakim Mihov and Boris Tsankov

**Abstract**—This paper investigates the call-level performance of a cognitive radio network for dynamic spectrum access providing VoIP service to secondary users. The paper proposes a new cognitive network paradigm, where only a part of the licensed channels are available to the unlicensed users by cognitive functionality. Numerical results are presented and some conclusions are drawn.

**Keywords** – Call blocking probability, call dropping probability, cognitive radio network, limited availability, VoIP.

## I. INTRODUCTION

The term Cognitive radio (CR) was first introduced by J. Mitola [1]. One of the most popular applications of CR is in dynamic spectrum access (DSA) networks, as a means to mitigate the artificially created scarcity of spectrum resources caused by the traditional static approach for spectrum regulation. Hierarchical spectrum overlay is a promising method for DSA. It allows secondary (unlicensed or cognitive) users (SUs) to temporarily utilize spectrum resources assigned to primary (licensed or incumbent) users (PUs) if these resources are not currently being used for PU transmission. SUs have to release the occupied resources as soon as PUs start reusing them, i.e. PUs have preemptive priority over SUs. The cognitive network utilizes opportunistically the available unoccupied spectrum of the primary network on a non-interference basis. Spectrum handover is an essential function of CR since it enables and facilitates the Quality of Service (QoS) provisioning of the SUs.

A popular and often quoted overview of CR is presented in [2]. Due to the great interest in using CR networks for DSA, there are numerous publications in the literature. Spectrum sensing is studied in [3], [4], [5]. Multiuser spectrum selection schemes for spectrum sharing and resolving channel contention are analyzed in [6]. Spectrum handover is investigated in [7], [8], [9]. Various QoS-related issues in CR networks are studied in [10], [11], [12], [13]. The resource allocation problem in a multiuser orthogonal frequency division multiplexing (OFDM) based CR system concerning the QoS provisioning for both real-time and non-real-time applications is investigated in [14]. An overview of the general methodology for cross-layer design and some cross-layer optimization schemes and algorithms are presented in [15]. The voice traffic service is of a particular interest. Some examples are [16], [17], [18], [19], [20], [21], [22].

<sup>1</sup> Yakim Mihov is with the Faculty of Telecommun. at TU-Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, yakim\_mihov@abv.bg.

<sup>2</sup> Boris Tsankov is with the Faculty of Telecommun. at TU-Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria. bpt@tu-sofia. bg

Arrangement of “cognitive” channels over given spectrum has a certain cost due to realization of functions like spectrum hole detection; detection of PU call arrival over channel occupied by a SU; spectrum handover realization etc. In case the secondary traffic volume does not need all of the licensed channels to be served, it looks rationally the secondary traffic to have access only to some part of the channels of the primary network. Therefore, the “cognitivity” will be arranged to these channels only. As a consequence, operations such as scanning, detection of idle/busy channel conditions, spectrum handover will take place faster as they are performed over a limited number of channels. This paper investigates the voice traffic performance under the circumstances described above.

## II. THE TELETRAFFIC SYSTEM

The corresponding teletraffic serving system is shown on Fig. 1. The offered PU traffic is denoted with  $A_p$  and the offered SU traffic is denoted with  $A_s$ . The PU calls have access to all of the  $N$  channels of the primary network. The SU calls have access only to  $N_c$  cognitive channels (limited availability). The channels  $N_o = N - N_c$  are not subject to cognitive activity, i.e. the secondary CR network is allowed to utilize only the predetermined  $N_c$  channels of the primary network. The bandwidth of a SU call is assumed to be equal to the bandwidth of a PU call, i.e. one channel is occupied by one PU or SU call. Perfect spectrum sensing and spectrum handover procedures are assumed. The service of PU calls is independent of the service of SU calls.

A slight system modification is proposed on Fig. 2 where the primary traffic  $A_p$  is first directed to the group of  $N_o$  channels devoted to PU calls only. Calls rejected from that group of channels represent an overflow traffic directed to the cognitive channels  $N_c$ . The cognitive channels form a serving system with PUs and cognitive SUs in accordance with the hierarchical spectrum overlay approach for DSA. PUs have preemptive priority over SUs. If a PU starts transmitting on a

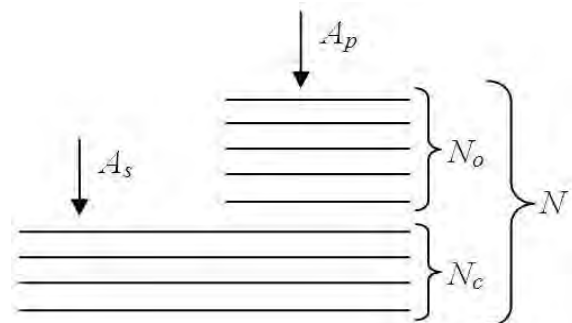


Fig. 1. Illustration of the teletraffic serving system.

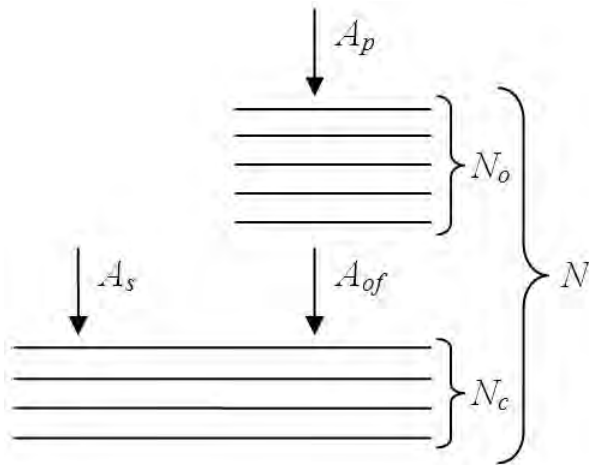


Fig. 2. Illustration of the proposed modified teletraffic serving system.

cognitive channel which is occupied by a SU call, the cognitive channel has to be vacated immediately. In this case, the SU performs spectrum handover to another idle cognitive channel in order to ensure successful call service completion. If there are no idle cognitive channels, the SU call is dropped.

### III. PERFORMANCE ANALYSIS

The offered PU and SU traffic is modeled by two Poisson random processes with arrival rates  $\lambda_p$  and  $\lambda_s$ , respectively. The PU and SU call durations follow a negative exponential distribution with mean  $1/\mu$ . Because of the limited availability of channels to the secondary CR network, it is impossible to apply the traditional method of building a 2-D Markov chain [23] or to find exact and simple closed-form solution of the corresponding steady-state equations and derive important QoS characteristics, such as the SU call blocking probability  $B_s$  and the SU call dropping probability  $B_d$ .

In the more practical and efficient arrangement proposed on Fig. 2, there is an overflowing traffic  $A_{of}$ , which is not a Poisson traffic at all. Because of the preemptive priority of the PU calls over the SU calls, it is impossible to apply the well known *equivalent random theory* [24] used for overflow traffic.

There is not any difficulty to obtain the PU call blocking probability  $B_p$  as the service of PU calls is affected neither by the secondary traffic, nor by the cognitive functionality. However, the application of limited availability influences the service of the secondary traffic. This is investigated in the paper by simulations.

### IV. SIMULATION RESULTS

In this section, the performance of the secondary CR network with limited channel availability is analyzed by

simulation experiments and some insightful conclusions are drawn.

A simulation model has been developed which takes into account all the essential factors necessary for performance evaluation of the described teletraffic system (Figs. 1 and 2), such as the Poisson PU and SU call arrival flows, the random service time of a PU or SU call with negative exponential distribution, the preemptive priority of PU calls over SU calls, and the application of limited channel availability for the CR network. Moreover, the proposed system modification (see Fig. 2) has also been implemented as an option in the simulation model.

We first analyze the effect of the number of available channels to the secondary CR network (i.e. the number of cognitive channels)  $N_c$  on the cognitive traffic capacity when some predefined level of SU QoS provisioning (in terms of SU call blocking probability  $B_s$  and SU call dropping probability  $B_d$ ) has to be guaranteed. As  $N_c$  decreases, the traffic capacity of the CR network decreases as well (see Fig. 3). Therefore, the limitation of the availability of the PU channels for DSA comes at the price of reduced cognitive traffic capacity, which is undesirable if the offered SU traffic that has to be served by the CR network is relatively large. However, when the offered SU traffic is relatively small, a reasonable decrease in the capacity of the CR network due to limited channel availability would not degrade the service of SU calls. The limitation of the cognitive channels, i.e. the use of a predefined subset of PU channels for DSA, may be desirable since the procedures and operations for supporting the “cognitivity” of the secondary network do not have to be performed on all of the channels of the primary network, i.e. the cognitive processing load (including procedures such as spectrum sensing, spectrum analysis, spectrum handover, etc.) can be reduced significantly, which is especially favorable in a resource-constrained cognitive environment.

Next, we analyze the effect of the offered PU traffic  $A_p$  on the traffic capacity of the secondary CR network when given

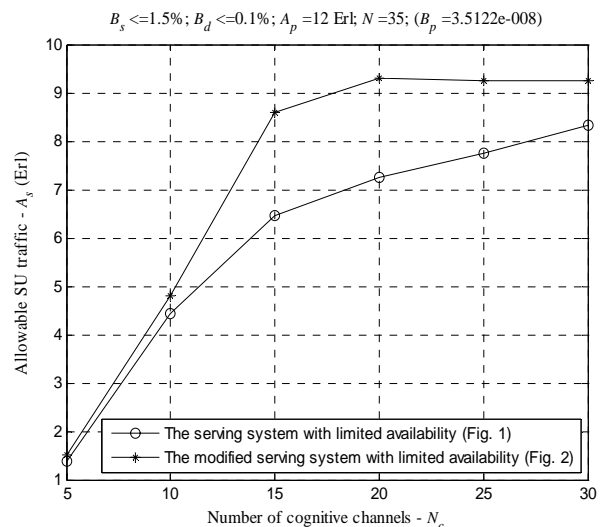


Fig. 3. Cognitive traffic capacity versus the number of channels available to the secondary CR network.

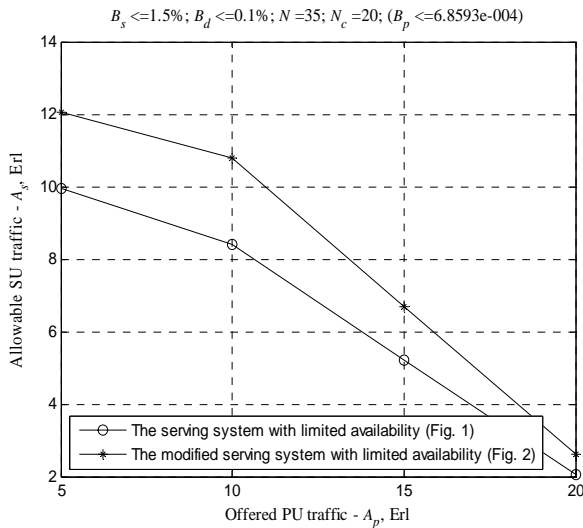


Fig. 4. Cognitive traffic capacity versus the offered PU traffic.

SU call blocking probability  $B_s$  and SU call dropping probability  $B_d$  have to be maintained. As  $A_p$  increases, the maximum allowable offered SU traffic  $A_s$  decreases (see Fig. 4). Consequently, the traffic capacity of the CR network depends on the PU traffic load and the use of CR for DSA is rational in primary networks with sufficiently underutilized transmission resources.

Finally, we analyze the effect of applying the proposed modification of the serving system (see Fig. 2) on the performance of the secondary CR network. Figs. 3 and 4 show that the modified serving system provides greater cognitive traffic capacity. Because of the suggested modification of the serving system, the channels  $N_o = N - N_c$  (that are not used for DSA) are more likely to be occupied by PU calls than the  $N_c$  cognitive channels, which leads to a reduction in the SU call dropping probability  $B_d$ . Similarly, the probability for spectrum handover of ongoing SU calls decreases as well, which facilitates the QoS provisioning in the CR network, since under certain circumstances spectrum handover could cause intolerable transmission delay. Therefore, the application of the proposed slight modification to the serving system considered herein improves considerably the performance of the secondary CR network.

## V. CONCLUSION AND FUTURE WORK

In this paper, the call-level performance of a VoIP secondary CR network operating over a VoIP primary network in accordance with the hierarchical spectrum overlay approach for DSA is investigated in the specific case when DSA is performed only on a predefined subset of primary channels. The simulation results presented in this paper prove the feasibility of the proposed new paradigm for DSA with limited availability. The main advantage of DSA with limited availability is that the cognitive processing load can be significantly reduced. Moreover, the risk of causing intolerable interference to the primary network due to spectrum sensing

errors is considerably reduced, since some channels of the primary network are never occupied by SU calls.

The suggested modification of the investigated serving system with limited availability leads to considerable performance enhancement of the secondary CR network in terms of increased cognitive traffic capacity and reduced probability for spectrum handover of ongoing SU calls.

For future research work, the authors plan to develop algorithms for determining the optimal number of primary channels available for DSA based on different design criteria.

## REFERENCES

- [1] J. Mitola III and G. Q. Maguire, "Cognitive Radio: Making Software Radios More Personal", IEEE Personal Communications, vol. 6, no. 4, pp. 13-18, August 1999.
- [2] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "A Survey on Spectrum Management in Cognitive Radio Networks", IEEE Communications Magazine, vol. 46, no. 4, pp. 40-48, April 2008.
- [3] W.-Y. Lee and I. F. Akyildiz, "Optimal Spectrum Sensing Framework for Cognitive Radio Networks", IEEE Transactions on Wireless Communications, vol. 7, no.10, pp. 3845-3857, 28 October 2008.
- [4] A. A. El-Saleh, M. Ismail and M. A. M. Ali, "Optimizing Spectrum Sensing Parameters for Local and Cooperative Cognitive Radios", 11<sup>th</sup> International Conference on Advanced Communication Technology (ICACT), pp. 1810-1815, Gangwon-Do, South Korea, 15-18 Feb. 2009.
- [5] A. Ghasemi and E. S. Sousa, "Collaborative Spectrum Sensing for Opportunistic Access in Fading Environments", 1<sup>st</sup> IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN), pp. 131-136, Baltimore, MD, USA, 8-11 Nov. 2005.
- [6] C.-W. Wang, L.-C. Wang and F. Adachi, "Modeling and Analysis of Multi-User Spectrum Selection Schemes in Cognitive Radio Networks", The 20th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), pp. 828-832, Tokyo, Japan, September 2009.
- [7] L.-C. Wang and C. Anderson, "On the Performance of Spectrum Handoff for Link Maintenance in Cognitive Radio", The 3rd International Symposium on Wireless Pervasive Computing (ISWPC), pp. 670-674, Santorini, Greece, May 2008.
- [8] O. Jo, H. H. Choi and D.-H. Cho, "Seamless Spectrum Handover Improving Cell Outage in Cognitive Radio Systems", 4<sup>th</sup> International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM), pp. 1-6, Hannover, Germany, June 2009.
- [9] L. Giupponi and A. I. Perez-Neira, "Fuzzy-based Spectrum Handoff in Cognitive Radio Networks", The 3rd International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM), pp.1-6, Singapore, May 2008.
- [10] L.-C. Wang and C.-W. Wang, "Spectrum Management Techniques with QoS Provisioning in Cognitive Radio Networks", The 5<sup>th</sup> IEEE International Symposium on Wireless Pervasive Computing (ISWPC), pp. 116-121, Modena, Italy, May 2010.

- [11] Y. Mihov and B. Tsankov, "QoS Provisioning via Channel Reservation in Cognitive Radio Networks", in Proc. IEEE COMCAS, pp. 1-5, Tel Aviv, Israel, November 2011.
- [12] T. Jin, C. Chigan and Z. Tian, "Game-theoretic Distributed Spectrum Sharing for Wireless Cognitive Networks with Heterogeneous QoS", IEEE Global Telecommunications Conference (GLOBECOM '06), pp. 1-6, San Francisco, CA, USA, December 2006.
- [13] B. Ishibashi, N. Bouabdallah and R. Boutaba, "QoS Performance Analysis of Cognitive Radio-based Virtual Wireless Networks", IEEE INFOCOM 2008, The 27<sup>th</sup> Conference on Computer Communications, pp. 2423-2427, April 2008.
- [14] Y. Zhang and C. Leung, "Cross-Layer Resource Allocation For Mixed Services in Multiuser OFDM-based Cognitive Radio Systems", IEEE Transactions on Vehicular Technology, vol. 58, no. 8, pp. 4605-4619, October 2009.
- [15] Q. Xin and J. Xiang, "Joint QoS-aware Admission Control, Channel Assignment, and Power Allocation for Cognitive Radio Cellular Networks", IEEE 6<sup>th</sup> International Conference on Mobile Adhoc and Sensor Systems (MASS), pp. 294-303, Macau, 12-15 Oct. 2009.
- [16] H. Lee and D.-H. Cho, "VoIP Capacity Analysis in Cognitive Radio System", IEEE Communications Letters, vol. 13, no. 4, pp. 393-395, June 2009.
- [17] H. Lee and D.-H. Cho, "Capacity Improvement and Analysis of VoIP Service in a Cognitive Radio System", IEEE Transactions on Vehicular Technology, vol. 59, no. 4, pp. 1646-1651., May 2010.
- [18] P. Wang, D. Niyato and H. Jang, "Voice Service Support over Cognitive Radio Networks," Proc. IEEE ICC, pp. 1-5, Dresden, Germany, June 2009.
- [19] P. Wang, D. Niyato and H. Jang, "Voice-Service Capacity Analysis in Cognitive Radio Networks", IEEE Trans. on Vehicular Techn., vol. 59, no. 4, pp. 1779-1790, May 2010.
- [20] Gunawardena, S. and W. Zhuang, "Voice Capacity of Cognitive Radio Networks," IEEE ICC, pp. 1-5, Cape Town, South Africa, May 2010.
- [21] K. Ali and W. Zhuang, "Link-layer Resource Allocation for Voice Users in Cognitive Radio Networks", in Proc. IEEE ICC, pp. 1-5, Kyoto, Japan, June 2011.
- [22] Z. Wang, T. Jiang, L. Jiang, and X. He, "VoIP Capacity Analysis in Cognitive Radio System with Single/Multiple Channels", 6<sup>th</sup> International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM), pp. 1-4, Chengdu, China, September 2010.
- [23] Y. Mihov and B. Tsankov, "Cognitive System with VoIP Secondary Users over VoIP Primary Users", COGNITIVE: The Third International Conference on Advanced Cognitive Technologies and Applications, pp.30-35, Rome, 2011.
- [24] R. I. Wilkinson. "Theories for toll traffic engineering in the USA", Bell System Technical Journal, 1956, vol.35, no.2, pp.421-514.

# BEP Performance of DE-QPSK and DE-OQPSK over composite fading channels in the presence of imperfect signal extraction

Milica I. Petković<sup>1</sup>, Bojana Nikolić<sup>2</sup>, Bata V. Vasić<sup>3</sup> and Goran T. Đorđević<sup>4</sup>

**Abstract** – In this paper we consider partially coherent detection of differentially encoded quadrature phase-shift keying (DE-QPSK) and differentially encoded offset quadrature phase-shift keying (DE-OQPSK) transmitted over the gamma-shadowed Nakagami- $m$  fading channel. The imperfect carrier signal recovery has been taken into account through the phase error that occurs in the phase-locked loop. The phase error is described by Tikhonov distribution. The expressions for bit-error probability (BEP) are analyzed and numerical results are presented.

**Keywords** – DE-QPSK, DE-OQPSK, BEP, phase noise, phase locked loop.

## I. INTRODUCTION

In wireless communication one of the main problem is fading, which is described by several statistical models as Rayleigh, Rician or Nakagami model. Nakagami model is more general than Rayleigh and Rician. Because of that, it is very often used in observations. The basis in all these fading models is the assumption that the average signal power measured is constant. However, the existence of multiple scattering may lead to the case where the received average power becomes random. This phenomenon is called shadowing. At first the shadowing was modeled using lognormal distribution. However, it was often inconvenient for further analyses because the obtained composite probability density function (PDF) is in integral form. Newly, gamma distribution was proposed because it is mathematically more corresponding model. Since fading and shadowing occur simultaneously in wireless systems, it is necessary to have models that can describe the faded-shadowed channel [1]-[7]. In this paper, we considered the composite signal modeled by gamma-shadowing Nakagami- $m$  fading distribution.

Quadrature phase-shift keying (QPSK) and offset quadrature phase-shift keying (OQPSK) are widely used modulation techniques. Because of the symmetry in two

dimensional signal constellations, ambiguities about the exact phase orientation of the received signal set exist at the receiver. In PSK systems, this ambiguity is resolved using differential encoding (DE). Differentially encoded QPSK (DE-QPSK) and differentially encoded OQPSK (DE-OQPSK) were introduced by researchers as a mean of resolving the carrier phase ambiguity of the data before transmission. One of the methods for detecting DE signals is coherent detection [8]. The phase-locked loop (PLL) is used for carrier signal recovery. The phase error is a difference between the incoming signal phase and the recovered carrier signal phase in PLL. It is a statistical process described by Tikhonov distribution. When the receiver is not ideal, a certain phase error appears.

The expressions for bit-error probability (BEP) of DE-QPSK and DE-OQPSK for the ideal carrier-synchronization over additive white Gaussian noise (AWGN) channel was shown and compared in [9]. After the ideal case, the conditional BEPs of both modulations are given and combined with the different statistics of the phase error in the maximum *a posteriori* (MAP) estimation carrier-synchronization loops to obtain the expressions of average BEPs for both modulations [9].

In this paper, we consider DE-QPSK and DE-OQPSK signals transmission over the gamma-shadowed Nakagami- $m$  fading channel. The expressions for BEP of DE-QPSK and DE-OQPSK, while the imperfect carrier signal recovery has been taken into account through the phase error, are analyzed and numerical results are presented.

## II. CHANNEL MODEL

As mentioned, we consider transmission of the signal over the channel affected by gamma-shadowed Nakagami- $m$  fading.

Let the received signal envelope  $r$  has Nakagami distribution given by [1]

$$p_{r/\Omega}(r/\Omega) = \frac{2m^m r^{2m-1} e^{-\frac{m}{\Omega}r^2}}{\Gamma(m)\Omega^m}, \quad r > 0, \quad (1)$$

where  $m$  is the Nakagami parameter,  $\Omega$  is the average power  $\Omega = E[r^2]$  with  $E$  denoting mathematical expectation and  $\Gamma(\cdot)$  is the gamma function. The  $m$  parameter refers to the fading severity. With lower values of  $m$ , the fading is stronger. In the case  $m=1$ , we have Rayleigh fading, and  $m=\infty$  is the no-fading case.

In the case when the shadowing is present,  $\Omega$  is random variable and has gamma distribution given by [1]

<sup>1</sup> Milica I. Petković is with the Faculty of Electronic Engineering at the University of Nis, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: milicapetkovic86@gmail.com.

<sup>2</sup> Bojana Nikolić is with the Faculty of Electronic Engineering at the University of Nis, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: bojana.nikolic@elfak.ni.ac.rs.

<sup>3</sup> Bata V. Vasić is with the Faculty of Electronic Engineering at the University of Nis, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: bata.vasic@silicon-studio.com.

<sup>4</sup> Goran T. Đorđević is with the Faculty of Electronic Engineering at the University of Nis, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: goran@elfak.ni.ac.rs.

$$p_{\Omega}(\Omega) = \frac{m_s^{m_s} \Omega^{m_s-1} e^{-\frac{m_s}{\Omega_s} \Omega}}{\Gamma(m_s) \Omega_s^{m_s}}, \quad \Omega > 0, \quad (2)$$

where  $\Omega_s = E[x]$  is the gamma shadow area mean power. The parameter  $m_s$  refers to the shadowing severity. With lower values of  $m_s$ , the shadowing influence is stronger. In the case  $m_s = \infty$ , shadowing is not exist.

The composite envelope  $r$  of the gamma-shadowed Nakagami- $m$  faded signal is:

$$p(r) = \int_0^{\infty} p_{r/\Omega}(r/\Omega) p_{\Omega}(\Omega) d\Omega. \quad (3)$$

Substituting (1) and (2) in (3), we have:

$$p(r) = \frac{4}{\Gamma(m) \Gamma(m_s)} \left( \frac{m m_s}{\Omega_s} \right)^{\frac{m+m_s}{2}} \times r^{m+m_s-1} K_{m_s-m} \left( 2r \sqrt{\frac{m m_s}{\Omega_s}} \right) \quad (4)$$

where  $K_{\nu}(\cdot)$  is the modified Bessel function of the second kind and order  $\nu$  and  $\Omega_s = E[r^2] = \overline{r^2}$  is the average power.

The instantaneous SNR per symbol,  $\rho$ , and the average SNR per symbol,  $\rho_0$ , are related by:

$$\frac{\mu^2}{\rho_0} = \frac{\rho}{\rho_0}, \quad \mu > 0, \quad \rho > 0. \quad (5)$$

The distribution of the SNR in gamma-shadowed Nakagami- $m$  fading channel can be obtained using (4) and (5) by applying standard technique of transforming random variables:

$$p(\rho) = \frac{2}{\Gamma(m) \Gamma(m_s)} \left( \frac{m m_s}{\rho_0} \right)^{\frac{m+m_s}{2}} \times \rho^{\frac{m+m_s-2}{2}} K_{m_s-m} \left( 2 \sqrt{\frac{m m_s}{\rho_0}} \rho \right). \quad (6)$$

Remark that  $\rho$  is the instantaneous SNR per symbol, and  $\rho_0$  is the average SNR per symbol. The average SNR per bit is  $\rho_{0b} = \rho_0 / \log_2 M = \rho_0 / 2$  in the case of quadrature modulation formats.

### III. AVERAGE BEP PERFORMANCE IN THE PRESENCE OF CARRIER PHASE ERROR OVER THE GAMMA-SHADOWED NAKAGAMI FADING CHANNEL

#### A. DE-QPSK

For DE-QPSK, the expression for conditional BEP in the presence of a phase error  $\phi_c$ , due to imperfect carrier

synchronization, in the channel under the influence of fading is (similarly as in [9]):

$$P_b(\phi_c; \rho) |_{DE-QPSK} = \frac{1}{2} \operatorname{erfc} \left( \sqrt{\frac{\rho}{2}} (\cos \phi_c - \sin \phi_c) \right) \times \left[ 1 - \frac{1}{2} \operatorname{erfc} \left( \sqrt{\frac{\rho}{2}} (\cos \phi_c - \sin \phi_c) \right) \right] + \frac{1}{2} \operatorname{erfc} \left( \sqrt{\frac{\rho}{2}} (\cos \phi_c + \sin \phi_c) \right) \times \left[ 1 - \frac{1}{2} \operatorname{erfc} \left( \sqrt{\frac{\rho}{2}} (\cos \phi_c + \sin \phi_c) \right) \right] \quad (7)$$

where  $\rho$  represents SNR per symbol.

In order to evaluate the average BEP for DE-QPSK, the PDF of phase error should be known. The conditional BEP must be averaged over that PDF.

The appropriate PDF of loop's phase error in the form of the Tikhonov distribution for DE-QPSK is given by [9]:

$$p_{\phi_c}(\phi_c; \rho) |_{QPSK} = \frac{4 \exp(\rho_{eq}(\rho) \cos 4\phi_c)}{2\pi I_0(\rho_{eq}(\rho))},$$

$$\rho_{eq}(\rho) = \frac{\rho_{PLL}(\rho) S_L |_{QPSK}(\rho)}{16}, \quad -\frac{\pi}{4} \leq \phi_c \leq \frac{\pi}{4}, \quad (8)$$

where  $\rho_{eq}$  represents an equivalent loop's SNR,  $\rho_{PLL}$  is loop's SNR that can be expressed as  $\rho_{PLL}(\rho) = P/N_0 B_L = (E_b/N_0)/(B_L T_b) = (\rho/2)/(B_L T_b)$  ( $B_L$  denotes the one-sided loop bandwidth). Degradation term referred to as "squaring loss",  $S_L$  is given by [9]:

$$S_L |_{QPSK}(\rho) = \frac{\left[ \operatorname{erf} \left( \sqrt{\frac{\rho}{2}} \right) - 2 \sqrt{\frac{\rho}{2\pi}} \exp \left( -\frac{\rho}{2} \right) \right]^2}{1 + \rho - 2 \left[ \sqrt{\frac{\rho}{2}} \operatorname{erf} \left( \sqrt{\frac{\rho}{2}} \right) + \frac{1}{\sqrt{\pi}} \exp \left( -\frac{\rho}{2} \right) \right]^2}. \quad (9)$$

To obtain average BEP, it is necessary to average conditional BEP (7) over PDF of phase error (8). Under the influence of fading, the instantaneous SNR is random variable. It is also required to do averaging of (7) over PDF of instantaneous SNR per symbol (6). The BEP of DE-QPSK is:

$$P_b |_{DE-QPSK} = \int_{\rho=0}^{\infty} \int_{\phi_c=-\pi/4}^{\pi/4} P_b(\phi_c; \rho) |_{DE-QPSK} \times p_{\phi_c}(\phi_c; \rho) |_{QPSK} p(\rho) d\phi_c d\rho. \quad (10)$$



B. DE-OQPSK

For DE-QPSK, the expression for conditional BEP in the presence of a phase error  $\phi_c$ , due to imperfect carrier synchronization, in the channel under the influence of fading is (similarly as in [9]):

$$\begin{aligned}
 P_b(\phi_c; \rho) |_{DE-OQPSK} = & \left[ \frac{1}{2} \operatorname{erfc} \left( \sqrt{\frac{\rho}{2}} \cos \phi_c \right) \right. \\
 & + \frac{1}{4} \operatorname{erfc} \left( \sqrt{\frac{\rho}{2}} (\cos \phi_c - \sin \phi_c) \right) \\
 & + \frac{1}{4} \operatorname{erfc} \left( \sqrt{\frac{\rho}{2}} (\cos \phi_c + \sin \phi_c) \right) \left. \right] \\
 & \times \left[ 1 - \frac{1}{4} \operatorname{erfc} \left( \sqrt{\frac{\rho}{2}} \cos \phi_c \right) \right. \\
 & - \frac{1}{8} \operatorname{erfc} \left( \sqrt{\frac{\rho}{2}} (\cos \phi_c - \sin \phi_c) \right) \left. \right] \\
 & - \frac{1}{8} \operatorname{erfc} \left( \sqrt{\frac{\rho}{2}} (\cos \phi_c + \sin \phi_c) \right) \left. \right]. \quad (11)
 \end{aligned}$$

The appropriate PDF of loop's phase error in the form of the Tikhonov distribution for DE-OQPSK is given by [9]:

$$\begin{aligned}
 p_{\phi_c}(\phi_c; \rho) |_{OQPSK} &= \frac{2 \exp(\rho_{eq}(\rho) \cos 2\phi_c)}{2\pi I_0(\rho_{eq}(\rho))}, \\
 \rho_{eq}(\rho) &= \frac{\rho_{PLL}(\rho) S_L |_{OQPSK}(\rho)}{4}, \quad -\frac{\pi}{2} \leq \phi_c \leq \frac{\pi}{2}. \quad (12)
 \end{aligned}$$

“Squaring loss”,  $S_L$ , for DE-OQPSK is given by [9]:

$$\begin{aligned}
 S_L |_{OQPSK}(\rho) &= \\
 &= \frac{\left[ \operatorname{erf} \left( \sqrt{\frac{\rho}{2}} \right) - \sqrt{\frac{\rho}{2\pi}} \exp \left( -\frac{\rho}{2} \right) \right]^2}{1 + \frac{\rho}{2} - \left[ \sqrt{\frac{\rho}{2}} \operatorname{erf} \left( \sqrt{\frac{\rho}{2}} \right) + \frac{1}{\sqrt{\pi}} \exp \left( -\frac{\rho}{2} \right) \right]^2}. \quad (13)
 \end{aligned}$$

Similar to DE-QPSK, it is necessary to average conditional BEP (11) over PDF of phase error (12) to obtain average BEP for DE-OQPSK. Under the influence of fading, the instantaneous SNR is random variable, and it is also required to do averaging of (11) over PDF of instantaneous SNR per symbol (6). The BEP of DE-QPSK is:

$$\begin{aligned}
 P_b |_{DE-OQPSK} &= \int_{\rho=0}^{\infty} \int_{\phi_c=-\pi/2}^{\pi/2} P_b(\phi_c; \rho) |_{DE-OQPSK} \\
 &\times p_{\phi_c}(\phi_c; \rho) |_{OQPSK} p(\rho) d\phi_c d\rho. \quad (14)
 \end{aligned}$$

IV. NUMERICAL RESULTS

Fig.1. shows DE-QPSK BEP dependence on SNR per bit for different values of  $B_L T_b$ . The system has better performance for lower value of  $B_L T_b$ . In Fig.1. the best performance is for the case when  $B_L T_b$  has the lowest value ( $B_L T_b=0.01$ ) and the parameters  $m$  and  $m_s$  are higher. It is noticed that influence of  $B_L T_b$  is more expressed for higher values of  $m$  and  $m_s$ , i.e. with the lower influence of fading and shadowing.

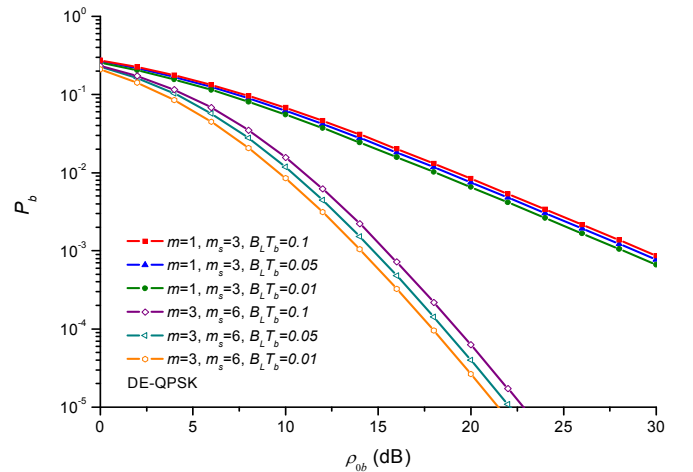


Fig. 1. DE-QPSK BEP dependence on SNR per bit for different values of  $B_L T_b$

The same BEP dependence for DE-OQPSK is shown in Fig.2. The conclusion is the same: System has better performance for lower value of  $B_L T_b$ .

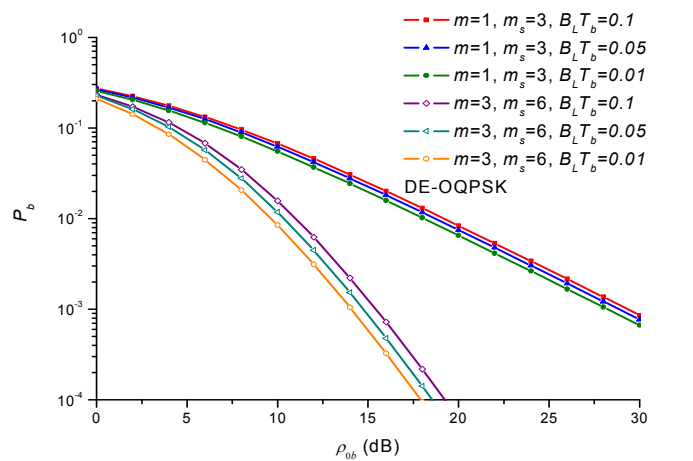


Fig. 2. DE-OQPSK BEP dependence on SNR per bit for different values of  $B_L T_b$

Fig.3. shows DE-QPSK BEP dependence on SNR per bit, for different values of the shadowing parameter  $m_s$ . When the

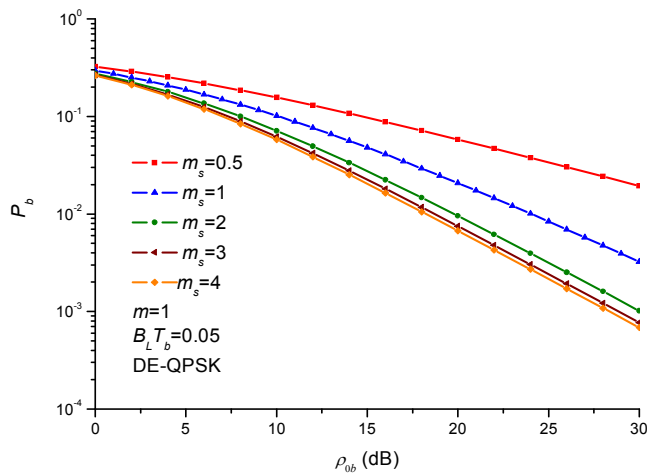


Fig. 3. DE-QPSK BEP dependence on SNR per bit for different values of  $m_s$

values of the parameter  $m_s$  is lower, the influence of shadowing is bigger and we have worse system performance.

Fig.4. shows DE-OQPSK BEP dependence on SNR per bit, for different values of the parameter  $m$ . The performance of the system is the worst when the  $m=0.5$ . With decreasing

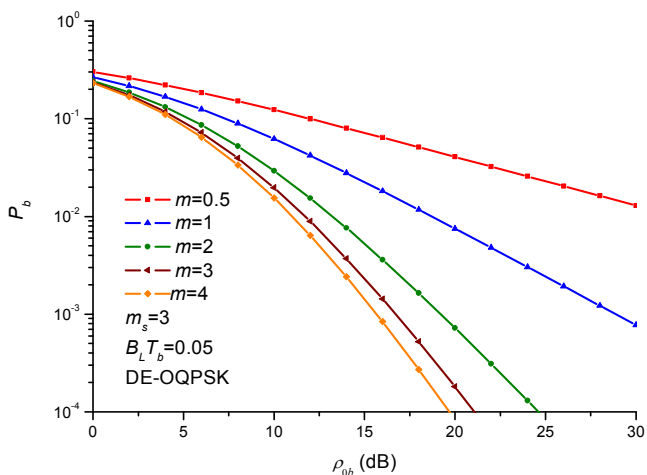


Fig. 4. DE-OQPSK BEP dependence on SNR per bit for different values of  $m$

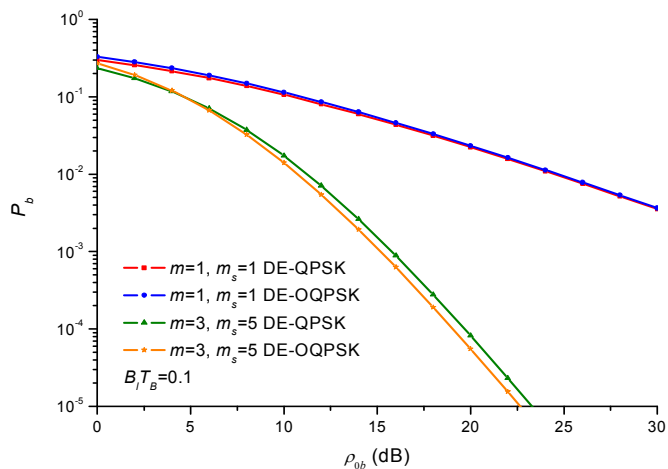


Fig. 5. DE-OQPSK and DE-QPSK BEP dependence on SNR per bit

value of fading parameter  $m$ , we have severe fading.

The DE-QPSK and DE-OQPSK BEP dependences on SNR per bit are shown in Fig.5. We have greater difference between DE-QPSK and DE-OQPSK when the parameters are higher, i.e. when the impacts of fading and shadowing are lower. When the impact of fading and shadowing ( $m=3$  and  $m_s=5$ ) is lower, the DE-QPSK and DE-OQPSK BEPs have same value when SNR per bit is 4.5dB. For lower values of SNR, system has better performance for DE-QPSK. When the SNR per bit is higher than 4.5dB, the system has better performance for DE-OQPSK.

### V. CONCLUSION

In this paper, we have derived the expressions for BEP of DE-QPSK and DE-OQPSK when the phase error introduced by loop's finite SNR is taken into account and the signal is transmitted over the gamma-shadowed Nakagami- $m$  fading channel. The effects of the parameter  $B_L T_b$  and the fading and shadowing parameters on the BEP have been noted.

### ACKNOWLEDGEMENT

This paper was supported in part by the Ministry of Science of Republic of Serbia under grant TR-32028 and in part by the Norwegian Ministry of Foreign Affairs within the project Norwegian, Bosnian and Serbian cooperation platform for ministry and industry in ICT R&D.

### REFERENCES

- [1] I. M. Kostic, "Analytical approach to performance analysis for channel subject to shadowing and fading", IEE Proceedings Communications, vol. 152, no. 6, pp. 821-827, December 2005.
- [2] P. S. Bithas, N. C. Sagias, P. T. 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
- [3] P. M. Shankar, "Error Rates in Generalized Shadowed Fading Channels", Wireless Personal Communications, vol. 28, no.3, pp. 233-238, February 2004.
- [4] V. Milenkovic, N. Sekulovic, M. Stefanovic, M. Petrovic, "Effect of microdiversity and macrodiversity on average bit error probability in gamma shadowed Rician fading channels", ETRI Journal, vol. 32, no. 3, pp. 464-467, Jun. 2010.
- [5] N. Sekulovic, M. Stefanovic, "Performance analysis of system with micro- and macrodiversity reception in correlated gamma shadowed Rician fading channels", Wireless Personal Communications, accepted for publication (published online 12. Feb. 2011.), DOI: 10.1007/s11277-011-0232-8.
- [6] A. Goldsmith, *Wireless Communications*, New York, Cambridge University Press, 2005.
- [7] M. K. Simon, M. S. Alouini, *Digital Communication over Fading Channels: A Unified Approach to Performance Analysis*, New York, John Wiley & Sons, Inc., 2000.
- [8] J. G. Proakis, *Digital Communications*, 4<sup>th</sup> ed., New York: McGraw-Hill, Inc., 2001.
- [9] M. K. Simon, "On the Bit-Error Probability of Differentially Encoded QPSK and Offset QPSK in the Presence of Carrier Synchronization", IEEE Transaction on Communications, vol. 54, no. 5, May 2006.

# Quality of Service (QoS) – main principles and managing tools

Miroslav Slavov<sup>1</sup> and Pencho Penchev<sup>2</sup>

**Abstract** – In this paper the basic concepts behind Quality of Service (QoS), the need for it will be discussed, and several of the types of QoS mechanisms available will be introduced. Quality of Service itself is not something that you configure on a Cisco router, rather it is an overall term that refers to a wide variety of mechanisms used to influence traffic patterns on a network. QoS has already proven itself as the enabling technology for the convergence of voice, video and data networks. As business needs evolve, so do demands on QoS technologies.

**Keywords** – Quality of Service (QoS), Cisco QoS Toolset, QoS levels.

## I. INTRODUCTION

Quality of Service (QoS) is a set of capabilities that allows delivering differentiated services for network traffic, thereby providing better service for selected network traffic. QoS expedites the handling of mission-critical applications, while sharing network resources with noncritical applications.

QoS also ensures the available bandwidth and minimum delays required by time-sensitive multimedia and voice applications. This allows using expensive network connections more efficiently, and to establish service level agreements with customers of the network.

QoS features provide better and more predictable network service by:

- Supporting dedicated bandwidth for critical users and applications.
- Controlling jitter and latency (required by real-time traffic).
- Avoiding and managing network congestion.
- Shaping network traffic to smooth the traffic flow.
- Setting traffic priorities across the network [2][3].

## II. DEFINING QUALITY OF SERVICE

### A. What is QoS?

QoS is the measure of transmission quality and service availability of a network (or internetworks).

Service availability is a crucial foundation element of QoS. The network infrastructure must be designed to be highly

available before you can successfully implement QoS. The transmission quality of the network is determined by the following factors:

- Loss – a relative measure of the number of packets that were not received compared to the total number of packets transmitted. Loss is typically a function of availability.
- Delay – the finite amount of time it takes a packet to reach the receiving endpoint after being transmitted from the sending endpoint.
- Delay variation (Jitter) – the difference in the end-to-end delay between packets [1][4].

### B. Why is QoS Important for Enterprise Networks?

QoS technologies refer to the set of tools and techniques to manage network resources and are considered the key enabling technology for network convergence.

QoS tools are not only useful in protecting desirable traffic, but also in providing deferential services to undesirable traffic such as the exponential propagation of worms [4].

The WAN devices can limit the bandwidth available to the traffic, or give the traffic priority, or even change the classification of the traffic. In this way, you can provide end-to-end QoS in your network.

Figure 1 shows an example of an enterprise network. Typically, you classify traffic in the LAN before sending it to the WAN. The devices on the WAN then use the classification to determine the service requirements for the traffic [3].

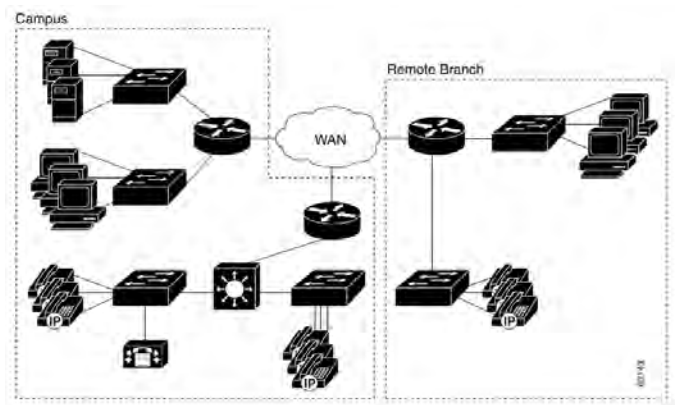


Fig. 1. Example of an Enterprise Network

### C. Which Applications Need QoS?

#### *Understanding the Characteristics of Applications*

It is important to understand the characteristics of the applications that need protection. Some applications tend to be sensitive to latency or packet loss, while others are considered

<sup>1</sup>Miroslav Slavov is with the Faculty of Electrical Engineering and Electronics at Technical University of Gabrovo, 4 H. Dimitar str., Gabrovo 5300, Bulgaria, e-mail: miroslav\_slavov@mail.bg.

<sup>2</sup>Pencho Penchev is with the Faculty of Electrical Engineering and Electronics at Technical University of Gabrovo, 4 H. Dimitar str., Gabrovo 5300, Bulgaria.

"aggressive" because they are bursty or consume a lot of bandwidth. If the application is bursty, determine if there is a constant burst or a small burst. Is the packet size of the application large or small? Is the application TCP or UDP based [5]?

TABLE I  
APPLICATIONS THAT REQUIRED QoS

Characteristic	Guideline
Application that is delay- or loss-sensitive. (Voice and Real Time Video)	Do not use weighted random early detection (WRED), traffic shaping, fragmentation (FRF-12 (describes the method of fragmenting Frame Relay frames into smaller frames)), or policing. For this kind of traffic, you should implement Low Latency Queuing (LLQ) and use a priority queue for the delay-sensitive traffic.
Application that is consistently bursty or is a bandwidth hog. (FTP and HTTP)	Use WRED, policing, traffic shaping, or class-based weighted fair queuing (CBWFQ) to guarantee bandwidth.
Application that is TCP-based.	Use WRED since lost packets cause TCP to back off and then ramp up again using the slow-start algorithm. If the traffic is UDP-based and does not change its behaviour when packets are dropped, do not use WRED. Use Policing if you need to rate-limit the application; otherwise just let the packets tail-drop.

### III. CISCO QoS TOOLSET

This section describes the main categories of the Cisco QoS toolset and includes the following topics (Figure 2):

- Classification and Marking tools
- Policing and Markdown tools
- Scheduling tools
- Link-specific tools
- AutoQoS tools
- Call Admission Control tools

#### A. Classification and Marking Tools

Classification and marking tools set this trust boundary by examining any of the following:

- Layer 2 parameters—802.1Q Class of Service (CoS) bits, Multiprotocol Label Switching Experimental Values (MPLS EXP).
- Layer 3 parameters—IP Precedence (IPP), Differentiated Services Code Points (DSCP), IP Explicit Congestion Notification (ECN), source/destination IP address

- Layer 4 parameters— L4 protocol (TCP/UDP), source/destination ports
- Layer 7 parameters— application signatures via Network Based Application Recognition (NBAR)

NBAR is a Cisco proprietary technology that identifies application layer protocols by matching them against a Protocol Description Language Module (PDLM), which is essentially an application signature. The NBAR deep-packet classification engine examines the data payload of stateless protocols against PDLMs. There are over 98 PDLMs embedded into Cisco IOS software 12.3 code.

Additionally, Cisco IOS software 12.3(4)T introduces the ability to define custom PDLMs which examine user-defined strings within packet payloads.

PDLMs can be added to the system without requiring an IOS upgrade because they are modular. NBAR is dependent on Cisco Express Forwarding (CEF) and performs deep-packet classification only on the first packet of a flow. The remainder of the packets belonging to the flow is then CEF-switched.

Within an enterprise, marking is done at either Layer 2 or Layer 3, using the following fields:

- 802.1Q/p Class of Service (CoS)—Ethernet frames can be marked at Layer 2 with their relative importance by setting the 802.1p User Priority bits of the 802.1Q header. Only three bits are available for 802.1p marking. Therefore, only 8 classes of service (0-7) can be marked on Layer 2 Ethernet frames.
- IP Type of Service (ToS) byte—Layer 2 media often changes as packets traverse from source to destination, so a more ubiquitous classification occurs at Layer 3. The second byte in an IPv4 packet is the ToS byte. The first three bits of the ToS byte are the IPP bits. These first three bits combined with the next three bits are known collectively as the DSCP bits.
- DSCPs and Per-Hop Behaviors (PHBs)—DSCP values can be expressed in numeric form or by special standards-based names called Per-Hop Behaviors. There are four broad classes of DSCP PHB markings: Best Effort (BE or DSCP 0), RFC 2474 Class Selectors (CS1-CS7, which are identical/backwards-compatible to IPP values 1-7), RFC 2597 Assured Forwarding PHBs (AFxy), and RFC 3268 Expedited Forwarding (EF).

DSCP values can be expressed in decimal form or with their PHB keywords. For example, DSCP EF is synonymous with DSCP 46, and DSCP AF31 is synonymous with DSCP 26.

- IP Explicit Congestion Notification (IP ECN)—IP ECN, as defined in RFC 3168, makes use of the last two bits of the IP ToS byte, which are not used by the 6-bit DSCP markings, as shown in Figure 3.

#### B. Policing and Markdown Tools

Policing tools (policers) determine whether packets are conforming to administratively-defined traffic rates and take action accordingly. Such action could include marking, remarking or dropping a packet.

A basic policer monitors a single rate: traffic equal to or below the defined rate is considered to conform to the rate, while traffic above the defined rate is considered to exceed the rate. On the other hand, the algorithm of a dual-rate policer (such as described in RFC 2698) is analogous to a traffic light.

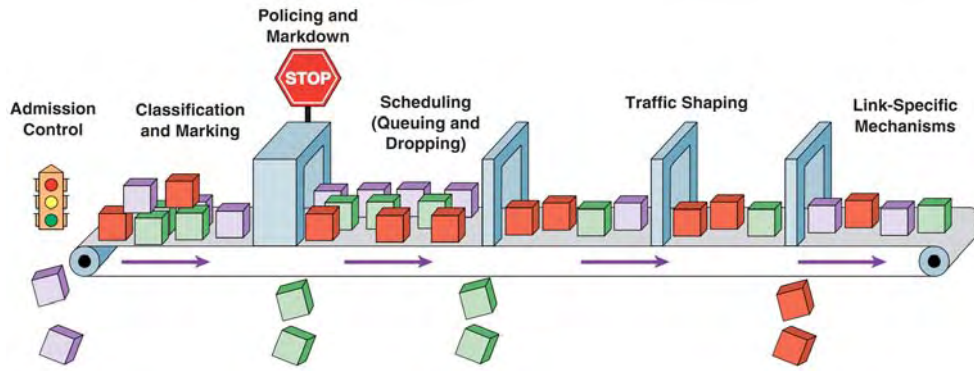


Fig. 2. The Cisco QoS Toolset

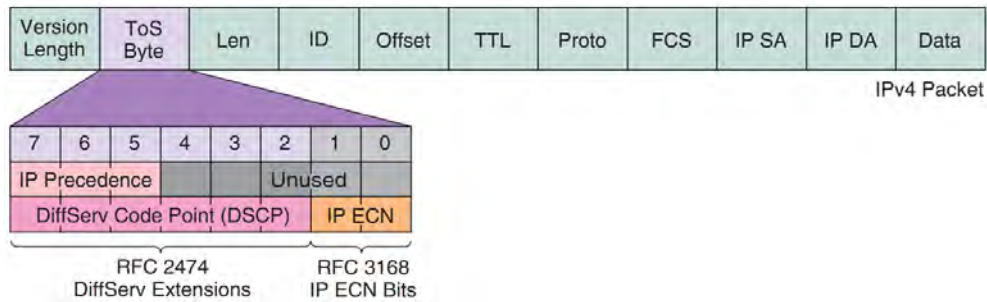


Fig. 3. The IP ToS Byte (DSCP and IP ECN)

C. Scheduling Tools

Scheduling tools determine how a frame/packet exits a device. Whenever packets enter a device faster than they can exit it, such as with speed mismatches, then a point of congestion, or bottleneck, can occur. Devices have buffers that allow for scheduling higher-priority packets to exit sooner than lower priority ones, which is commonly called queuing.

Queuing algorithms are activated only when a device is experiencing congestion and are deactivated when the congestion clears. Figure 4 shows the Layer 3 and Layer 2 queuing subsystems of the Cisco IOS (Internetwork Operating System) software LLQ/CBWFQ algorithm.

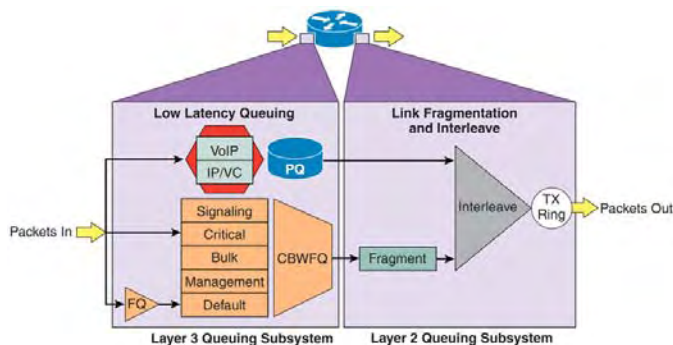


Fig. 4. The IP ToS Byte (DSCP and IP ECN)

Selective dropping of packets when the queues are filling is referred to as congestion avoidance. Congestion avoidance

mechanisms work best with TCP-based applications because selective dropping of packets causes the TCP windowing mechanisms to “throttle-back” and adjust the rate of flows to manageable rates.

The principle IOS congestion avoidance mechanism is WRED, which randomly drops packets as queues fill to capacity.

D. Link-Specific Tools

Link-specific tools include the following:

- Shaping tools—A shaper typically delays excess traffic above an administratively-defined rate using a buffer to hold packets and shape the flow when the data rate of the source is higher than expected.

- Link Fragmentation and Interleaving tools—With slow-speed WAN circuits, large data packets take an excessively long time to be placed onto the wire.

- Compression tools—Compression techniques, such as compressed Real-Time Protocol (cRTP), minimize bandwidth requirements and are highly useful on slow links. At 40 bytes total, the header portion of a VoIP packet is relatively large and can account for nearly two-thirds or the entire VoIP packet (as in the case of G.729 VoIP).

- Transmit ring (Tx-Ring) tuning—The Tx-Ring is a final interface First-In-First-Out (FIFO) queue that holds frames to be immediately transmitted by the physical interface. The Tx-Ring ensures that a frame is always available when the interface is ready to transmit traffic, so that link utilization is driven to 100 % of capacity. The size of the Tx-Ring is

dependent on the hardware, software, Layer 2 media, and queuing algorithm configured on the interface. E. AutoQoS Tools

The richness of the Cisco QoS toolset inevitably increases its deployment complexity. To address customer demand for simplification of QoS deployment, Cisco has developed the Automatic QoS (AutoQoS) features. AutoQoS is an intelligent macro that allows an administrator to enter one or two simple AutoQoS commands to enable all the appropriate features for the recommended QoS settings for an application on a specific interface.

For Campus Catalyst switches, AutoQoS automatically performs the following tasks:

- Enforces a trust boundary at Cisco IP Phones.
- Enforces a trust boundary on Catalyst switch access ports and uplinks/downlinks.
- Enables Catalyst strict priority queuing for voice and weighted round robin queuing for data traffic.
- Modifies queue admission criteria (CoS-to-queue mappings).
- Modifies queue sizes as well as queue weights where required.
- Modifies CoS-to-DSCP and IP Precedence-to-DSCP mappings.

For Cisco IOS routers, AutoQoS is supported on Frame Relay (FR), Asynchronous Transfer Mode (ATM), High-Level Data Link Control (HDLC), Point-to-Point Protocol (PPP), and FR-to-ATM links.

The AutoQoS Enterprise feature consists of two configuration phases, completed in the following order:

- Auto Discovery (data collection)—Uses NBAR-based protocol discovery to detect the applications on the network and performs statistical analysis on the network traffic.
- AutoQoS template generation and installation—Generates templates from the data collected during the Auto Discovery phase and installs the templates on the interface. These templates are then used as the basis for creating the class maps and policy maps for your network. After the class maps and policy maps are created, they are then installed on the interface.

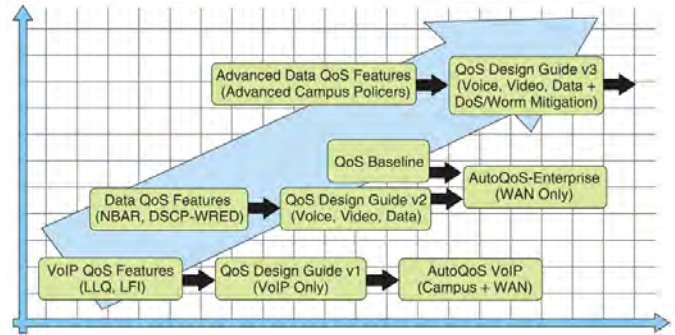


Fig. 5. Cisco QoS Feature, Design Guide and AutoQoS Evolution

Figure 5 shows the relationship between Cisco QoS features, Design Guides, and AutoQoS.

#### IV. CONCLUSION

The purpose of this paper is to describe the terms and the main concept of Quality of Service (QoS). In the paper a set of tools, used for maintaining the QoS are described. These are the basics, needed for future researches of QoS and finding a way to improve it in different types of networks.

#### ACKNOWLEDGEMENT

This paper has been sponsored by E 1102 project of Technical University of Gabrovo.

#### REFERENCES

- [1] M. Flannagan, etc "Administering Cisco QoS in IP Networks", Syngress, Rockland, 2001, ISBN: 1-928994-21-0.
- [2] T. Szigeti, "End-to-End QoS Network Design", Cisco Press, Indianapolis, 2005 ISBN 1-58705-176-1.
- [3] "User Guide for CiscoWorks QoS Policy Manager" Software Version 4.1.5, Cisco Systems Inc, 2010.
- [4] "Enterprise QoS Solution Reference Network Design Guide" Version 3.3, Cisco Systems Inc, 2008.
- [5] "Implementing Quality of Service", Document ID: 13747, Cisco Systems Inc, 2008 – 2009.

# Pitch perception in complex sound

Marko Janković<sup>1</sup> and Dejan Ćirić<sup>2</sup>

**Abstract** – Sound of an instrument represents a typical example of a complex sound. Such a sound can be decomposed in a set of simple (pure) sounds. In a number of cases, an instrument sound can be considered to consist of a fundamental and harmonics that give it a unique characteristic, depending on their number, distribution, amplitude and envelope. The ability of a human auditory system to perceive pitch in a complex sound is associated with the harmonics. In this paper, we will show how distribution of harmonics in complex sound affects the pitch perception. For that purpose, spectral analysis is performed and listening tests are applied. Special attention is paid to the phenomenon known as “missing fundamental”.

**Keywords** – complex sound, harmonics, auditory system, pitch perception.

## I. INTRODUCTION

Ability that a human auditory system perceives pitch in a complex sound is of extreme importance for understanding music and speech, and for analysing sound images [1]. Many instruments generate specific sounds, whose main attribute to create a melody is pitch. In verbal communication, pitch carries prosodic information in languages such as English and semantic information in tonal languages such as Mandarin and Chinese. On the basis of differences in fundamental frequencies, we can differentiate sounds and order them in appropriate music scales [1].

From an acoustical point of view, music instruments represent sources that generate sounds of specific characteristics including spectral content. Majority of instrument sounds consist of a fundamental (the lowest partial) and upper partials that can be harmonically related to fundamental or not [2]. Factors affecting perception of such sounds are related to the number, distribution, amplitude and envelope of the partials. Very interesting feature that has attracted attention during the previous decades is pitch perception, and especially a phenomenon known as “missing fundamental”. This has been a motive to investigate the pitch perception in complex sounds and the mentioned phenomenon in more detail here.

For the purpose of investigating pitch perception, various complex sounds are first generated and then analyzed. The pitch perception in such complex sounds is examined through the listening tests. The impact of harmonics distribution on the pitch perception is observed. Emphasis is placed on studying the perceptual consequences of the phenomenon “missing fundamental”.

<sup>1</sup>Marko Janković is a PhD student at the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: marestudio2004@gmail.com

<sup>2</sup>Dejan Ćirić is with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: dejan.ciric@elfak.ni.ac.rs

## II. SOUND AND PERCEPTION

The simplest possible sound is a pure sound representing a sinusoidal signal (sound). When a sound consists of two or more pure sounds (partials), it represents a complex sound that can be given by:

$$s(t) = \sum_{n \geq 1} A_n \cos(2\pi f_n t + \theta_n), \quad (1)$$

where  $A_n$  is the amplitude,  $f_n$  is the frequency and  $\theta_n$  is the phase. The sound with the lowest frequency is called fundamental, while other sounds whose frequencies are higher are called overtones. An interesting example of sound is obtained when all of the frequencies are integer multiples of a given (fundamental) frequency  $f$ ,  $f_n = nf$ . The sound signal is then periodic with period  $T = 1/f$ . The fundamental (caused by the first mode of vibration) determines the pitch of the corresponding complex sound (tone). The frequency  $f_n$  (sound at that frequency) is called the  $n$ -th harmonic. Accordingly, the fundamental is the first harmonic. Perceptually, a sound containing only few harmonics is perceived as “poor”, while a sound containing a number of harmonics is perceived as “rich”.

Sounds of all music instruments consist of many partials. Some of them are harmonically related to fundamental and some are not [2]. According to Hornbostel-Sachs system, all music instruments can be classified into five groups depending on what actually makes the sound [3]. These groups are: ideophones, membranophones, chordophones, aerophones and electrophones. In both aerophones and chordophones, there is something long and thin that basically vibrates in one dimension: a vibrating string or column of air. This produces strong resonances, and sound tends to be pure with a specific pitch. In membranophones, the membrane can basically vibrate in two dimensions, while in idiophones, the body of the instrument can vibrate in three dimensions. As the number of dimensions goes up, the resonances become more complex and weaker, and therefore, the sounds become more complex and with more diffuse pitch [3].

Generally, pitch represents an auditory sensation defined in various ways. The most common one given by American National Standards Institute (ANSI) in 1994 describes pitch as an “attribute according to which the sounds can be arranged in a musical scale, from lowest to highest” [1]. Pitch is closely related to frequency, but there is no absolute equivalency between those two since the frequency is an objective concept, while pitch is a subjective.

With modern methods for measuring brain activity (fMRI, EEG, MEG), it becomes possible to locate the area of auditory cortex that corresponds to pitch perception called *Heschl's gyrus*, and it is positioned in the right cerebral hemisphere [1,4]. In the auditory cortex, there are two types of neurons:

multi-peak and single-peak neurons [1,4]. Multi-peak neuron can be excited with multiple frequencies that are harmonically related. A frequency required to have minimum energy to excite a multi-peak neuron is called the central frequency and it is equal to frequency of the first harmonic (fundamental). The other frequencies are required to have more energy to make the same excitation [1,4]. Single-peak neuron is excited only by a single frequency [1]. In the auditory cortex, about 20 % of neurons represent multi-peak neurons (30 Hz-5 kHz), whereas 80 % are single-peak neurons (5 kHz-20 kHz) [1,4].

Important frequencies for understanding human speech are ranged up to about 4 kHz [1]. On the other hand, the highest frequency of a fundamental tone of symphony orchestra has a piccolo flute, and it is about 4.5 kHz [1]. The tests have shown that people best understand the melody up to 5 kHz; beyond this frequency understanding of melody is difficult or impossible [1].

When the first harmonic (fundamental) is excluded from a complex sound whose partials are harmonically related, we still can perceive the pitch equal to the frequency of missing fundamental. This phenomenon is known as “missing fundamental” [1].

### III. METHODS OF INVESTIGATION

For the purpose of spectral analysis of various instrument sounds, the samples from Steinberg’s sounds library called *Hypersonic II* were used. Sample rate of all samples was 44.1 kHz while 16 bits were used for quantization.

Sounds of some instruments belonging to the groups chordophones (acoustic guitar, harp, piano and violin), aerophones (oboa and flute), membranophones (timpani, kick, snare and tambourine) and ideophones (closed hi-hat, open hi-hat, crash cymbal, finger snap, rattle and effects of wind) are spectrally analyzed. Instruments from the group electrophones are excluded from the analysis because they generate sounds identical to those of instruments from other groups.

Since pitch perception is a subjective attribute, it is investigated here through the listening tests. They were carried out in a typical listening situation (in a typical living room). In the listening tests, different complex sounds were presented to the subjects over the headphones (*Beyerdynamic DT-770 Pro*), while the subjects’ response was recorded by the condenser microphone (*AKG C444*) and stored in the corresponding audio tracks using *HP 110 Mini computer* with *Intel Atom N270* processor and integrated sound board. The task of the subjects was to repeat pitch of the reproduced sound by singing neutral syllable *La* or to give the answers to particular questions.

Ten subjects participated in the listening tests of which 7 were male and 3 were female. The mean age of the subjects was 26. Nine subjects have graduated from the Faculty of Arts (Music Arts) and one subject is a student of the secondary music school. In that regard, they can be considered to be expert listeners.

The listening tests are divided in five groups – test 1 to test 5. In the listening test 1, the complex sounds of duration of 2 seconds consisting of fundamental and 4 harmonics were presented to the subject by the headphones. The frequency of

fundamental is 440 Hz, while the frequencies of harmonics are: 880 Hz, 1320 Hz, 1760 Hz and 2200 Hz. The spectrum of such a complex sound is given in Fig. 1. The task of the subjects was to repeat the pitch of the reproduced sound by singing neutral syllable *La*.

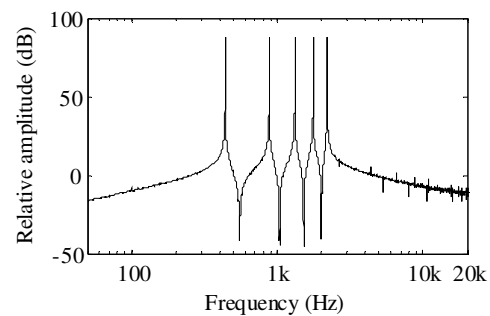


Fig. 1 Spectrum of generated complex sound used in the listening test 1

In the listening test 2, the complex sound used in the listening test 1, but without fundamental was presented to the subjects in the same manner as in the test 1. The task of the subjects was also to repeat the pitch of the reproduced sound by singing the neutral syllable *La*. The purpose of this test is to study the phenomenon “missing fundamental”.

Two complex sounds separated by silence of 1 second were presented to the subjects in the listening test 3. The first sound is the same as in the test 1, while the second one is the same as in the test 2. The spectrogram of complex sound applied in the test 3 is given in Fig. 2. This time, the task of the subjects was to answer a question: “Is the difference between two reproduced sounds in pitch or in timbre?”

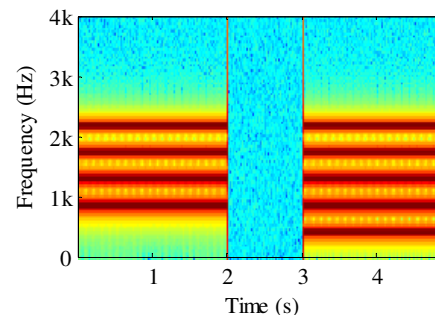


Fig. 2 Spectrogram of generated complex sound used in the listening test 3

In the listening test 4, the complex sound of duration of 2 seconds consisting of 5 harmonics with random distribution was presented to the subjects in the same manner as in other tests. The frequencies of harmonics are: 440 Hz, 475 Hz, 483 Hz, 502 Hz, 543 Hz. The spectrum of this complex sound is given in Fig. 3. The task of the subjects was the same as in the tests 1 and 2 - to repeat the pitch of the reproduced sound by singing the neutral syllable *La*.

Similarly as in the test 2, the first partial (with frequency of 440 Hz) of the signal from test 4 was removed, and such a complex signal was presented to the subjects in the listening test 5. The task of the subjects was the same - to repeat the pitch of the reproduced sound on the neutral syllable *La*. The



purpose of this test is to compare the results with those from the tests 1, 2 and 4 in studying the phenomenon “missing fundamental”.

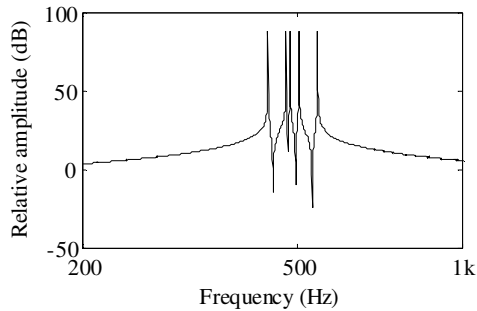


Fig. 3 Spectrum of complex sound consisting of 5 partials in non-harmonic distribution used in the listening test 4

#### IV. RESULTS

Performed spectral analysis shows that chordophones and aerophones instruments have harmonically related partials, while membranophones and ideophones instruments do not have a harmonic distribution of partials. Representative examples of instrument sound spectra are given in Fig. 4.

The spectrum of piano (Fig. 4(a)) has harmonically related partials. The frequency of the first harmonic (fundamental) is 215 Hz and frequencies of overtones are: 430 Hz, 645 Hz, 860 Hz, 1075 Hz, 1290 Hz, etc. The spectrum of oboa also has harmonically related partials (see Fig. 4(b)). The frequency of the fundamental is 656 Hz and the frequencies of overtones are: 1312 Hz, 1968 Hz, 2624 Hz, etc. On the other hand, snare and crash chinel do not generate harmonically related partials (see Fig. 4(c) and (d)).

In the listening test 1, all 10 subjects were able to successfully repeat the pitch of the reproduced complex sound containing fundamental and 4 harmonics. A representative example of a subject’s response is given in Fig. 5. From presented spectrum and spectrogram, it can be seen that frequency positions of fundamental and harmonics are close to those from the test sound.

The results of the listening test 2 are similar to those of the test 1. So, all 10 subjects were able to successfully repeat the pitch of the reproduced complex sound without fundamental (but with harmonically related partials), that is, they were able to reproduce the pitch of missing fundamental. A representative example of the response of a subject is given in Fig. 6.

In the listening test 3, where the subjects compared two complex sounds (with and without fundamental) separated by silence of 1 second, 9 subjects responded that there was a difference in timbre, and 1 subject responded that there was a difference in pitch of compared sounds.

The results of the listening test 4 are different than those from the tests 1 and 2. Since the partials of the presented complex sound are not in harmonic relation, the subjects were not able to perceive and reproduce (sing) the same pitch. The distribution of answers, that is, distribution of the frequencies of fundamental of sang neutral syllable *La* is shown in Fig. 7.

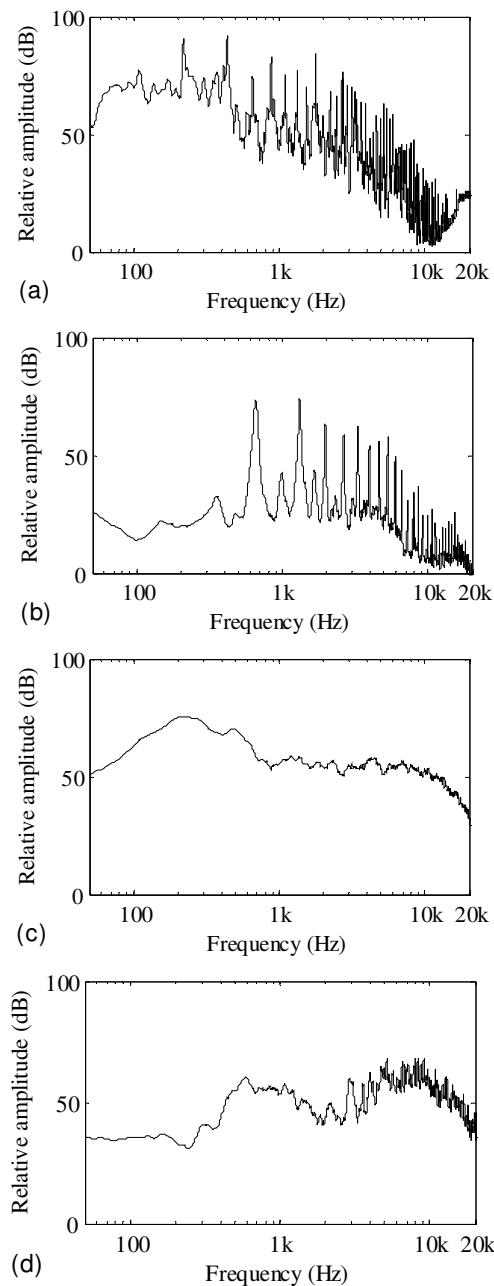


Fig. 4. Spectra of sounds of representative instruments: piano (a), oboa (b), snare (c) and crash cymbal (d)

Similarly to the results of the test 4, the perceived and reproduced pitch of complex sound without first partial presented in the listening test 5 is not the same for all subjects. The distribution of the fundamental frequency of the subjects’ responses (sang neutral syllable *La*) is shown in Fig. 8.

#### V. CONCLUSION

From the previously mentioned results, it can be concluded that the distribution of partials (harmonics) in complex sound plays a key role in the pitch perception. In the case of

harmonically distributed partials in a complex sound, the most energy excites the multi-peak neuron(s), so that the impression of pitch is unambiguously clear to the listener. In absence of fundamental frequency, other harmonics excite multi-peak neuron whose central frequency corresponds to the fundamental, and the listener's impression of pitch corresponds to that of the fundamental frequency. This principle can explain the phenomenon "missing fundamental".

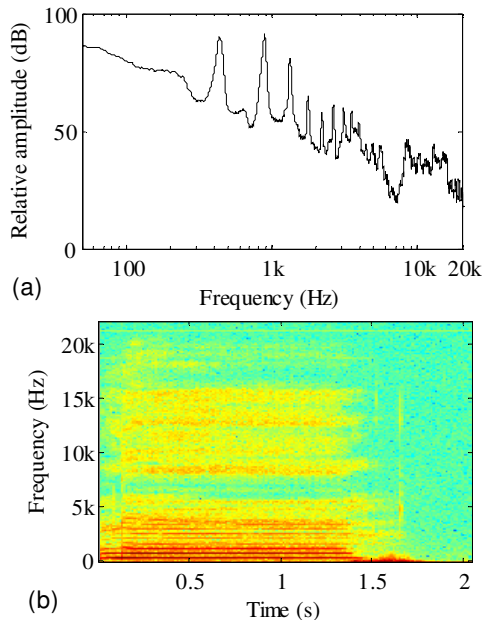


Fig. 5. Representative response of a subject in the listening test 1: spectrum (a) and spectrogram (b) of the sang neutral syllable *La*

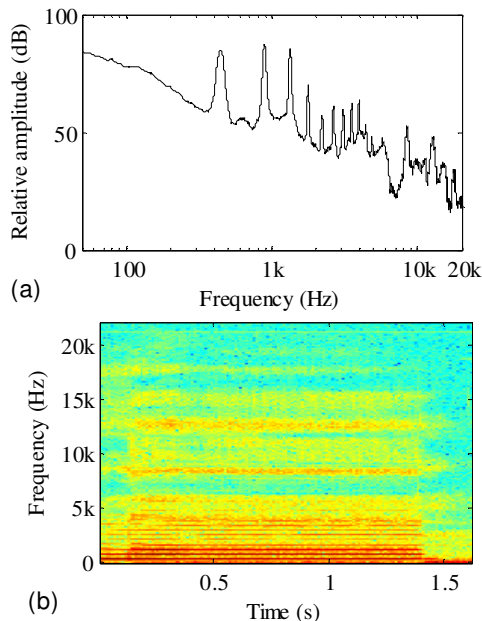


Fig. 6. Representative response of a subject in the listening test 2: spectrum (a) and spectrogram (b) of the sang neutral syllable *La*

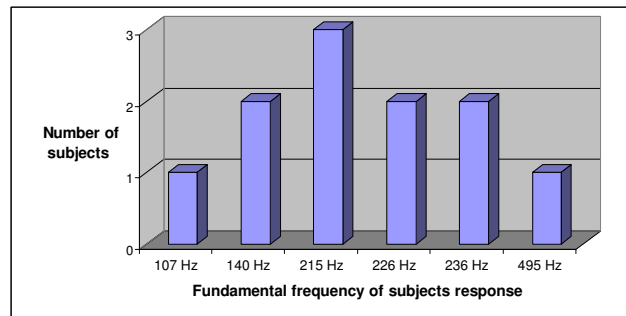


Fig. 7. Distribution of fundamental frequency of subjects' responses (sang neutral syllable *La*) from the test 4

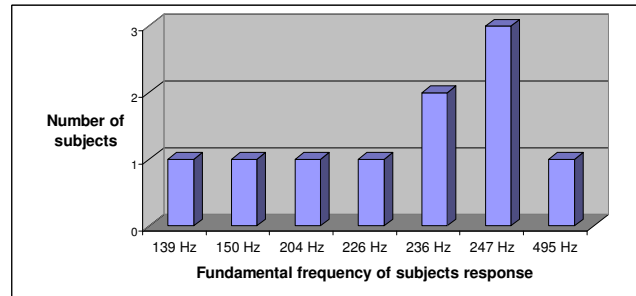


Fig. 9. Distribution of fundamental frequency of subjects' responses (sang neutral syllable *La*) from the test 5

ACKNOWLEDGEMENT

This research is partially supported by the Ministry of Science and Technological Development of Serbia through the project No. 44009.

REFERENCES

- [1] Xiaoqin Wang, "The Harmonic Organization of Auditory Cortex", *The Neurophysiological Bases of Auditory Perception*, pp. 211-222, 2010.
- [2] William A. Sethares, "Tuning, Timbre, Spectrum Scale", Second Edition, London, Springer-Verlag, 2005.
- [3] Martin Russ, "Sound Synthesis and Sampling", Third Edition, Oxford, Focal Press, 2009.
- [4] Jennifer K. Bizley, Kerry M. M. Walker, "Sensitivity and Selectivity of Neurons in Auditory Cortex to the Pitch, Timbre, and Location of Sounds", *The Neuroscientist*, Vol. 16, No. 4, pp.453-469, 2010.
- [5] B. Kostek, A. Czyzewski, "Representing Musical Instrument Sounds for their Automatic Classification", *J. Audio Eng. Soc.*, vol. 49, no. 9, pp. 768-785, 2001.
- [6] S. Ando, K. Yamaguchi, "Statistical Study of Spectral Parameters in Musical Instrument Tones", *J. Acoust. Soc. Am.*, vol. 94, no. 1, pp. 37-45, 1993.
- [7] J. Meyer, "The Sound of the Orchestra", *J. Audio Eng. Soc.*, vol. 41, no. 4, pp. 203-213, 1993.
- [8] J.C. Brown, "Musical Fundamental Frequency Tracking Using a Pattern Recognition Model", *J. Acoust. Soc. Am.*, vol. 92, no. 3, pp. 1394-1402, 1992.

# Digital Bandpass IIR Filters with High Selectivity

Peter Apostolov<sup>1</sup>

**Abstract** – This paper proposes an optimal third-degree polynomial, which approximates Kronecker’s delta function with high precision. The polynomial is obtained by a new approximation method, called “method with compressed cosines”. The method is based on Chebyshev’s optimality norm. The polynomial is used for narrow bandpass IIR filter design. The filter’s selectivity depends on the parameter  $Q$  without increasing the polynomial’s order. With the proposed method an IIR filter with 5(6) multipliers, a very narrow passband and a high stopband attenuation can be designed.

**Keywords** – IIR digital filters, Frequency response, Polynomial approximation.

## I. INTRODUCTION

The task of filter synthesis is a mathematical problem of approximating ideal functions with rectangular shape. The transfer function of the filter results from the approximation. The aim is to obtain a mathematical relationship which has the lowest computational complexity and approximation error. In approximations with polynomials, this indicator is the degree of the polynomial.

This paper will show a method for digital filter design based on a polynomial approximation with “compressed cosines”.

## II. BACKGROUND

In some practical cases the passband filter is required to have a very narrow bandwidth. The ideal characteristic of a super-narrowband filter is Kronecker’s delta function

$$\delta(x) = \begin{cases} 1, & x = 0.5 \\ 0, & x \neq 0.5 \end{cases}; x \in [0, 1]. \quad (1)$$

This is a transfer function of a filter that has a pass bandwidth equal to zero, stopband gain equal to zero and an infinite steepness of its characteristic. It cannot be realized in practice. Hence, Kronecker’s delta function needs to be approximated by another one, which can be realized. The approximation is carried out with a preset accuracy  $\epsilon > 0$ . The difference between the ideal function and the approximating polynomial defines the error function. The two most popular norms for the approximation are  $L_2$  - weighted integral least-

squares norm, and  $L_\infty$  - weighted Chebyshev’s norm. In the literature different polynomial approximation methods are proposed. Fig.1 shows approximations with Hausdorff (Chebyshev) polynomial [1] using  $L_\infty$  norm, with sinc(.)

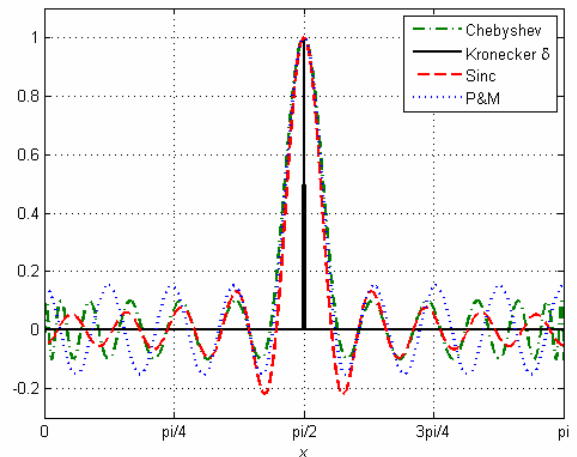


Fig. 1. Polynomial approximations of the Kronecker’s delta

function using  $L_2$  norm, method of Parks-McClellan [2] with trigonometric polynomial using  $L_\infty$  norm.

It is seen that a suitable trade-off between the flatness in the stopband and the bandwidth must be done. In all the criteria, the functions have the oscillations in the stopband. These oscillations are undesirable. The goal is to obtain a rectangular shape of the ideal function, that has maximally flat pass band and stop band, and narrowest possible bandwidth. In  $L_2$  case the oscillations increase near the main lobe. This is due to the Gibbs’ phenomenon [3]. In the approximations using  $L_\infty$  norm the oscillations are with equal amplitude. These approximations are known as optimal and equiripple.

The approximations with rational functions [4, 5], have better properties than the polynomials approximations. The most popular are Chebyshev, Butterworth and Cauer.

In [5, 6] a polynomial approximation method with compressed cosines is proposed. With this method a third degree polynomial with significantly better properties than the other polynomials approximations is derived. The approximation accuracy is close to the approximations with rational functions. The polynomial has the form

$$P_3 = \sum_{k=1}^4 b_k \cos [(k-1)\varphi], \quad (2)$$

with coefficients:

<sup>1</sup>Peter Apostolov is with the Department of Wireless Communications and Broadcasting at the College of Telecommunications and Posts, Sofia 1700, 1 Acad Stefan Mladenov St, Bulgaria. E-mail: p\_apostolov@abv.bg

$$b_1 = 0.5 - \varepsilon; b_2 = b_4 = 0; b_3 = -0.5. \quad (3)$$

The function

$$\varphi = -\pi - 2\text{arctg} \left[ Q \left( \frac{2f - 2f_0}{f_d} - \frac{f_d}{2f - 2f_0} \right) \right], \quad (4)$$

is the phase response off the allpass lattice filter with quality factor  $Q$ ,  $f \in [0, f_d/2]$  is frequency,  $f_d$  is the sampling

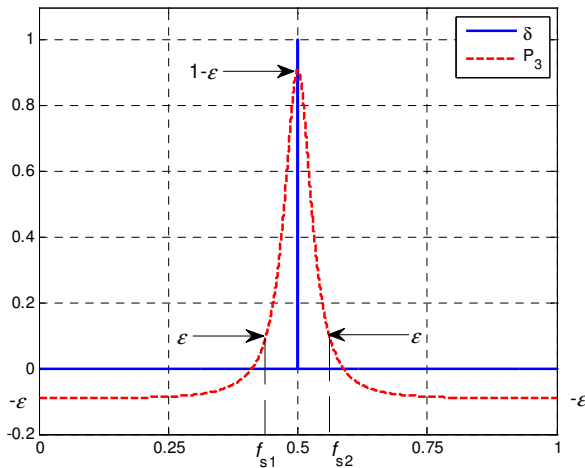


Fig. 2. Approximation of the Kronecker's delta by a third-degree optimal polynomial

frequency,  $f_0$  is the middle frequency of the passband.

Fig. 2 shows an optimal approximation of Kronecker's delta function by an optimal 3<sup>rd</sup> degree polynomial.  $f_{s1}$  and  $f_{s2}$  are the two normed stopband frequencies. Their difference defines the bandwidth  $\Delta f_{stop}$ . The passband is defined by  $\Delta f_{pass}$  - the bandwidth at level  $-3\text{dB}$ . The approximation error  $\varepsilon$  determines the stopband attenuation  $DS$ , and the quality factor  $Q$  the bandwidth  $\Delta f_{stop}$ . The filter's coefficients are obtained by those of the polynomial:

$$h_k = b_4/2, b_3/2, b_2/2, b_1, b_2/2, b_3/2, b_4/2. \quad (5)$$

The filter's transfer function has the form

$$H = -0.25 + (0.5 - \varepsilon) \exp(-j\varphi) - 0.25 \exp(-j2\varphi). \quad (6)$$

### III. DESIGN EXAMPLE

The realization will be demonstrated with an example of a bandpass digital IIR filter design with the following specification: middle frequency in the pass band  $f_0 = 800$  Hz;  $\Delta f_{pass} = 100$  Hz; sampling rate  $f_d = 8000$  Hz; degree of the polynomial  $m = 3$ ; attenuation in the stopband  $DS \geq 20$  dB.

A normalization of the frequencies is done as

$$\Delta f_n = 2\Delta f_{pass} / f_d = 0.025; f_{0n} = 2f_0 / f_d = 0.2. \quad (7)$$

The optimal approximation error is determined as

$$\varepsilon = \frac{1}{1 + 10^{DS/20}} = 0.0909. \quad (8)$$

When the normalized bandwidth  $\Delta f_n$  is set, the  $Q$ -factor can be determined approximately by the following five equations:

$$Q = f_{0n} \sum_{k=1}^{11} a_k \Delta f_n^{11-k}; \Delta f_n \in [0.1, 0.0127); \quad (9)$$

$a_1=5.6075\text{e}14$	$a_2=-3.348\text{e}14$	$a_3=8.7823\text{e}13$	$a_4=-1.3321\text{e}13$
$a_5=1.2944\text{e}12$	$a_6=8.4311\text{e}10$	$a_7=3.7445\text{e}9$	$a_8=-1.1304\text{e}8$
$a_9=2.2671\text{e}6$	$a_{10}=-2.8705\text{e}4$	$a_{11}=1885.71$	

$$Q = f_{0n} \sum_{k=1}^{11} a_k \Delta f_n^{11-k}; \Delta f_n \in [0.0127, 1.272\text{e}-3); \quad (10)$$

$a_1=1.2838\text{e}25$	$a_2=-9.0581\text{e}23$	$a_3=2.79655\text{e}22$	$a_4=-4.9689\text{e}20$
$a_5=5.6239\text{e}18$	$a_6=-4.2396\text{e}16$	$a_7=2.1628\text{e}14$	$a_8=-7.435\text{e}11$
$a_9=1.6817\text{e}9$	$a_{10}=-2.3764\text{e}6$	$a_{11}=206.1499$	

$$Q = f_{0n} \sum_{k=1}^{11} a_k \Delta f_n^{11-k}; \Delta f_n \in [1.272\text{e}-3, 3.14\text{e}-4); \quad (11)$$

$a_1=9.934\text{e}34$	$a_2=-7.838\text{e}32$	$a_3=2.7414\text{e}30$	$a_4=-5.6059\text{e}27$
$a_5=7.4403\text{e}24$	$a_6=-6.7267\text{e}21$	$a_7=4.2269\text{e}18$	$a_8=-1.8471\text{e}15$
$a_9=5.5072\text{e}11$	$a_{10}=-1.0692\text{e}8$	$a_{11}=1.2226\text{e}4$	

$$Q = f_{0n} \sum_{k=1}^{11} a_k \Delta f_n^{11-k}; \Delta f_n \in [3.14\text{e}-4, 5.965\text{e}-5); \quad (12)$$

$a_1=-1.6808\text{e}41$	$a_2=2.5067\text{e}38$	$a_3=-1.4578\text{e}35$	$a_4=3.5739\text{e}31$
$a_5=1.3647\text{e}27$	$a_6=-3.3608\text{e}24$	$a_7=1.016\text{e}21$	$a_8=-1.6545\text{e}17$
$a_9=1.6554\text{e}13$	$a_{10}=-1.0265\text{e}9$	$a_{11}=36979.1$	

$$Q = f_{0n} \sum_{k=1}^{11} a_k \Delta f_n^{11-k}; \Delta f_n \in [5.965\text{e}-5, 5.5\text{e}-6). \quad (13)$$

$a_1=-4.7623\text{e}49$	$a_2=1.2294\text{e}46$	$a_3=-1.1141\text{e}42$	$a_4=2.0403\text{e}37$
$a_5=3.8601\text{e}33$	$a_6=-3.5797\text{e}29$	$a_7=1.5185\text{e}25$	$a_8=-3.7635\text{e}20$
$a_9=5.697\text{e}15$	$a_{10}=-5.176\text{e}10$	$a_{11}=263267.1$	

By substituting the defined in (7)  $\Delta f_n = 0.025$  into (9),  $Q = 5.0714$  is obtained. This allows for defining the transfer function of the allpass lattice filter. The coefficients of the denominator of the transfer function are determined by those of the denominator of the Butterworth bandpass filter of first order with bandwidth

$$\Delta f_{Butt} = f_0 / Q = 157.7469 \text{ Hz}. \quad (14)$$

The coefficients in the numerator are the same as in the denominator, but in reverse order.

$$H_{AP}(z) = \frac{c_1 + c_2 z^{-1} + z^{-2}}{1 + c_2 z^{-1} + c_1 z^{-2}}; z = \exp(-j\omega);$$

$$c_1 = 0.8832; c_2 = -1.5265. \quad (15)$$

Then

$$\exp(-j\varphi) = H_{AP}(z) = \frac{0.8832 - 1.5265z^{-1} + z^{-2}}{1 - 1.5265z^{-1} + 0.8832z^{-2}}. \quad (16)$$

In Fig.3, the diagram of the designed filter is shown. A criterion for comparing the selectivity of the digital filters is the number of multipliers with which they are realized.

The coefficients having the same value are realized with

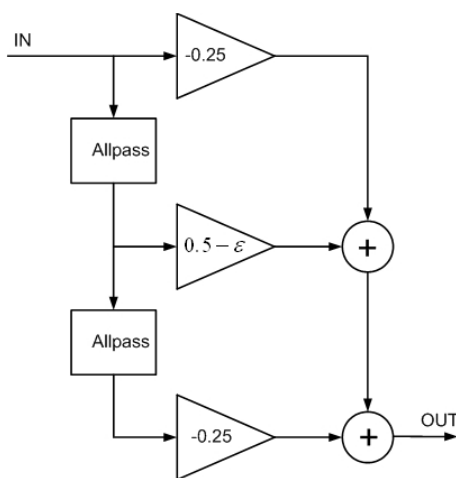


Fig. 3. Functional diagram of the filter

one multiplier to reduce the power consumption. As it is known, the allpass lattice filters are realized with 4 multipliers. Therefore, the total number of multipliers in the diagram is 11. This scheme can be realized with only six multiplier as both allpass filters are the same and  $h_1 = h_3 = -0.25$ . If the filter's coefficients are multiplied by 4, then  $h_1 = h_3 = -1$ . Then the filter will be realized with 5

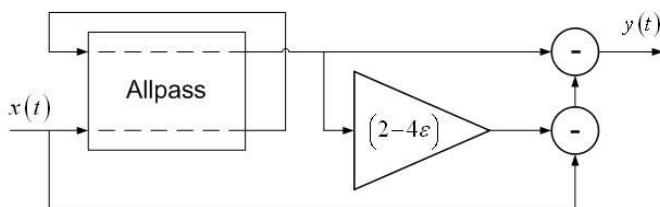


Fig. 4. Functional diagram of the filter with 5 multipliers

multipliers, as in digital signal processing the change of the sign with the operation  $x = -x$  is performed. In this case the filter will amplify the signal four times (12dB).

The scheme shows that all signals are summed. Therefore in the design the sequence of the operations is irrelevant, in

accordance with the commutative law. This allows the input signal to pass twice through one allpass lattice filter. The bandpass filter is implemented with the scheme in Fig.4

Fig. 5 shows the magnitude response in dB ( $10 \lg x$ ). Fig. 6 shows an output response of a computer simulation of the filter with 5 multipliers. The filter input is fed with a discretized at 8000Hz linear chirp signal with frequency sweep from 1 to 4000Hz, amplitude of  $\pm 0.25V$  and duration of 10 seconds. It is seen that the filter's output response corresponds to the input specification.

This implementation requires frame signal processing with

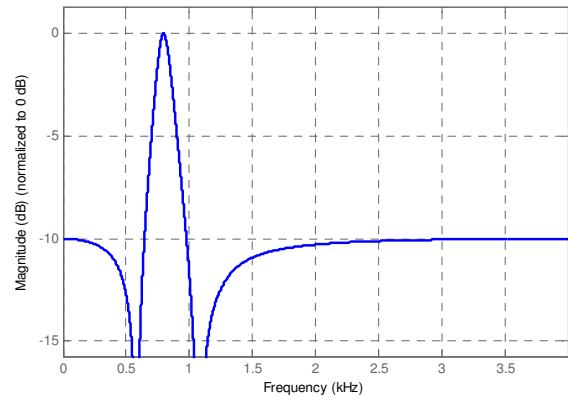


Fig. 5. Magnitude response

buffers. The concatenation of two neighbor fragments is treated with "overlap" to remove the uncertainties, which result from the filtration at the beginning of each fragment.

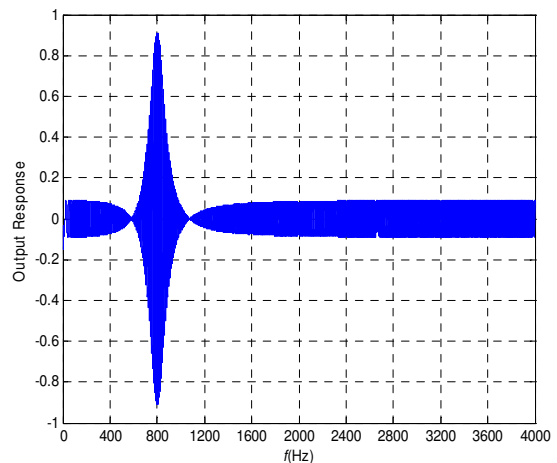


Fig. 6. Computer simulation – output response

#### IV. DISCUSSION

An advantage of the method of compressed cosines is that the approximation is carried out with third-degree polynomial. The polynomial's coefficients are calculated easily. To obtain a high selectivity, it is not necessary to increase the degree of

the polynomial, as in other polynomial approximations, but to use an allpass lattice filter with high  $Q$ -factor. The bandwidth of the stopband  $\Delta f_{stop}$  is a result of the approximation. It can not be defined in the input specification. With a third-degree

2. Butterworth, Chebyshev and Cauer filters of second order are implemented by 7-9 multipliers. Fig.9 shows a similar comparison. In this case the magnitude response of the filter using compressed cosines with 5 multipliers has a lower selectivity.

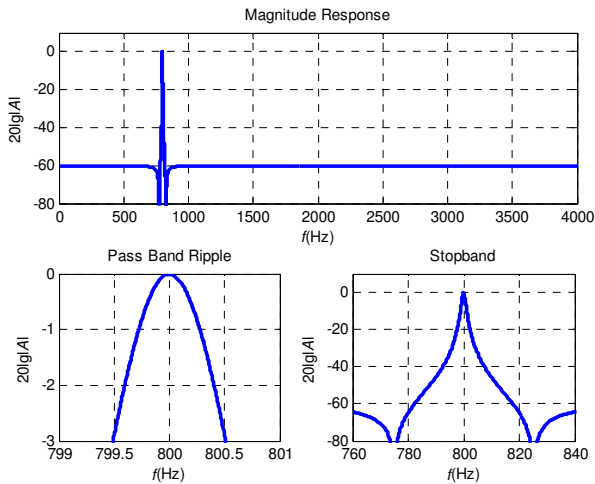


Fig.7. Bandpass filter - magnitude response

polynomial, a filter with an arbitrary bandwidth  $\Delta f_{pass}$  and a stopband attenuation  $DS$  can be realized. For example, Fig. 7 shows the magnitude response of a filter with a passband of 1Hz and a stopband attenuation of 60dB.

The most commonly used IIR digital filters are those of Butterworth, Chebyshev and Cauer. A criterion for comparing the selectivity is the number of multipliers.

1. The scheme of the digital filter of Fig. 4 is realized always with 5 multipliers, regardless of the filter's specification. This comes at the expense of using a larger volume of the memory, which is not a serious disadvantage. The filters of Butterworth, Chebyshev and Cauer of first order are implemented by 4 multipliers. Due to the low order, their magnitude responses are identical. Fig.8 compares magnitude responses of bandpass filters with equal bandwidth  $\Delta f_{pass}$ . The magnitude response of the filter using compressed cosines has better selectivity.

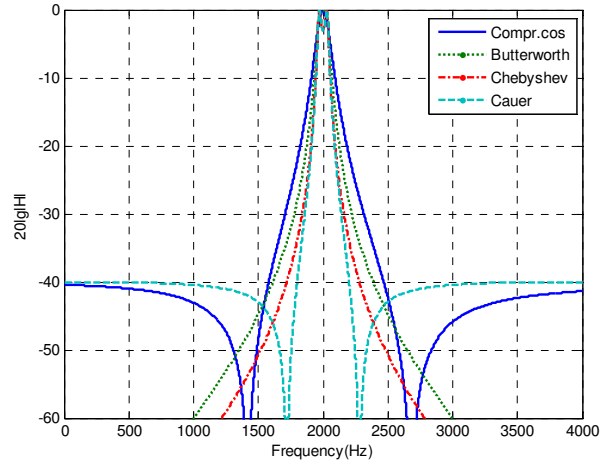


Fig. 9. Comparison of magnitude responses

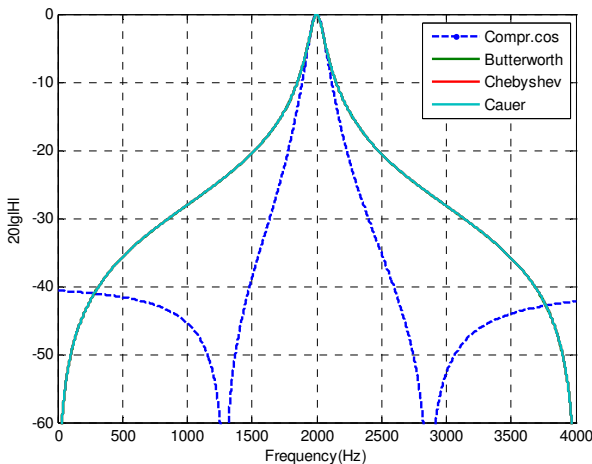


Fig. 8. Comparison of magnitude responses

### V. CONCLUSION

The obtained results show that the selectivity of the filters with “compressed cosines” is determined by the steepness of the  $S$ -curve of the allpass filter's phase response ( $Q$ -factor). From (4) it is seen that it is the function  $\text{arctg}(\cdot)$ . To obtain a high selectivity it is necessary to use a function with a greater gradient, e.g.  $\tanh(\cdot)$ . Unfortunately, an allpass filter with such a phase response has not been realized until now.

The proposed method may be a good alternative in several applications in IIR bandpass filter design.

### REFERENCES

- [1] Sendov, B. *Hausdorff Approximations*. Kluwer Academic Publishers London 1990, ISBN: 0792309014.
- [2] Parks, T. W. and J. H. McClellan. A Program for the Design of Linear Phase FIR Digital Filters. *IEEE Trans. on Audio and Electroacoustics*, Vol. AU – 20, №3, pp. 196-199, August 1972.
- [3] B. Porat, *A Course in Digital Signal Processing*. New York: Wiley, 1997.
- [4] Daniels, R., *Approximation Methods for Electronic Filter Design*. McGraw Hill, 1974.
- [5] Schaumann, R., M.E. van Valkenburg. *Design of Analog Filters*, Oxford University Press 2001.
- [6] Apostolov, P. S. Linear Equidistant Antenna Array with Improved Selectivity, *IEEE Transaction on Antennas and propagation*, Vol.59, Issue10, pp.3940-3943, Aug. 2011.
- [7] Apostolov, P. S. Method for FIR filter design with compressed cosine using Chebyshev's norm. *Signal Processing Elsevier*, Vol. 91, Issue 11, pp. 2589-2594, Nov.2011.

# Features of time-frequency analysis visualization of large dynamic range signals

Tihomir Trifonov<sup>1</sup>, Ivan Simeonov<sup>2</sup> and Rosen Dzhakov<sup>3</sup>

**Abstract** – In this paper practical time-frequency analysis of large dynamic range signals is made. The acoustic signals from unique Bulgarian bells recorded by PULSE 12 data acquisition hardware are analyzed.

The mechanical and acoustical properties of these objects are discussed and their features are determined. Some specific techniques of visualization are proposed.

**Keywords** – time-frequency analysis, acoustic signal visualization, conformal map

## I. INTRODUCTION

At work on the project about research of the valuable Bulgarian bells the database was created, in which almost all characteristics of the bells were included [1]. Database includes acoustical records, obtained via unique measurement set. The used measurement and processing methods can be implemented to the other purposes.

The bell is a complicated sound source with a very wide frequency range and an unique dynamic range of the transmitted signal. Its spectrum contains infrasound, sound and ultrasound partials. The dynamic range is very large too and it cannot be detected entirely by human ear whose dynamic range of perception is about 120dB. The best all over the world measurement set with a corresponding measurement microphone at this moment was used because of this [2]. For example, this set is able to process the signal without distortions with dynamic range up to 160dB.

The most modern processing methods and integrated system of computer mathematics MatLab are used [3]. The features of the source and the raw records require this way of measurement [4].

In this paper we propose a new method of presentation of some transformations (Fourier Transform and Wavelet Transform for example). Wavelet Transform gives the improvement for analysis and reception via conversion of 2D signal into pseudo 3D signal. Conform transformation improves these possibilities in addition. We introduce “sound print” as analog of the “fingerprint”, used in the criminology

<sup>1</sup>Tihomir Trifonov is with the Dep. of Algebra and Geometry at Veliko Tarnovo University "St. St. Cyril and Methodius", 3 Arch. Georgi Kozarov Str., V. Turnovo BG-5000, Bulgaria, e-mail: tihomirtrifonov@ieee.org

<sup>2</sup>Ivan Simeonov is with the Dep. of CIS at National Military University Vasil Levski of Veliko Turnovo, 76 Bulgaria Blvd, V. Turnovo BG-5006, Bulgaria, e-mail: ivanov\_ivan@nvu.bg.

<sup>3</sup>Rosen Dzhakov is with the Faculty of Artillery, Air-Defense and CIS Faculty of NMU, 1 Karel Shkorpil Str., Shumen BG-9713 Shumen

and biometrics. The detection and classification of the transmitted object becomes easily. This way the methods of biometric iris recognition can be used for process automation.

## II. METHOD OF “SOUNDPRINT” VISUALIZATION

### A. Wavelet transform applications

Every signal (or function of time)  $\varphi(t)$ , can be described by

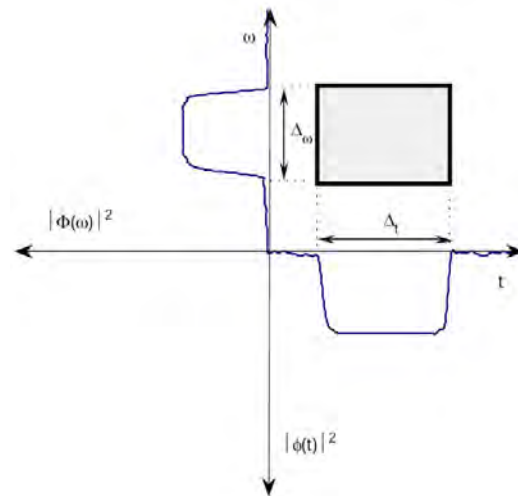


Fig. 1 Localization characteristics of  $\varphi(t)$

interval  $\Delta_t$  in the time axis and interval  $\Delta_\omega$ , in frequency which are including 90% of his energy, concentrated around center of mass of functions  $|\varphi(t)|^2$  and  $|\Phi(\omega)|^2$ . The modulation on this function is translation of the rectangle across axis  $\omega$ , while the scaling of function (her contraction or



Fig. 2. The disposition of unique bells, denoted as **Melnik1** -1270AD and **Melnik2**-1220 AD, in the National Historical Museum in Sofia.

stretching) changes the rectangle proportions.

In this case, the function  $\varphi(t)$  can be represented as rectangle on the plane  $Ot\omega$ , as shown in fig. 1.

Unlike to Fourier transform - FT and Short time fourier transform – STFT, wavelet transform will alters the rectangle type for analysis according to the frequency, area of rectangle will stay constant. An illustration of local properties of wavelets in frequency area is shown on figure 3. This is a kind of analyze where, the relation  $\omega_0/2\Delta_\omega$  is constant or the quality factor Q is equal. The time-frequency window area stay  $4\Delta_t\Delta_\omega$  for a different scales, where  $\Delta_m^2$  и  $\Delta_\omega^2$  are the second central moments on the functions  $\psi(t)$  and  $\Psi(\omega)$ .

More precisely, suppose that  $a \in \mathbb{R}^+, b \in \mathbb{R}$ , or (a,b) determine one point in right-half plane, then the continuous wavelet transform (CWT) of a continuous, square-integrable function is expressed by:

$$CWT_f(a,b) = \langle f(t), \psi_{a,b} f(t) \rangle = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \psi^* \left( \frac{t-b}{a} \right) dt, \quad (1)$$

where  $\langle , \rangle$  denotes the inner product.

The wavelet transform of a one-dimensional signal is a two-dimensional time-scale joint representation, [7]. So the resolution of identity must be satisfied, that is expressed as

$$f(t) = \frac{1}{c_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} CWT_f(a,b) \frac{1}{\sqrt{|a|}} \psi \left( \frac{t-b}{a} \right) db \frac{da}{|a|^2}, \quad (2)$$

where,

$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi \left( \frac{t-b}{a} \right)$  is basis that satisfy the conditions of admissibility (the mean value equal to zero), regularity (has exponential decay, so that its first low order moments are equal to zero), and orthogonality, see fig. 3,

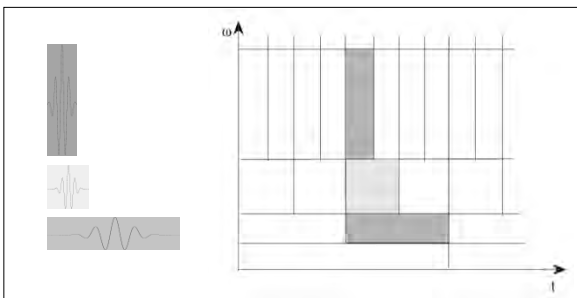


Fig. 3 Wavelet basis functions  $\Psi_{a,b}(t)$ , and time-frequency plane image, Continuous Wavelet Transform – CWT

$$C_\psi = \int_{-\infty}^{\infty} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < \infty - \text{the admissibility constant,}$$

$$a = 2^k - \text{scale parameter, } k = 0,1,2, \dots,$$

$$b - \text{time shift parameter.}$$

### B. Experimental set-up

An experimental set-up was realized to record the sound of unique bells in the National Historical Museum, Sofia, [4,5].

These objects denoted as **Melnik2**-1220 AD. and **Melnik1** -

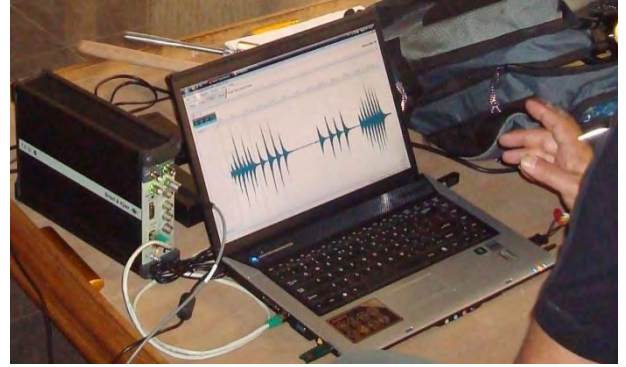


Fig. 4. Data Acquisition Unit 3560B Brüel & Kjær

1270AD are shown in Fig.2 [2]. The measuring microphone 4193 Brüel&Kjær and Data Acquisition Unit 3560B Brüel&Kjær [2] is illustrated in Fig.4.

Experimental set-up includes:

- Pressure-field Microphone Type 4193 Brüel&Kjær,[2] available in *Transducer Electronic Data Sheet (TEDS)* combinations with the classical Preamplifier Type 2669 with an individual calibration; Dynamic Range: 19 ... 162 dB, Sensitivity: 12.5mV/Pa.
- Vibration Transducer TRV-01 SPM Instrument;
- Compact Data Acquisition Unit 3560B Brüel & Kjær, [2] for outdoor use that consist: Dyn-X input modules with a analysis range exceeding 160 dB and automatic detection of front-end hardware and transducers – supports IEEE 1451.4-capable transducers with TEDS (Transducer Electronic Data Sheet); output TCP/IP protocol communication - RJ 45 connector complying with IEEE-802.3100baseX; Multiframe Control option;
- Base software PULSE 12 for CPB (Constant Percentage Band) analysis 2 channels; 5-channel Time Capture; PULSE Bridge to MATLAB®
- MathWorks Software - MatLab&Simulink, toolboxes for FFT and Wavelet analysis.[3]

It can be seen, that the hardware equipment and the software manufacturers are known for theirs high quality all over the world. A part of equipments are shown in Fig.4.

### C. Conformal map soundprint visualization

Wavelet analysis as a tool allows a deeper analysis of sound frequencies. The using of scalogram plots was providing new pictures for complex sounds structure.

In this section we analyze the Bell sounds structures. The preprocessed signals named “melnik1-1.mat” from **Melnik1** - 1270AD Bell and „melnik2.mat“ **Melnik2**-1220 AD are obtained by Brüel & Kjær’s Data Acquisition Unit 3560B,[6].

The Continuous Wavelet Transform signals calculations were produced in MatLab, Continuous Wavelet 1-D tool [3].



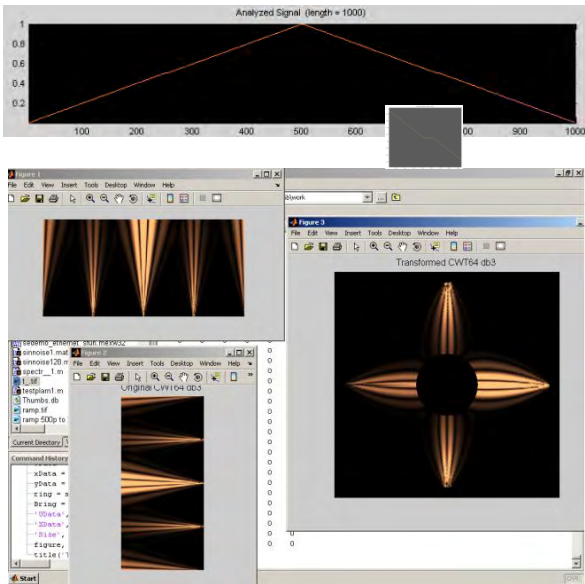


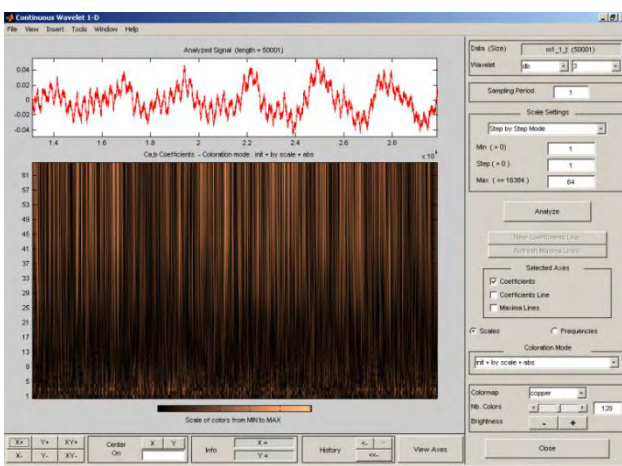
Fig. 5 An conformal mapping example, “NearBreaks I” test signal.

If we make known conformal mapping - logarithm function  $\ln(z)$ , the rectangular graph will be transformed to a circular graph. An example is given in Fig 5, where the signal is “Near Breaks I” test signal.

Fragments of bell strikes Melnik1 and Melnik2 in the time scale are shown in Fig.6 and Fig.8 respectively. The images that result from continuous wavelet transformations are illustrated to Fig.6 and Fig.8 in the bottom.

These scalogram coefficients are calculated by Daubechies wavelets order 3.

On the figures 7 and 9 are illustrated the conformal map visualizations of the same bell strikes (Melnik1\_10 and Melnik2\_8) for various number of coefficients



Фиг. 6 Fragment of the signal Melnik1\_10 - the strike tail, as well as its respective scalogram - CWT coefficients.

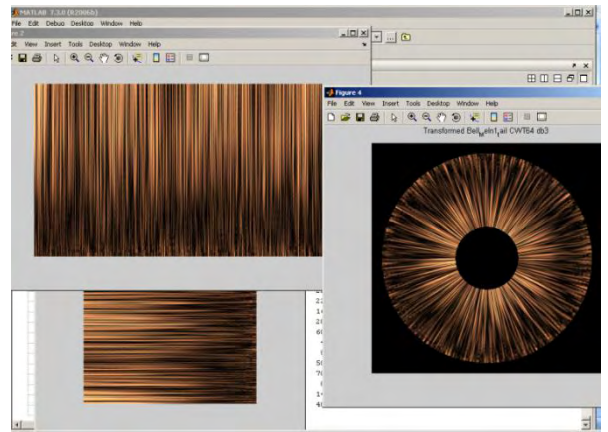


Fig.7 Conformal map visualization of the signal Melnik1\_10

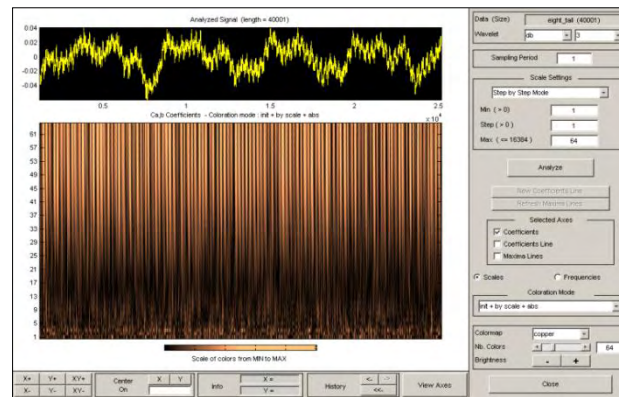


Fig. 8 Fragment of the eight strike tail and its respective scalogram (Melnik2\_8) - CWT coefficient. Daubechies wavelets order 3.

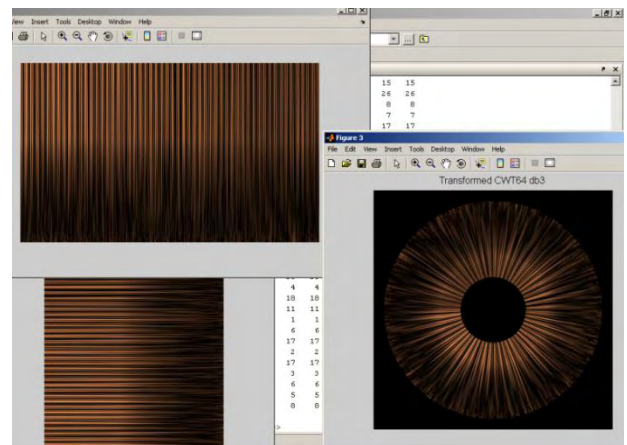


Fig. 9 Conformal map visualization of the signal Melnik2\_8

In the figures 10 was shown fragments of the strikes (in tails) in left and in the right was shown their conformal maps.

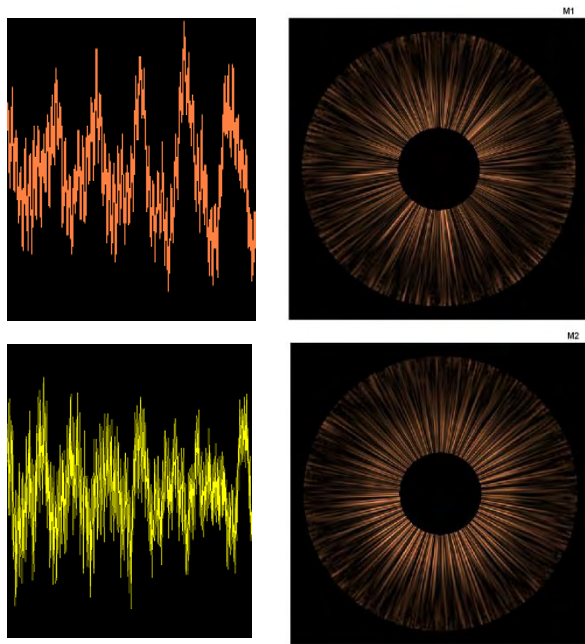


Fig. 10 Visualization of the signals Melnik1\_10( top) vs Melnik2\_8(bottom).

### III. ADDITIONAL REMARKS

The possibility for analyze and perception is improved by transformation on two-dimensional acoustic signal image to pseudo three-dimensional (scalogram).

Additional improvement in perception is achieved by using a conformal mapping of the obtained scalogram, because there are well-known iris recognition techniques that can be applied.

### IV. CONCLUSION

In the presented paper we propose a new method of presentation of some transformations (Fourier Transform and Wavelet Transform for example). Wavelet Transform, except its well known advantages, mentioned above, gives the improvement for analysis and reception via conversion of 2D signal into pseudo 3D signal (scalogram). Conform transformation improves these possibilities in addition. We introduce “sound print” as analog of the “fingerprint”, used in the criminology and biometrics. The detection and classification of the transmitted object becomes easily. This way the methods of biometric iris recognition can be used for process automation.

Our future work will be pointed to obtain more reasons of practical implementation of proposed method.

### REFERENCES

- [1] BELL: Research and Identification of Valuable Bells of the Historic and Culture Heritage of Bulgaria and Development of Audio and Video Archive with Advanced Technologies), KIN-1009, Institute of Mathematics and Informatics, Bulgarian Academy of Sciences , (<http://www.math.bas.bg/bells>)
- [2] Brüe&Kjar, [www.bksv.com](http://www.bksv.com)
- [3] MathWorks, [www.mathworks.com](http://www.mathworks.com)
- [4] Tihomir Trifonov, Tsvetanka Georgieva, Web based approach to managing audio and video archive for unique Bulgarian bells, Proceedings of the Tenth International Conference on System Analysis and Information Technologies, Kyiv, Ukraine, 2008, page 325
- [5] G. Dimkov, Al. Aleksiev, I. Simeonov, T. Trifonov, K. Simeonov, Acoustical researches on historical valuable Bulgarian bells, In Proceedings of the National Scientific Conference on Acoustics, Sofia, Bulgaria, 2008, pp. 115-124 (in Bulgarian).
- [6] Tihomir Trifonov, Georgi Dimkov, Rosen Dzhakov, Ivan Simeonov, Research and identification of valuable bells of the national historic and cultural heritage of Bulgaria, In Proceedings of the XXII Conference with International Participation "Noise and Vibration", Nish, Serbia, 2010, pages 103-107.
- [7] Alexander D. Poularikas, *The Transforms and Applications Handbook*, Second Edition, CRC Press, 1999

### APPENDIX

Bell’s donation inscriptions XIII century:

**Bell Melnik2**, 1211-1216 year

Material: bronze, Place: Tower-belfry on the metropolitan church of St. Nicholas, Melnik.

*"† The bell /is/ forged out of copper, a gift from despot Alexii, † pious Slav to St. Nicholas, he who is from Mira."*

**Bell Melnik1**, 1270

Material: bronze, Place: Belfry of the monastery of St. Charalambius - Saints Achangels, Melnik.

*"Lord, help your servant Theodosii monk who for the first time has created (sanctified) a bell for strategus Mihail, the one that is in Melnik, this one that has been fixed under the reign of Michael Paleologus, the new Konstantine. In March, indiction 12 year 6778 (=1270)."*

# Accuracy Improvement of Allpass-based Digital Hilbert Transformers

Kamelia Nikolova<sup>1</sup> and Georgi Stoyanov<sup>1</sup>

**Abstract** – A design procedure to reduce the deviation of the phase from 90° of allpass-based digital Hilbert transformers is proposed. This is achieved by introducing the phase sensitivity minimization of each individual allpass section in the cascade realizations of the two branches of the structure used. The effectiveness of the proposed design is experimentally proven.

**Keywords** – digital filters, allpass filters, Hilbert transformers, sensitivity minimizations.

## I. INTRODUCTION

Hilbert transformers (HT) are very important building blocks in both, analog and digital signal processing. They are used in telecommunications for generation of analytic and single-sideband signals [1] [2] and in many other modulation and demodulation schemes (mainly for splitting the narrow-band signals to two (*I* and *Q*) components), in complex signals processing, in audio and video signal processing, and even in fields like mechanical vibration signal processing. Many approaches and methods of design of digital HTs have been developed in the last 50 years and most of them have been well systematized in [3]. The FIR based HTs are providing easily a linear phase response and unconditional stability but at the price of a very high transfer function (TF) order (say, several hundred), producing quite a high total delay and requiring higher power consumption. These disadvantages are eliminated in the IIR realizations, most often based on the usage of allpass structures. The theory of the allpass-based HTs is quite mature and several design methods using real or complex allpass structures have been summarized in [3]. Many new optimization-based methods for design of half-band filters and HTs have been proposed since then (including even frequency response masking technique [4]), but no specific methods for accuracy improvement have been reported. Meanwhile the practical importance of the HTs grew considerably with the extension of the frequency ranges and the growth of the proportion of the narrow-band signals, described as analytic, in telecommunications. The problem with the accuracy of the realization of the HTs is of paramount importance in many of these telecommunication applications, like in the maintenance of *I* and *Q* channels balance in a wide frequency range. When the HTs are realized using a fixed-point arithmetic (what is often the case in the portable and mobile communication equipment), the limited word-length may reduce considerably that accuracy and

special measures have to be taken to prevent that. Higher accuracy could be achieved by designing the HTs with higher TF order, but the portability of the equipment is imposing another constraint – the power supply limitation. The main aim of this work is to try to improve the accuracy of the allpass-based HTs throughout minimization of their sensitivities. It will reduce the computational load and will permit shorter word-length and lower power consumption for given accuracy. The design procedures should be straightforward, without iterative and complicated optimization steps, in order to be easily used by practicing engineers and the structures have to be with the lowest possible TF order and complexity.

## II. DESIGN PROCEDURE

An ideal Hilbert transformer (also known as a 90-degree phase shifter) is described in frequency domain as [5]

$$H_{HT}(e^{j\omega}) = \begin{cases} -j, & 0 \leq \omega < \pi \\ j, & -\pi \leq \omega < 0 \end{cases} \quad (1)$$

A way to synthesize an IIR Hilbert transformer (called also a complex half-band filter) is to start with an odd-order half-band filter with specifications  $F_p, F_s, \delta_p$  and  $\delta_s$ , interconnected by the relations [3]

$$\delta_s = \sin(\Delta\varphi_{\max}/2); \delta_p = 1 - \sqrt{1 - \delta_s^2}; F_p = 0.5 - F_s; \quad (2)$$

and with a TF  $G(z)$  that may be represented as a sum of two allpass TFs [3] [5]

$$G(z) = 0.5[A_1(z^2) + z^{-1}A_2(z^2)]. \quad (3)$$

An "even-odd" decomposition (Fig. 1) and the substitution

$$H(z) = 2G(-jz) \quad (4)$$

must be applied in order to obtain the real allpass TFs. Thus

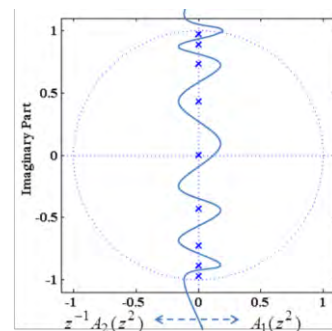


Fig. 1. "Even-odd" decomposition of the TF poles.

$$H_{HT}(z) = 2G(-jz) = [A_1(-z^2) + jz^{-1}A_2(-z^2)] \quad (5)$$

represents the HT as a complex sum of two real allpass functions, whose realization (for real input signal  $x(n)$ ) is given in Fig. 2. Details about the design are given in [3] [5].

<sup>1</sup>The authors are with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria; e-mails: [ksi@tu-sofia.bg](mailto:ksi@tu-sofia.bg), [stoyanov@ieee.org](mailto:stoyanov@ieee.org).

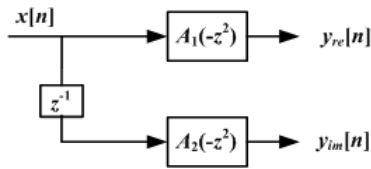


Fig. 2. HT realization.

### III. ALLPASS SECTIONS REALIZATIONS

The allpass TFs in Eq. (3) are having all their poles on the imaginary axes, while those in Eq. (5) are all on the real axes. In order to obtain higher accuracy in the 90° phase shifting in case of a limited word-length environment, the allpass TFs in Fig. 2 could be realized as cascades of special second-order allpass sections. It follows from Fig. 1 that if a cascade realization would be used, as the possible real pole positions are scattered all around the real axes, the allpass sections with low sensitivities for all these positions will be needed.

We have studied [6] all known (about 20) first order sections and it was found that several low-sensitivity sections for every real pole position could be found. We select to use the most typical four of them, namely the *ST1* section, providing low-sensitivity for poles near  $z=1$ , *MH1* and *SC*, having low sensitivity for poles near  $z=0$  and *SV* section for poles near  $z=-1$ . The special sections are obtained from these real first order sections by changing the signs of the coefficients of the allpass TFs in Eq. (3) and by replacing  $z^{-1}$  by  $z^{-2}$  as it is shown in Fig. 3. We denote these new second order allpass sections as *MH1-2*, *ST1-2*, *SV-2* and *SC-2*.

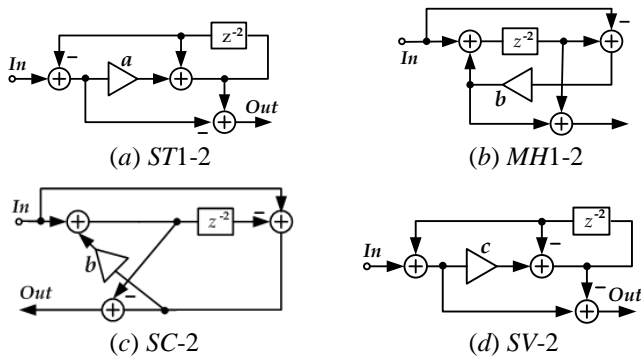


Fig. 3. Different special second-order allpass sections.

Their TFs are:

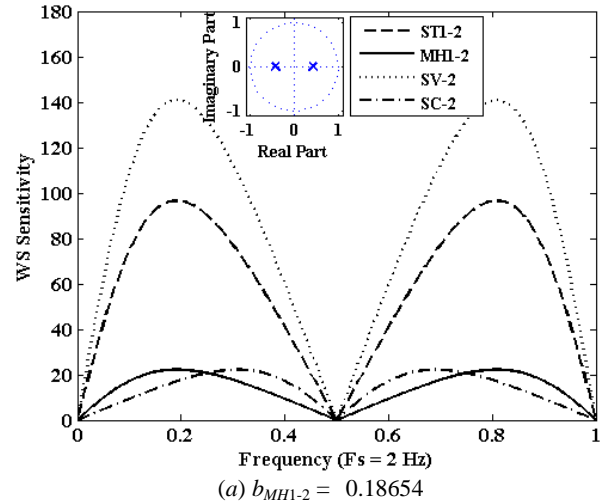
$$H_{ST1-2}(z) = \frac{-(1-a) + z^{-2}}{1 - (1-a)z^{-2}}; \quad H_{MH1-2}(z) = \frac{-b + z^{-2}}{1 - bz^{-2}}; \quad (6)$$

$$H_{SC-2}(z) = \frac{-b - z^{-2}}{1 + bz^{-2}}; \quad H_{SV-2}(z) = \frac{1 - c + z^{-2}}{1 + (1-c)z^{-2}}. \quad (7)$$

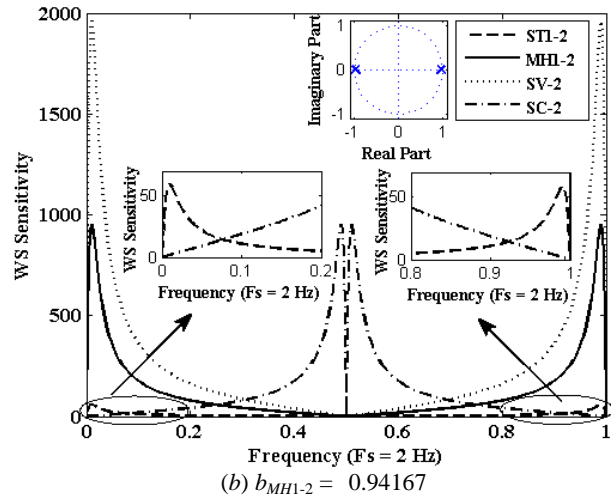
### IV. ALLPASS SECTIONS SENSITIVITY INVESTIGATIONS

In Fig. 4 a, b the worst-case (WS) phase-response sensitivities of the above mentioned four special sections are given for realizations with two different TF pole positions. The sensitivities are obtained by using the package PANDA [7]. By

comparing the results with our previous investigations in [8] [9], it can be noted that the WS sensitivity behavior of the special second order sections is very similar to that of the corresponding first order sections but with the symmetry around the frequency  $f = 0.5$ . It is clearly seen that there exists a proper selection of the sections for every given TF pole position because of the significant difference between the maximal values of the sensitivities (in some cases it can reach more than 100 times especially for the poles near  $\pm 1$ ).



(a)  $b_{MH1-2} = 0.18654$



(b)  $b_{MH1-2} = 0.94167$

Fig. 4. Worst-case phase-sensitivities of second-order allpass sections (Fig. 3) for two different TF poles positions.

### V. OVERALL SENSITIVITY INVESTIGATIONS

In order to estimate how the proper choice of the special sections will affect the behavior of the HT realization in a limited word-length environment, we have designed and investigated a ninth order HT having the TF poles positions given in Fig. 1 (the initial elliptic half-band filter specifications are: passband frequency  $F_p = 0.24$  and stopband attenuation  $\delta_s = 0.01$  ( $R_s = 40$  dB), producing  $\Delta\phi_{max} = 1.15^\circ$ ).

Then, we have designed 4 different HT realizations (Fig. 2). The first one was realized using the standard way (using only *MH1-2* sections) and it is marked in the figures as "4*MH1-2*". The allpass sections selection for the other realizations is based on the sensitivity minimization of the individual sec-

tions depending on their poles positions. Thus, in the second HT realization (denoted with "4 ST1-2") four ST1-2 sections were used. In the third and fourth implementations, two MH1-2 and two ST1-2 sections have been selected. In the first case, we have a special section of each type in every branch of the realization, while in the second case – two MH1-2 sections are used in the upper branch (the real output) and two ST1-2 sections – for the imaginary output. The results for the overall sensitivity of the two branches are shown in Fig. 5.

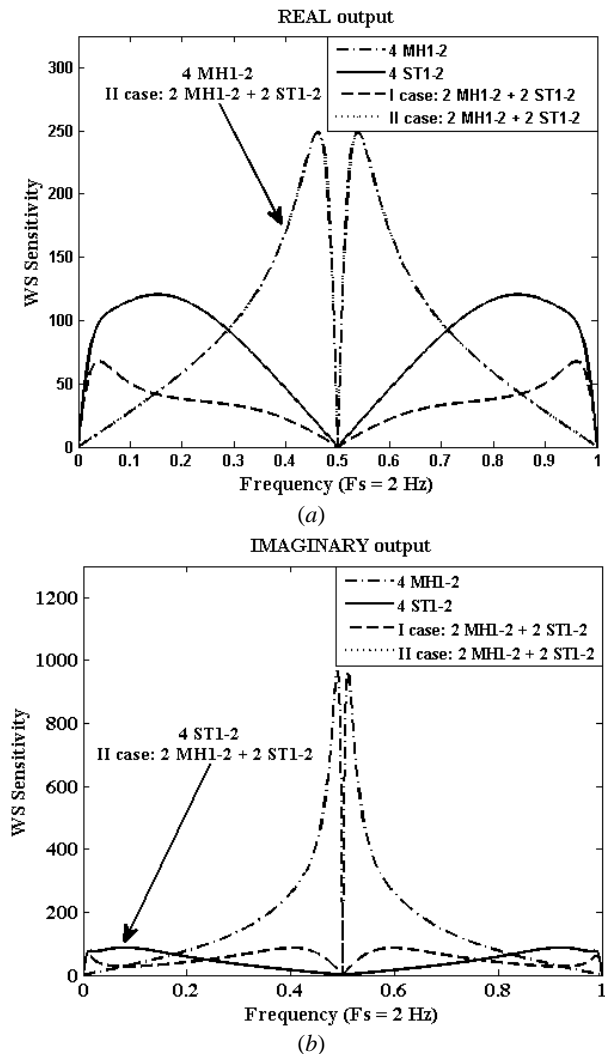


Fig. 5. Worst-case phase-sensitivities of the HT (Fig. 2) realized with different sets of allpass sections (for a 9<sup>th</sup> order HT).

It appeared that the best configuration is with two MH1-2 and two ST1-2 sections, each in every branch (I case), providing the lowest overall sensitivity in both paths.

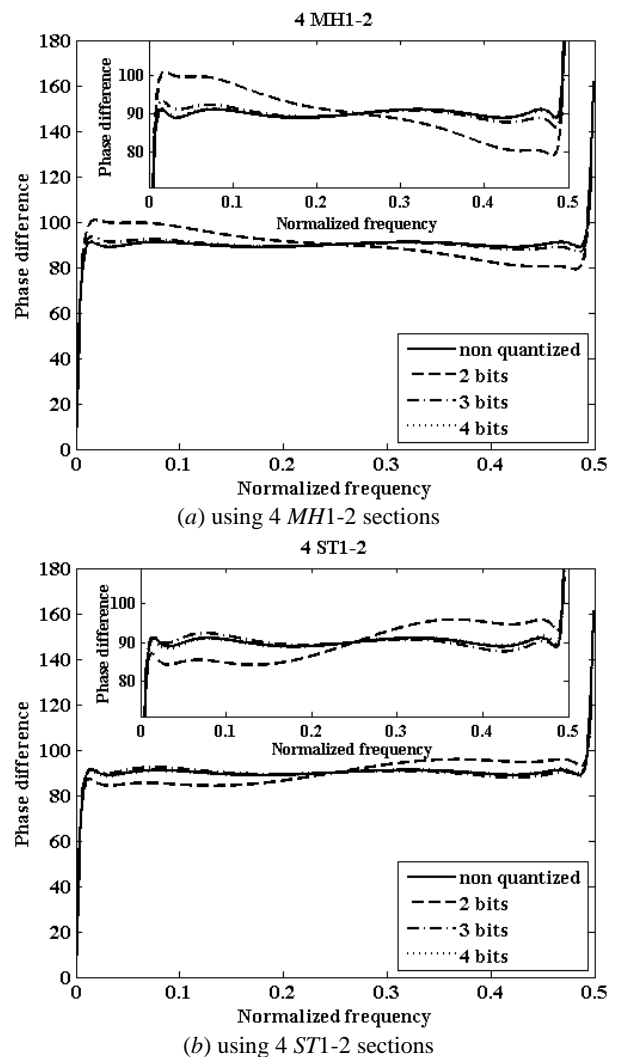
## VI. INVESTIGATION OF THE INFLUENCE OF THE SECTIONS COMBINATIONS IN THE BRANCHES

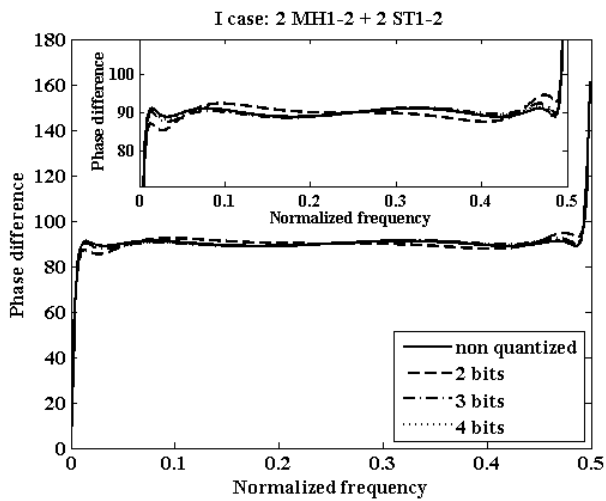
The phase difference between the two outputs in Fig. 2 will not be exactly 90°. Over some frequency range (narrower than half of the sampling frequency) it will alternate around this value with amplitudes  $\Delta\phi_{max}$  depending ideally only on the selected value of  $\delta_s$  Eq. (2), but in reality – also on the design accuracy and on the parasitic effects of the digital realization.

These additional deviations should be kept as lower as possible mainly by reducing the influence of the parasitic effects (by minimizing the sensitivities to the variations of the multiplier coefficients values). It will appear from what follows, that it might not be an easy straightforward procedure.

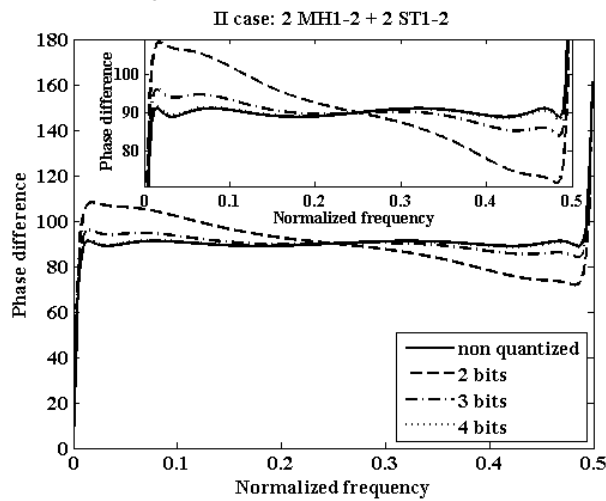
The accuracies of the HT realizations (the phase difference between the two branches) in a limited word-length environment are compared in Fig. 6. Based on the results shown in Fig. 5, it is natural to have a high sensitivity (to small changes in the two branches) of the phase difference between the two outputs in Fig. 2 for 4 MH1-2 HT realization, but the results shown in Fig. 6a are quite surprising, compared to these in Fig. 6b,c,d (with minimized sensitivity). We suppose that this might be an effect due to some internal compensation between the parasitic effects in the branches, explained with the different signs of the sensitivities. The worst-case sensitivity WS, used in our investigations, is not able to reveal these mutual compensations, because it is eliminating the signs of the individual sensitivities.

The highest accuracy, as it is shown in Fig. 6, is achieved when we have two MH1-2 and two ST1-2 sections each in every branch (I case) of the HT. In this case, the selection of the sections and their placement in the branches are made under the above mentioned observations.





(c) using 2 MH1-2 and 2 ST1-2 sections - I case



(d) using 2 MH1-2 and 2 ST1-2 sections - II case

Fig. 6. Word-length dependence of the accuracy of the HT phase difference for realizations with different allpass sections.

As it can be seen after quantization to 2 bit (in CSD code) not only the fluctuations of the phase difference in Fig. 6a,b,d are growing very much above the ideal, but the range of frequencies over which this difference is approximately constant, is sharply reduced, while in Fig. 6c these parameters are practically unchanged.

The main conclusion of these investigations is that besides the sensitivity minimization, an additional step, consisting of a study of all possible combinations of the selected allpass sections within the branches, has to be introduced. A more general solution of this problem will be a derivation of a formula about the sensitivity of the phase quadrature to the changes of the multipliers' values, but it may appear to be a very difficult task.

### VII. LOW-SENSITIVITY DESIGN PROCEDURE

Taking into account all results so obtained, we propose the following design procedure:

1. Obtain  $H_{HT}(z)$  Eq. (5) by applying the standard design procedure from Sect. 2.

2. Decompose the TFs  $A_1(z^{-2})$  and  $A_2(z^{-2})$  to special second-order allpass TFs and find where their poles are situated.
3. Select (from Fig. 3) or develop new allpass sections realizing each couple of poles with the lowest sensitivity and verify this by sensitivity studies as these in Fig. 4.
4. Investigate the overall sensitivities in the two branches of Fig. 2 for all possible combinations of the selected allpass sections realizations in order to select the best set.
5. In case of a very high accuracy design, verify the selection by simulating the structure in a limited word-length environment (as in Fig. 6).

We have applied this procedure for different sets of specifications and it was always possible to find an implementation clearly outperforming all the others as the case in Fig. 6 c.

### VIII. CONCLUSION

A new approach to improve the accuracy of the allpass based Hilbert transformers (realized as two parallel branches) through sensitivity minimizations of each individual special second-order allpass section in the cascade realizations of the two branches was proposed in this paper. The design procedure is simple and straightforward, without iterative and complicated optimization steps and is achieving accuracy of realizations close to the ideal case (nonquantized coefficients). The low sensitivities so attained permit also a very short coefficients word-length, a higher processing speed and lower power consumption.

### REFERENCES

- [1] C. S. Turner, "An efficient analytic signal generator", *Signal Proc. Magazine*, vol. 23, pp. 91-94, July 2009.
- [2] C. B. Rorabaugh, "Notes on Digital signal processing: practical recipes for design, analysis, and implementation", Notes 58-66, Prentice Hall, 2011.
- [3] P. A. Regalia, "Special Filter Designs", Chapter 13 in the book *Handbook for Digital Signal Processing* (Editors S. K. Mitra and J. F. Kaiser), John Wiley & Sons, NY, pp. 909-931, 1993.
- [4] L. Milic, J. Certic, M. Lutovac, "A class of FRM-based allpass digital filters with applications in half-band filters and Hilbert transformers", *Proc. Intern. Conf. on Green Circuits and Syst. (ICGCS)* 2010, Shanghai, China, pp. 273-278, June 2010.
- [5] S. Mitra, *Digital signal processing: A computer based approach*, McGraw-Hill, 2006.
- [6] G. Stoyanov and H. Clausert, "A comparative study of first order digital all-pass structures", *Frequenz*, vol. 48, No 9/10, pp. 221-226, Sept./Oct. 1994.
- [7] H. Sugino and A. Nishihara, "Frequency-domain simulator of digital networks from the structural description", *Trans. of the IEICE of Japan*, vol. E73, No.11, pp. 1804-1806, Nov. 1990.
- [8] G. Stoyanov, Z. Nikolova, K. Ivanova, V. Anzova, "Design and realization of efficient IIR digital filter structures based on sensitivity minimizations", *Proc. 8th IEEE Conf. TELSIS'2007*, Nish, Serbia, vol. 1, pp. 299-308, Sept. 2007.
- [9] G. Stoyanov, K. Nikolova, and M. Kawamata, "Low-sensitivity design of allpass based fractional delay digital filters", Chapter 7 in the book *Digital filters*, F. P. Márquez (Ed.), Intech Publ. House, pp. 155-178, 2011.

# Acoustic Standing Waves in Closed Cylindrical Enclosures

Ekaterinoslav Sirakov<sup>1</sup> and Hristo Zhivomirov<sup>2</sup>

**Abstract** – The work presented in this paper provides a theoretical analysis of acoustic standing waves inside cylindrical enclosures with rigid walls. Mathematical relationships are given for the calculation and researching of modal frequencies and standing sound waves in a cylindrical box. The results from the calculating and measuring of modal frequencies and the box response are shown graphically and in a table.

**Keywords** – Acoustic standing waves, Closed cylindrical enclosures.

## I. ACOUSTIC STANDING WAVES

In the paper are discussed the acoustic processes in a closed cylindrical volume (Fig. 1). As a result of multiple reflections of the sound waves from the walls of the volume three-dimensional sound field arises [1], an example of which is given in Fig. 2. Depending on the shape, dimensions and their ratios in the enclosed volume fluctuations occur with a different set of natural frequencies [2, 3].

The acoustic processes in a closed cylindrical volume can be represented by the wave equation in cylindrical coordinate system  $(r, \phi, z)$  [4]:

$$\nabla^2 p = c \cdot \left[ \frac{1}{r} \frac{\partial}{\partial r} \left( r \frac{\partial p}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2 p}{\partial \phi^2} + \frac{\partial^2 p}{\partial z^2} \right] \quad (1)$$

where:  $\phi$  - azimuth angle of the source.

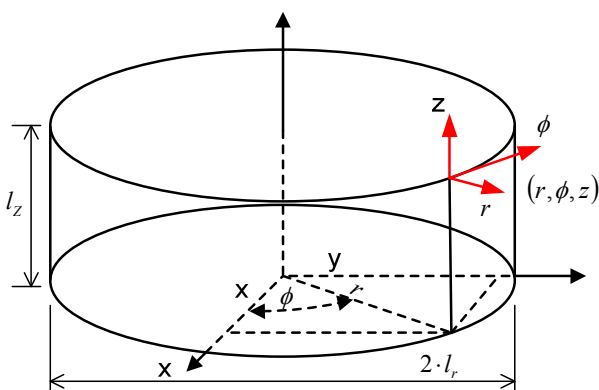


Fig. 1. Cylindrical acoustic volume [5]

<sup>1</sup>Ekaterinoslav Sirakov is with the Department of Communication Engineering and Technologies, Faculty of Electronics, Technical University-Varna, Studentska Street 1, Varna 9010, Bulgaria, E-mail: katosirakov@abv.bg

<sup>2</sup>Hristo Zhivomirov is a Ph.D. student with the Department of Communication Engineering and Technologies, Faculty of Electronics, Technical University-Varna, Studentska Street 1, Varna 9010, Bulgaria, E-mail: hristo\_car@abv.bg.

The solution of the wave equation is [6, 7]:

$$p = \begin{cases} \cos(j \cdot \theta) \\ \text{or} \\ \sin(j \cdot \theta) \end{cases} \cdot \cos\left(\frac{n_z \cdot \pi \cdot x}{l_z}\right) \cdot J_j\left(\frac{\lambda_{j,k} \cdot r}{l_r}\right) \quad (2)$$

$n_z, j, k = 0, 1, 2, \dots$

where:  $J_j$  - Bessel function.

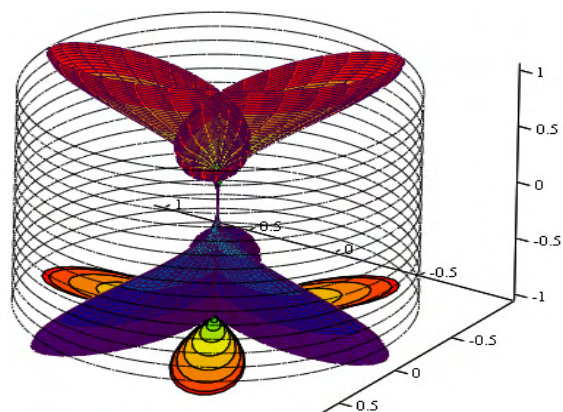


Fig. 2. Distribution of the magnitude of sound pressure in a cylindrical box:  $j$  axial sound wave,  $f=3.538$  kHz,  $j = 3, k = 0, \lambda_{j,k} = 4.201189, \cos(j \cdot \theta)$ , mode  $(0, 3, 0)$

The natural frequencies for the corresponding values of  $n_z, j$  and  $k$  can be found by [5]:

$$f = \frac{c}{2} \cdot \sqrt{\left(\frac{n_z}{l_z}\right)^2 + \left(\frac{\lambda_{j,k}}{\pi \cdot l_r}\right)^2}, \text{ Hz} \quad (3)$$

$n_z, j, k = 0, 1, 2, \dots$

The natural frequency of the cylindrical box, calculated in accordance with mathematical dependence (3) is presented in Fig. 3 and Table I.

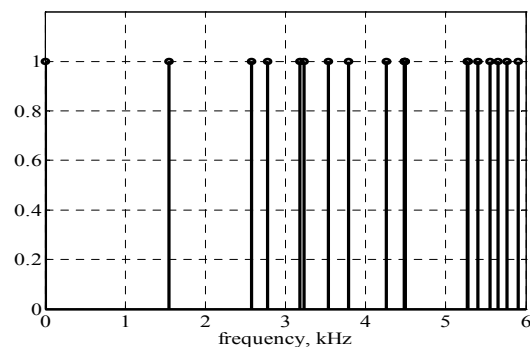


Fig. 3. Plot of mode distribution

TABLE I THE EIGHTEEN LOWEST NORMAL MODES AND THEIR NATURAL FREQUENCIES FOR A CILINDRICAL BOX WITH RIGID WALLS.

№	mode	$n_z, j, k$	frequency, kHz
1	j axial	0, 1, 0	1.551
2	j axial	0, 2, 0	2.573
3	z axial	1, 0, 0	2.774
4	z, j tangential	1, 1, 0	3.178
5	k axial	0, 0, 1	3.227
6	j axial	0, 3, 0	3.538
7	z, j tangential	1, 2, 0	3.784
8	z, k tangential	1, 0, 1	4.256
9	j axial	0, 4, 0	4.479
10	j, k tangential	0, 1, 1	4.491
11	z, j tangential	1, 3, 0	4.497
12	z, j tangential	1, 4, 0	5.268
13	z, j, k oblique	1, 1, 1	5.278
14	j axial	0, 5, 0	5.404
15	z axial	2, 0, 0	5.548
16	j, k tangential	0, 2, 1	5.648
17	z, j tangential	2, 1, 0	5.761
18	k axial	0, 0, 2	5.909

## II. ENCLOSURE RESPONSE MEASUREMENTS IN MODEL CLOSED CYLINDRICAL BOX

The measured characteristics of the sound pressure in cylindrical enclosure with the application software Realtime Analyzer [8] are presented in graphical form in Fig. 4.

Measurements were made in a cylindrical box with dimensions: height 6.2 cm, diameter 13 cm and wall thickness 0.1 cm. The program allows the data from the measured values of sound pressure in dB to be stored in tabular and text format for further analysis.

To examine the modal structure of the enclosure box response at the center of the volume was measured.

## III. CONCLUSION

In the cylindrical acoustic volumes as with rectangular [2] "axial" and "tangential" natural frequencies can be defined. The  $z$ -axial natural frequency for  $j = k = 0$  is given in Table I, № 3 and № 15. When  $n_z = 0$  by analogy with the rectangular speaker enclosure  $z, \phi$ -tangential natural frequencies can be defined – Table I, № 10 and № 16. If  $n_z = 0$  and  $j = 0$  the sound is focused along the axis of the cylinder, the sound wave propagates radially and the natural frequencies are  $r$ -axial – Table I, № 5 and № 18. When  $n_z = 0$  and  $k = 0$  the natural frequencies can be called  $\phi$ -axial [4] (perpendicular to  $z$  and  $r$ ) as shown in Table I, № 1, 6, 9 and № 14.

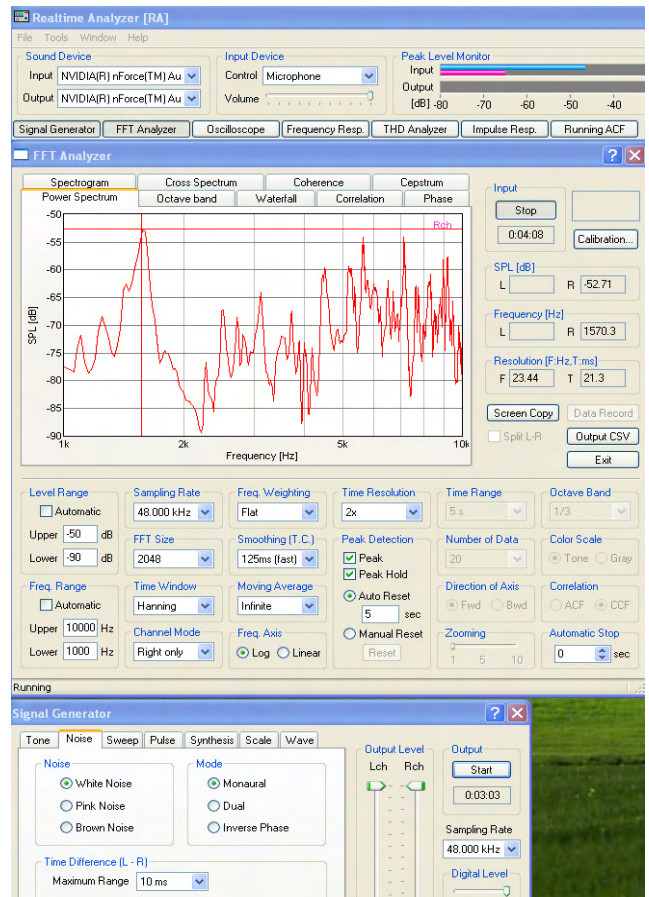


Fig. 4. The measured sound pressure in the closed cylindrical enclosure (dimensions: height 6.2 cm, diameter 13 cm and wall thickness 0.1 cm)

## REFERENCES

- [1] Ekaterinoslav Sirakov, Hristo Zhivomirov, Boris Nikolov, "Green's Function and Acoustic Standing Waves in Rectangular Loudspeaker Enclosures", ICEST 2011, Proceedings of Papers, pp. 721-724, Nis, Serbia, 2011.
- [2] Екатеринаслав С. Сираков, "Собствени резонансни честоти на правоъгълно озвучително тяло", Национална конференция с международно участие „Акустика 1“, Варна, 2009, Списание „Акустика“, год. XI, бр.11, 2009 г., стр. 110-118.
- [3] Екатеринаслав Сираков, Борис Николов, Любомир Камбуров, "Влиянието на съотношението на размерите на правоъгълна акустичен обем върху разпределението на собствените честоти", Национална конференция с международно участие „Акустика 1“, Варна, 2010, Списание „Акустика“, год. XII, бр.12, 2010 г., стр. 94-103.
- [4] Philip M. Morse, *Vibration and Sound*, New York, McGraw-Hill, 1936.
- [5] Екатеринаслав Сираков, "Акустичен модел на цилиндрични затворени обеми", Национална конференция с международно участие „Акустика 1“, Варна, 2011.
- [6] Frank Fahy, *Foundations of Engineering Acoustics*, Academic Press, San Diego, 2005.
- [7] István L. Vér, Leo L. Beranek, *Noise and Vibration Control Engineering: Principles and Applications*, Wiley, 2006.
- [8] <http://www.ymec.com/products/dssf3e/>



# Control of Radiation Directivity Applying Independent Element Dodecahedral Loudspeaker

Marko Jelenković<sup>1</sup>, Dejan Ćirić<sup>2</sup>, Jelena Zdravković<sup>3</sup> and Stefan Tomić<sup>4</sup>

**Abstract** – Control and synthesis of sound source directivity have a significant position in modern acoustics. For this purpose, special sound sources that consist of independent element loudspeaker array are used. Sound source analyzed in this paper is a loudspeaker array in the form of dodecahedron. To generate different directivities, the loudspeaker array is fed independently with a few characteristic sets of signals. Special setup was established for radiation directivity measurements. For analyzing and presenting the results, a MATLAB application is developed. The measured directivities obtained by feeding of some of the loudspeakers with particular signals show interesting patterns of radiation.

**Keywords** – Sound source, loudspeaker array, dodecahedron, radiation directivity, swept sine technique.

## I. INTRODUCTION

Radiation directivity pattern is one of the most important properties of a real sound source [1], [2]. It reflects source's interaction with the environment and therefore it is a significant parameter in various acoustics researches. A sound source can generally be characterized by three properties: timbre related to spectral and temporal attributes, intensity, and directivity providing spatial information on the sound radiated from the source [3]. Today's loudspeaker systems can faithfully reproduce the sound tone (timbre) and intensity, but they have typical directivities that are significantly different from the directivities of natural sound sources [4]. Therefore, there is a need for simulating (synthesizing) directivity characteristics of the real sound sources [5]. This can be realized using the three-dimensional sound sources, such as dodecahedral loudspeaker [6-9].

Another problem with modern sound sources comes from the requirement for omni-directivity. Sources are typically not omni-directional in the whole frequency range, but only at lower frequencies [4], [5]. That gives importance to researches dealing with radiation directivity control.

In this paper, a sound source based on a loudspeaker array

<sup>1</sup>Marko Jelenković is a student at the Faculty of Electronic Engineering at University of Niš, Aleksandra Medvedeva 14, Niš 18000, Serbia, E-mail: [virtus@elfak.rs](mailto:virtus@elfak.rs).

<sup>2</sup>Dejan Ćirić is with the Faculty of Electronic Engineering at University of Niš, Aleksandra Medvedeva 14, Niš 18000, Serbia, E-mail: [dejan.ciric@elfak.ni.ac.rs](mailto:dejan.ciric@elfak.ni.ac.rs).

<sup>3</sup>Jelena Zdravković is a student at the Faculty of Electronic Engineering at University of Niš, Aleksandra Medvedeva 14, Niš 18000, Serbia, E-mail: [jelena.z@elfak.rs](mailto:jelena.z@elfak.rs).

<sup>4</sup>Stefan Tomić is a student at the Faculty of Electronic Engineering at University of Niš, Aleksandra Medvedeva 14, Niš 18000, Serbia, E-mail: [stefan@elfak.rs](mailto:stefan@elfak.rs).

in dodecahedron form is used for radiation directivity control. Method for analyzing the radiation directivity is partially developed in earlier researches [4], [5], but it is modified here in order to enable using all twelve channels independently. Special attention is paid on establishment of unique, easy to use, measurement system, and synthesis of basic radiation directivity patterns.

## II. MEASUREMENT SYSTEM

The computer-based measurement system was established especially for the purpose of radiation directivity measurement. It consists of a computer (laptop), M-Audio ProFire 610 audio interface (sound card), M-Audio Fast Track Pro audio interface, twelve-channel audio amplifier, spherical sound source with independent elements in the form of dodecahedron (Fig. 1), condenser measuring microphone with preamplifier, and power supply.



Fig. 1. Sound source with independent elements (loudspeakers) in the form of a dodecahedron

Since the dodecahedral loudspeaker array consists of twelve loudspeakers, independent signals were fed via two audio interfaces to every of these twelve channels. Eight channels of M-Audio ProFire and four of M-Audio Fast Track Pro provide the needed twelve channels, which could be controlled independently. The diagram of the measurement system is shown in Fig. 2. Due to the different characteristics of the interfaces, setting up of the system was more complicated, and a latency problem was inevitable. The problems were resolved by applying special calibration method that includes measuring of impulse responses for both sound cards independently and determination of latency difference

between two impulse responses. Based on this information, silence of appropriate duration is automatically added before the measurement signal with smaller latency. In this way, signals from both sound cards have the same latency.

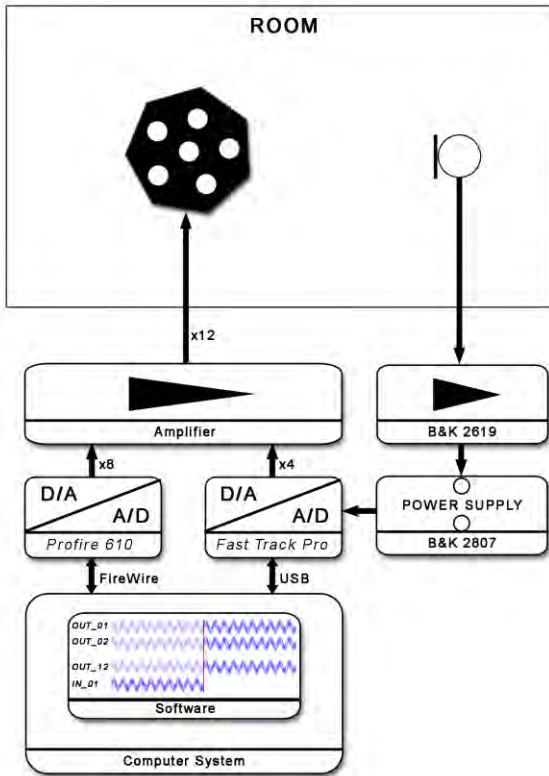


Fig. 2. Measurement system

Also, the measurement system has basic signal control options implemented in every channel, such as amplitude regulation and phase inversion. This provides a lot of combinations that could be used for achieving the wanted characteristics of radiation.

### III. RADIATION DIRECTIVITY MEASUREMENT

The measurements were carried out in pseudo anechoic conditions. Distance to the first obstacles was greater than 1.5 meters, so difference between direct sound and first reflection was about 10 ms, which allows us to identify and cut off all reflections, and analyze only direct sound. Microphone and sound source were placed on the same height and the distance between them was 1 meter.

Measurement process is based on swept sine technique [10], [11]. In our earlier research, it has been shown that this technique is more reliable than some other techniques (like MLS technique) [12], [13]. The excitation signal used in this technique is so called swept sine or sweep. It represents a sine signal with frequency varying in time, e.g. linearly (linear sweep) or exponentially (logarithmic sweep). Swept sine technique is considered to be immune to some disturbances such as time variance or non-linearities [10], [11]. Also, one of its advantages is that precise synchronization between the

clock applied for the test signal reproduction and recording of the response during the measurement is not required [12]. In the present research, an excitation swept sine signal of duration of about 24 seconds with frequency range from 10 Hz to 22 kHz sampled at 44.1 kHz is used.

Three excitation combinations are utilized. For the first one, six loudspeakers from one half of the sphere are fed with the same sweep signal, and the other six with opposite phase of the same signal. For the second and the third one, two opposite loudspeakers on the sphere are fed with corresponding signals. In one of the configurations, loudspeakers are fed with the same signals in phase and in the other one with same signals, but in opposite phases.

The radiation directivity is measured only in horizontal plane with resolution of  $10^\circ$ , so there are 36 impulse responses for each excitation combination. Measurements were performed in a way that the microphone was fixed, and sound source was rotated around its horizontal axis. Processing of the extracted part of the impulse response (free of reflections) is performed in the frequency domain, which is based on FFT and determination of impulse response spectrum. After determination of spectrum, next step is determination of direction of maximal radiation (reference axis), which is used for normalization and presentation of directivity patterns. Using this kind of processing, radiation pattern can be obtained for many different frequencies. Actually, the maximum number of frequencies (the frequency resolution) depends on the length of the extracted impulse response containing only direct sound. Summary of impulse response processing is given in Fig. 3.

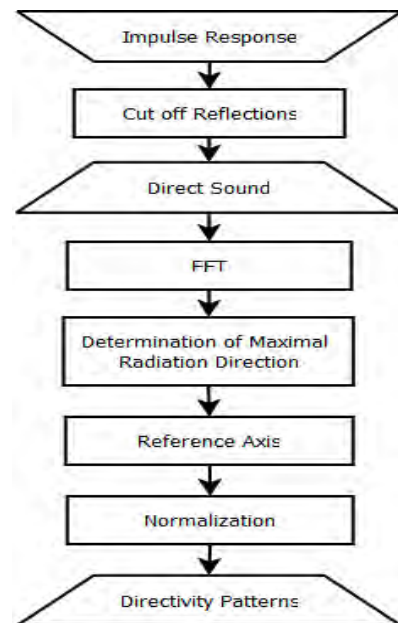


Fig. 3. Impulse response processing and determination of directivity radiation pattern

It should be kept in mind that processing of extracted parts of impulse responses, in order to obtain radiation directivity patterns, may also be performed in the time domain, as an alternative to processing in the frequency domain.

IV. RESULTS

For aforementioned configurations, the radiation directivities are obtained using the described procedures of measurement and processing in the frequency domain. Configuration of six independent elements from one half of dodecahedral loudspeaker array fed with same signal in phase, and the other six with opposite phase signal leads to the expected radiation patterns at lower frequencies, which tend to be close to bi-directional (Fig. 4(a)). By increasing frequency, the patterns begin to lose primary form, and in this case, at higher frequencies they get a flower shape (Fig. 4(b)).

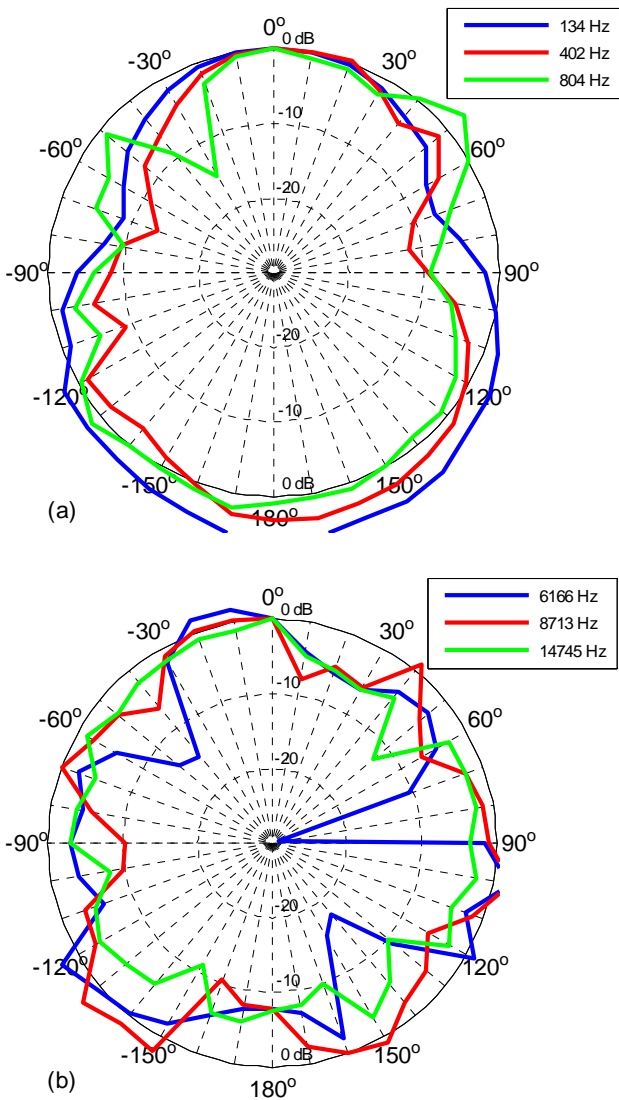


Fig. 4. Radiation directivity of dodecahedral loudspeaker array with six elements in phase and six elements in opposite phase at lower (a) and higher (b) frequencies

In the second configuration case, with two opposite elements fed with the same signal in phase, the obtained radiation characteristics are omni-directional at lower frequencies, while they get shape of narrowed figure-of-eight pattern at higher frequencies (Fig. 5.). For the last

measurement configuration, where elements (loudspeakers) on opposite sides of dodecahedron are fed with opposite phase signals, the radiation patterns at both lower and higher frequencies have the expected bi-directional shape (figure-of-eight), see Fig. 6. From the presented results, it can be seen that the pattern is somewhat wider at lower frequencies (Fig. 6(a)), while it is somewhat narrower at higher frequencies (Fig. 6(b)).

It should be noted that for some directions, curves presenting radiation directivity come out of the range of 0 dB. This occurs because chosen reference axis is not the axis that has absolute maximum of radiation at all frequencies. Hence, regardless the criterion of referent axis choice, there is no uniform solution that yields the direction of absolute maximum at all frequencies. Due to that, the described disturbance in presenting the directivity radiation is inevitable. In all presented figures, the reference axis is positioned in the direction of 0°.

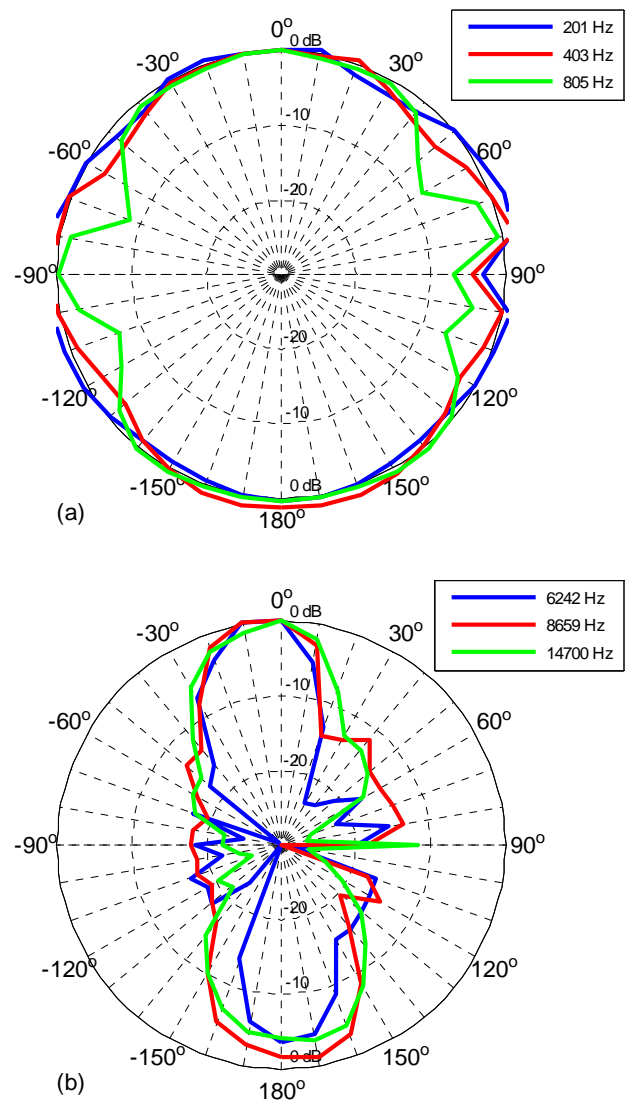


Fig. 5. Radiation directivity of dodecahedral loudspeaker array with two opposite elements in phase at lower (a) and higher (b) frequencies

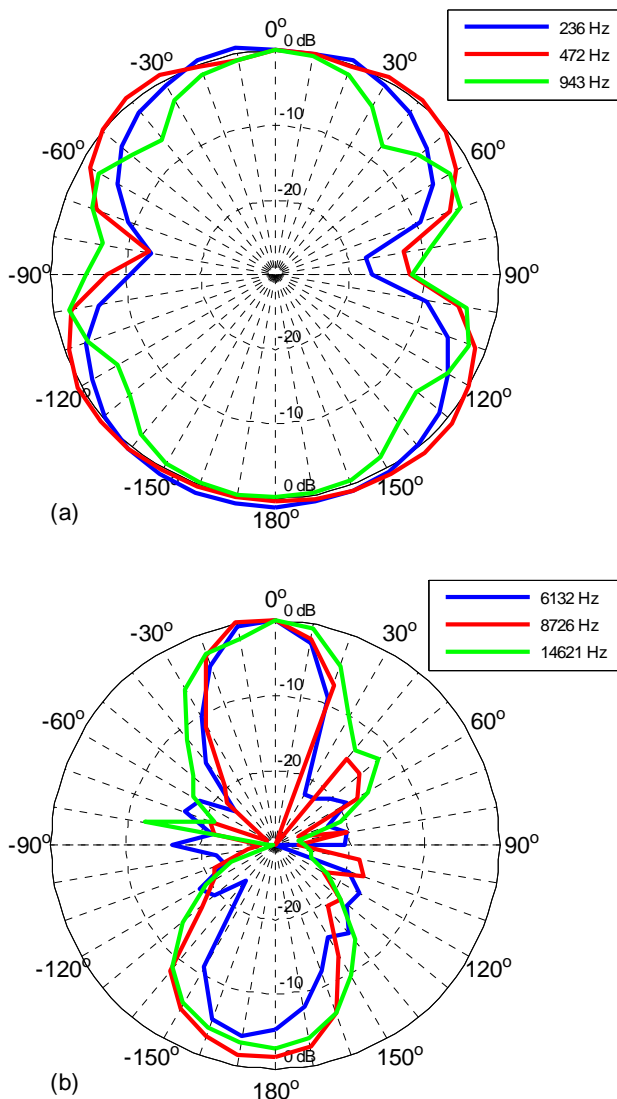


Fig. 6. Radiation directivity of dodecahedral loudspeaker array with two opposite elements, one in phase and one in opposite phase on lower (a) and higher (b) frequencies

V. CONCLUSION

In this paper, radiation directivity of spherical sound source with independent elements in the form of dodecahedron is analyzed. The presented results and up to now experience show that using develop procedure of measurement, radiation directivity characteristics of sound source could be obtained in an adequate way. Measurement system along with the developed software support and measuring and processing procedures are well established and provides sufficient flexibility and control.

For the observed configuration cases, applying various combinations of excitations and disposition of elements (loudspeakers), some specific radiation directivity patterns can be achieved. Most of these patterns have the expected forms at lower frequencies - one omni-directional, and two

approximately bi-directional patterns. At higher frequencies, their shape is changed and converges to the particular form.

The developed system for control of radiation directivity together with the system for measurement of this directivity will be involved in further research related to synthesis of some specific radiation directivity patterns similar to those of real sound sources.

ACKNOWLEDGEMENT

Presented results are obtained within the scope of the project no. 36026 financed by Ministry of Science and Technological Development of Republic of Serbia.

REFERENCES

- [1] J. H Rindel, F. Otondo, "The interaction between room and musical instruments studied by multi-channel auralization," in Proc. of Forum Acusticum, Budapest, Hungary, 2005.
- [2] K. Brian FG, D'A. Christophe, "Directivity measurements of singing voice," in Proc. 19th International Congress on Acoustic, Madrid, Spain, 2007.
- [3] O. Warusfel and N. Misdariis, "Directivity synthesis with a 3D array of loudspeakers application for stage performance," in Proc. Conf. Digital Audio Effects (DAFX-01), Limerick, Ireland, 2001.
- [4] B. Stojić, D. Ćirić, M. Marković, "Radiation directivity of dodecahedral sound source with independent elements", (in Serbian), in Proc. LIV ETRAN, paper AK3.5, Donji Milanovac, Serbia, 2010.
- [5] D. Ćirić, "Control of radiation directivity of multi-loudspeaker sound source", (in Serbian), in Proc. LIII ETRAN, paper AK3.3, Vrnjačka Banja, Serbia, 2009.
- [6] B. Stojić, D. Ćirić, M. Marković, "Radiation directivity of spherical sound source with independent element dodecahedral loudspeaker array", in Proc. 1st EAA-EuroRegion, paper S20-11, Ljubljana, Slovenia, 2010.
- [7] P. Kassakian, D. Wessel, "Caracterization of spherical loudspeaker arrays," presented at 11 th AES Convention, San Francisco, USA, 2004.
- [8] N. Misdariis, O. Warusfel, R. Causse, "Radiation control on a multi-loudspeaker device," in Proc. International Symposium on Musical Acoustic 2001, Perugia, Italy, 2001.
- [9] N. Misdariis, F. Nicolas, O. Warusfel, R. Causse, "Radiation control on multi-loudspeaker device: La Timée," in Proc. International Computer Music Conference, Havana, Cuba, 2001.
- [10] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," presented at 108th Convention Audio Eng. Soc., abstract in J. Audio Eng. Soc., vol. 48, no. 4, p. 350, 2000.
- [11] S. Müller and P. Massarani, "Transfer-Function Measurement with Sweeps," J. Audio Eng. Soc., vol. 49, no. 6, pp. 443-471, 2001.
- [12] M. Ličanin, A. Đorđević, M. Jelenković, Mentor: D. Ćirić, "Sensitivity of impulse response measurements with maximum length sequences and sweeps", in Proc XLVI ICEST, pp. 567-570, Niš, Serbia, 2011.
- [13] M. Jelenković, D. Ćirić, M. Ličanin, A. Đorđević, "Repeatability of room impulse response measurements", (in Serbian), in Proc. LV ETRAN, paper AK2.7, Banja Vrućica, Bosnia and Herzegovina, 2011.

# Modulated bandpass Farrow Decimators and Interpolators

Djordje Babic<sup>1</sup> and Vesa Lehtinen<sup>2</sup>

**Abstract** – The Farrow structure provides a flexible way for adjustable filtering, including adjustable fractional-delay filters and sample rate conversion (SRC) by arbitrary, non-integer, factors. Recent publications have pointed out its suitability to bandpass SRC. When the passband centre is located above the input sample rate, the interpolator complexity becomes approximately proportional to the centre frequency. We propose a new construct, the modulated Farrow structure, which allows arbitrarily high centre frequencies without increasing the polynomial degree of the prototype Farrow filter. The modulating function is constructed as a low-order polynomial in order to avoid costly generation of trigonometric functions.

**Keywords** – Digital Filters, Farrow Structure, Decimators, Interpolators, Band-pass Filtering.

## I. INTRODUCTION

Discrete-time filter structures are used in many applications to interpolate new sample values at arbitrary points between the discrete-time input samples [1], [2]. In those cases, it is beneficial to use polynomial-based interpolation filters. The advantage of the above system lies in the fact that the actual implementation can be efficiently performed by using the Farrow structure [1] or its modifications [2], [3]. In [2], [3], several design methods for polynomial-based interpolation filters have been developed in the time and in frequency domain. Using the frequency domain approach, it is possible to design the polynomial-based filter realized in the form of Farrow structure with an arbitrary baseband (zero center frequency) frequency response.

However, there are some applications where there is a need for a bandpass interpolation. Some example applications are: bandpass sampling, bandpass sampling rate conversion, rational filter banks, etc [4], [5]. It has been shown in [6], that if the center frequency of the bandpass interpolation filter is smaller than the half of the sample rate, for the given passband and stop-band requirements, the filter order does not increase compared to the baseband filter with the same requirements. The design methods presented in [2] and [3] can be effectively used for the design of bandpass polynomial interpolation filters with minor modifications [6].

However, when the center frequency is (far) higher than the input sample rate, the polynomial degree would become approximately proportional to the centre frequency. Therefore, we propose a new construct, the modulated Farrow structure, which allows arbitrarily high centre frequencies

without increasing the polynomial degree of the lowpass prototype Farrow filter. The modulating function is constructed as a low-order polynomial in order to avoid costly generation of trigonometric functions.

## II. POLYNOMIAL MODULATION

Bandpass filters can be implemented with lowpass filters as building blocks by applying frequency translation to the signal before and after filtering which is equivalent to multiplying the impulse response by cosine whose frequency is equal to the center frequency of the bandpass filter [4]. While this permits arbitrary and variable centre frequencies, generation of costly cosine function is needed, this increases the implementation complexity. The same principle can be applied to the case of polynomial-based filter, as well.

In order to minimize the implementation complexity, it may be beneficial to approximate the modulating function with a piecewise polynomial. The modulating polynomial is chosen first, and the modulation is taken into account in the optimized subfilter coefficients. The non-idealities of the modulating function can be compensated by the filter. This eliminates the need for costly (co)sine generation.

Using Taylor polynomials a function can be approximated as closely as desired by a polynomial provided that the function possesses sufficient number of derivatives [7], [8]. Based on the definition, the modulating cosine function  $\cos(\omega_0 t)$  can be approximated with following polynomial:

$$\cos(\omega_0 t) = \sum_{j=0}^{j=J_a} (-1)^j \left[ \cos(\omega_0 a) \frac{(\omega_0 t)^{2j}}{(2j)!} - \sin(\omega_0 a) \frac{(\omega_0 t)^{2(j+1)}}{(2j+1)!} \right]. \quad (1)$$

The above equation can be expressed in more suitable way for practical realization:

$$\cos(\omega_0 t) = \sum_{m=0}^{M_a} g_m t^m, \quad (2)$$

where

$$g_m = \begin{cases} (-1)^{m/2} \cos(\omega_0 a) \frac{(\omega_0)^m}{(m)!} & \text{for } m \text{ even, and} \\ -(-1)^{(m-1)/2} \sin(\omega_0 a) \frac{(\omega_0)^m}{(m)!} & \text{for } m \text{ odd,} \end{cases} \quad (3)$$

$M_a=2*J_a+1$  is a polynomial degree which is related to the approximation error, and  $t \in [0, T)$ . Here,  $T$  is the input sampling interval  $T_{in}$  in the interpolation case, and output sampling interval  $T_{out}$  in decimation case. Thus, we have obtained a polynomial that approximates the cosine function over one polynomial segment. It is possible to generalize the approximation of the cosine over a longer interval by making it a piecewise having  $N$  polynomial segments, and in each segment the order of the polynomial is  $M_a$ . The piecewise

<sup>1</sup> Djordje Babic is with the School of Computing at Union University, Belgrade, Serbia, E-mail: djbabac@raf.edu.rs.

<sup>2</sup> Vesa Lehtinen is with the Department of Communications Engineering at Tampere University of Technology, Tampere, Finland, E-mail: vesa.lehtinen.tut.fi.

polynomial approximation of the cosine can be written as:

$$\cos(\omega_0 t) = \sum_{n=0}^N \sum_{m=0}^{M_a} b_m(n) t^m, \quad (4)$$

where

$$b_m = \begin{cases} (-1)^{m/2} \cos(\omega_0 a(n)) \frac{(\omega_0)^m}{(m)!} & \text{for } m \text{ even, and} \\ -(-1)^{(m-1)/2} \sin(\omega_0 a(n)) \frac{(\omega_0)^m}{(m)!} & \text{for } m \text{ odd.} \end{cases} \quad (5)$$

Here,  $a(n)$  determines the initial phase of the approximated cosine function in each polynomial segment. The value of  $a(n)$  can be selected in such a manner that overall approximated cosine function is symmetric around  $N/2$ ,  $a(n)=n+0.5$ , for  $n=0, 1 \dots N-1$ . In this way, the impulse response of modulated lowpass prototype filter is also symmetric. In the actual implementation, the symmetry can be exploited in a similar way as for polynomial-based filters, the polynomial in  $t$  can be transformed to polynomial in  $(2t-1)$  thus we obtain

$$\cos(\omega_0 t) = \sum_{n=0}^N \sum_{m=0}^{M_a} b_m(n) (2t-1)^m = \sum_{n=0}^N \sum_{m=0}^{M_a} b_m(n) f_m(n, T, t). \quad (6)$$

The modulated impulse response of the polynomial based filter is obtained by multiplying the impulse response of the lowpass prototype filter by cosine:

$$h_m(t) = h_a(t) \cos(\omega_0 t), \quad (7)$$

yielding to:

$$h_m(t) = \sum_{n=0}^N \left[ \sum_{m=0}^M c_m(n) f_m(n, T, t) \sum_{m_1=0}^{M_a} b_{m_1}(n) f_{m_1}(n, T, t) \right]. \quad (8)$$

The modulated impulse response is a piecewise polynomial with the same number of polynomial segments as the lowpass prototype filter and polynomial order of  $M+M_a$  in each segment. The polynomial order increases as the central frequency of the passband increases. Furthermore, the obtained bandpass filter is not configurable, as the new set of coefficients is calculated for different value of passband central frequency.

If we calculate the output sample according to hybrid analogue/digital model, the  $l$ th output sample can be expressed, after some manipulations, in the following form [2], [3]:

$$y(l) = \sum_{n=0}^{N-1} \left[ \sum_{m=0}^M x(n_l - n + \frac{N}{2}) c_m(n) (2\mu_l - 1)^m \sum_{m_1=0}^{M_a} b_{m_1}(n) (2\mu_l - 1)^{m_1} \right] \quad (9)$$

where

$$n_l = \lfloor lT_{out}/T_{in} \rfloor \text{ and } \mu_l = lT_{out}/T_{in} - n_l. \quad (10)$$

Alternatively, the  $l$ th output sample can be expressed as:

$$y(l) = \sum_{m=0}^M \sum_{n=0}^N \eta_l(n) x(n_l - n + N/2) c_m(n) (2\mu_l - 1)^m, \quad (11)$$

with

$$\eta_l(n) = \sum_{m=0}^{M_a} b_m(n) (2\mu_l - 1)^m. \quad (12)$$

Here  $\eta_l(n)$  is a modulating term. There are altogether  $N$  modulating terms  $\eta_l(n)$ , one for each polynomial segment. Equations (12) and (13) can be used to define a bandpass modulated modified Farrow structure, which is shown in Fig. 1. The structure can be used for different passband central frequencies, with a single prototype filter.

### III. SPECIAL CASES OF MODULATED FARROW STRUCTURE

There are several special cases of polynomial modulation, in which the process of building the bandpass Farrow structure can be further simplified. These special cases are derived from mutual relation between the desired central frequency  $f_0 = \omega_0/2\pi$  and the sample frequency of the bandpass filter  $F_s = 1/T$ . If the central frequency is an integral multiple of the sample frequency, integral multiple of the half or quarter of the sample frequency then the modulating term  $\eta_l(n)$  has special values as shown in a sequel.

The most straightforward case of modulated Farrow structures is that with the modulating frequency of  $kF_s$ . All that is needed is to multiply the output of the lowpass Farrow structure with a cosine at  $kF_s$ , i.e., when the modulation frequency is an integral multiple of  $F_s$ , the cosine generation is simplified by:

$$b_m = \begin{cases} (-1)^{m/2} (-1)^k \frac{(\omega_0)^m}{(m)!} & \text{for } m \text{ even, and} \\ 0 & \text{for } m \text{ odd.} \end{cases} \quad (13)$$

Therefore,  $b_m$  depends only on polynomial order  $m$ , and integral multiple  $k$ , reducing (12) to

$$\eta_l = \sum_{m=0}^{M_a} b_m (2\mu_l - 1)^m. \quad (14)$$

Fig. 2. depicts the modulated Farrow structure for modulation frequencies  $kF_s$ . Impulse response modulation at  $kF_s$  is attained simply by implementing modulation of Farrow structure by  $\eta_l$ .

When the modulation frequency is an odd multiple of  $F_s/2$ , the modulating function has the same shape in every polynomial segment, but its sign alternates between adjacent segments. The calculation of the modulating term reduces to

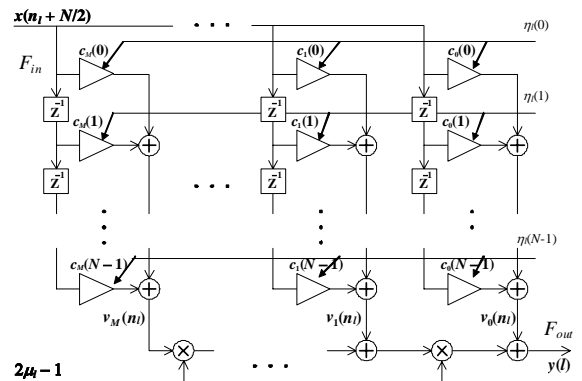


Fig. 1. The modulated modified Farrow structure

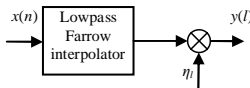


Fig. 2. The modulated Farrow structure for center frequencies  $kF_s$ .

$$b_m(n) = \begin{cases} 0 & \text{for } m \text{ even, and} \\ -(-1)^{(m-1)/2} (-1)^n (-1)^k \frac{(\omega_0)^m}{(m)!} & \text{for } m \text{ odd} \end{cases} \quad (15)$$

and, thus

$$\eta_l(n) = (-1)^n \sum_{m=0}^{M_q} b_m (2\mu_l - 1)^m. \quad (16)$$

Fig. 3. illustrates the modulated Farrow structure for modulation frequencies  $(2k+1)F_s/2$ . As above, modulation at  $(2k+1)F_s/2$  is attained simply by multiplying by  $\eta_l(n)$ . and by alternating the sign of coefficients  $c_m(n)$  for odd  $n$ .

Let us consider modulating the impulse response by a cosine at the frequency  $(2k+1)F_s/4$ . In each polynomial segment, the modulating function has a similar shape:(i) between the  $n$ th and  $(n+2)$ th segment, the shape is the same but the signs are opposite; (ii) adjacent segments are time-domain mirror images of each other with the same or opposite signs. The modulating functions can be written as

$$b_m(n) = \begin{cases} (-1)^{m/2} \cos((2k+1)(n + \frac{1}{2}) \frac{\pi}{2}) \frac{(\omega_0)^m}{(m)!}, & \text{for } m \text{ even,} \\ -(-1)^{(m-1)/2} \sin((2k+1)(n + \frac{1}{2}) \frac{\pi}{2}) \frac{(\omega_0)^m}{(m)!}, & \text{for } m \text{ odd.} \end{cases} \quad (17)$$

After applying series of trigonometric identities, the coefficients  $b_m(n)$  are reduced to:

$$b_m(n) = \begin{cases} (-1)^{\frac{m}{2}} (-1)^{\frac{(2k+1)n-1}{2}} \sin((2k+1) \frac{\pi}{4}) \frac{(\omega_0)^m}{(m)!}, & \text{for } m \text{ even,} \\ (-1)^{\frac{m+1}{2}} (-1)^{\frac{(2k+1)n-1}{2}} \cos((2k+1) \frac{\pi}{4}) \frac{(\omega_0)^m}{(m)!}, & \text{for } m \text{ odd,} \end{cases} \quad (18)$$

for odd  $n$ , and

$$b_m(n) = \begin{cases} (-1)^{\frac{m}{2}} (-1)^{\frac{(2k+1)n}{2}} \cos((2k+1) \frac{\pi}{4}) \frac{(\omega_0)^m}{(m)!} & \text{for } m \text{ even,} \\ (-1)^{\frac{m+1}{2}} (-1)^{\frac{(2k+1)n}{2}} \sin((2k+1) \frac{\pi}{4}) \frac{(\omega_0)^m}{(m)!} & \text{for } m \text{ odd.} \end{cases} \quad (19)$$

for even  $n$ . These coefficients can be further simplified based on the value of modulating integer  $k$  as shown in a sequel: for  $n$  even,  $k$  even in (20); for  $n$  even,  $k$  odd in (21); for  $n$  odd,  $k$  even in (22); for  $n$  odd,  $k$  odd in (23);

$$b_m(n) = \begin{cases} (-1)^{m/2} (-1)^{\frac{(2k+1)n}{2}} (-1)^{\frac{k}{2}} \cos(\frac{\pi}{4}) \frac{(\omega_0)^m}{(m)!} & \text{for } m \text{ even,} \\ -(-1)^{(m-1)/2} (-1)^{\frac{(2k+1)n}{2}} (-1)^{\frac{k}{2}} \sin(\frac{\pi}{4}) \frac{(\omega_0)^m}{(m)!} & \text{for } m \text{ odd.} \end{cases} \quad (20)$$

$$b_m(n) = \begin{cases} (-1)^{m/2} (-1)^{\frac{(2k+1)n}{2}} (-1)^{\frac{k-1}{2}} \sin(\frac{\pi}{4}) \frac{(\omega_0)^m}{(m)!} & \text{for } m \text{ even,} \\ -(-1)^{(m-1)/2} (-1)^{\frac{(2k+1)n}{2}} (-1)^{\frac{k-1}{2}} \cos(\frac{\pi}{4}) \frac{(\omega_0)^m}{(m)!} & \text{for } m \text{ odd.} \end{cases} \quad (21)$$

$$b_m(n) = \begin{cases} (-1)^{m/2} (-1)^{\frac{(2k+1)n-1}{2}} (-1)^{\frac{k}{2}} \sin(\frac{\pi}{4}) \frac{(\omega_0)^m}{(m)!} & \text{for } m \text{ even,} \\ -(-1)^{(m-1)/2} (-1)^{\frac{(2k+1)n-1}{2}} (-1)^{\frac{k}{2}} \cos(\frac{\pi}{4}) \frac{(\omega_0)^m}{(m)!} & \text{for } m \text{ odd.} \end{cases} \quad (22)$$

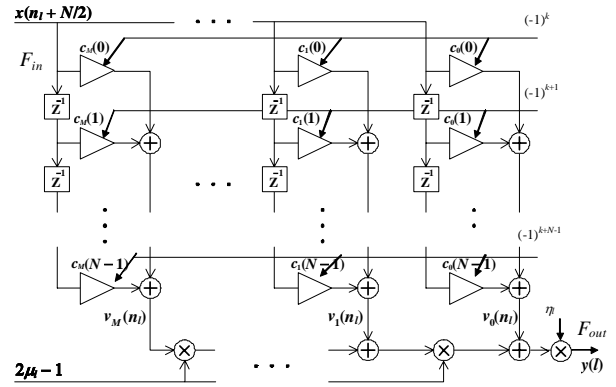


Fig. 3. The modulated Farrow structure for center frequencies  $(2k+1)F_s/2$ .

$$b_m(n) = \begin{cases} (-1)^{m/2} (-1)^{\frac{(2k+1)n-1}{2}} (-1)^{\frac{k-1}{2}} \cos(\frac{\pi}{4}) \frac{(\omega_0)^m}{(m)!} & \text{for } m \text{ even} \\ -(-1)^{(m-1)/2} (-1)^{\frac{(2k+1)n-1}{2}} (-1)^{\frac{k-1}{2}} \sin(\frac{\pi}{4}) \frac{(\omega_0)^m}{(m)!} & \text{for } m \text{ odd.} \end{cases} \quad (23)$$

It is possible to see from (20)-(23), that the values of coefficients  $b_m(n)$  in this case depend only on the polynomial order  $m$ , and sign is determined by  $n$ , and  $k$ . The main consequence is that all  $b_m(n)$ s are the same for given  $m$ , while the sign alternates according to (20)-(23). In order to reduce the number of operations needed for modulation, it is beneficial to decompose the  $m$ th-degree subfilter into two polyphase branches:

$$\begin{aligned} c_{m0}(n) &= c_m(2n) \\ c_{m1}(n) &= c_m(2n+1). \end{aligned} \quad (24)$$

In this way, two interleaved Farrow structures are obtained:

$$C_m(z) = C_{m0}(z) + z^{-1}C_{m1}(z). \quad (25)$$

In each interleaved Farrow structure the sign of coefficients is determined by  $n$  and  $k$ , and the output is multiplied by corresponding value of  $\eta_{l0}$  or  $\eta_{l1}$ . Fig. 4. shows the modulated Farrow structure for modulation frequencies  $(2k+1)F_s/4$ .

It is possible to see from (20)-(23), that the values of coefficients  $b_m(n)$  in this case depend only on the polynomial order  $m$ , and sign is determined by  $n$ , and  $k$ . The main consequence is that all  $b_m(n)$ s are the same for given  $m$ , while the sign alternates according to (20)-(23). In order to reduce the number of operations needed for modulation, it is beneficial to decompose the  $m$ th-degree subfilter into two polyphase branches:

$$\begin{aligned} c_{m0}(n) &= c_m(2n) \\ c_{m1}(n) &= c_m(2n+1). \end{aligned} \quad (26)$$

In this way, two interleaved Farrow structures are obtained:

$$C_m(z) = C_{m0}(z) + z^{-1}C_{m1}(z). \quad (27)$$

In each interleaved Farrow structure the sign of coefficients is determined by  $n$  and  $k$ , and the output is multiplied by corresponding value of  $\eta_{l0}$  or  $\eta_{l1}$ . Fig. 4. shows the modulated Farrow structure for modulation frequencies  $(2k+1)F_s/4$ .

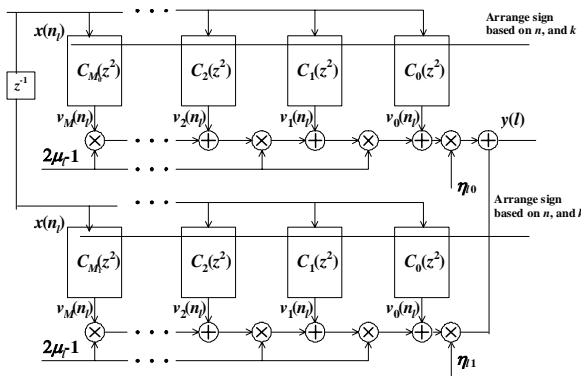


Fig. 4. The modulated Farrow structure for center frequencies  $(2k+1)F_s/4$ .

IV. DESIGN EXAMPLES

In order to illustrate the effectiveness of the proposed modulation method for the design of bandpass Farrow structure we use several illustrative examples. The first step is the design of the lowpass prototype filter with passband edge  $f_p=0.3F_s$ , stopband edge  $f_s=0.5F_s$ , required stopband attenuation  $A_s=60\text{dB}$ , and passband tolerance  $\delta_p=0.1$ . The filter, with performance shown in Fig 5., is designed using minimax design of [2], having  $N=14$  polynomial segments, and polynomial degree  $M=4$ .

The next step is to build the bandpass filter by modulating the obtained lowpass prototype. We use two example cases, with central frequencies of  $3F_s/4$  and  $2F_s$ . The corresponding polynomial-based modulation function that approximates cosine has the same length  $N=14$ , and polynomial order  $M_a=7$  for  $3F_s/4$ , and  $M_a=19$  for  $2F_s$ . According to Section 3, there are 10 additional multiplications to produce modulation by frequency  $3F_s/4$ , and also 10 to produce modulation by frequency  $2F_s$ . Frequency domain performances of modulated bandpass filters are shown in Figs. 5 and 7 respectively. Though there is a slight degradation of performance in stopband, the filtering requirements can be met by over designing the lowpass prototype filter.

V. CONCLUSIONS

We have presented a modulated Farrow structure, which allows band-pass realization with arbitrarily high centre frequencies without increasing the polynomial degree of the lowpass prototype Farrow filter. The modulating function is constructed as a low-order polynomial in order to avoid costly generation of trigonometric functions. The non-idealities of the modulating function are then mitigated by taking them into account when optimizing the filter coefficients.

ACKNOWLEDGEMENTS

This work was supported by the Serbian Ministry of Science under technology development projects: TR32028 – “Advanced Techniques for Efficient Use of Spectrum in Wireless Systems” and TR32023 – “Performance

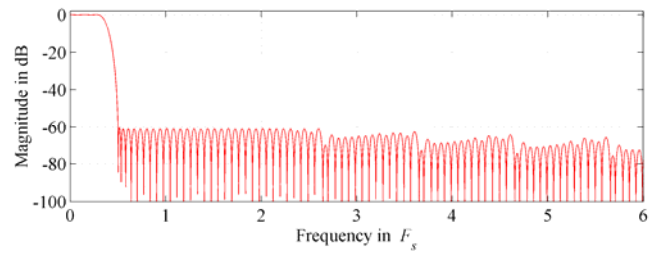


Fig. 5. The frequency response of the lowpass prototype.

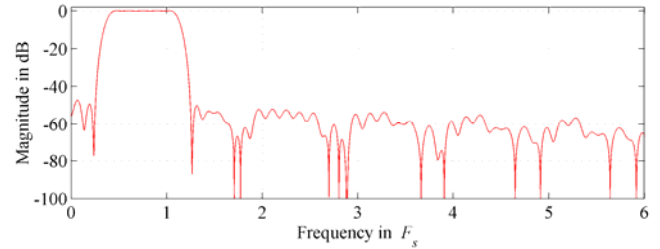


Fig. 6. The frequency response of the filter modulated by  $3F_s/4$ .

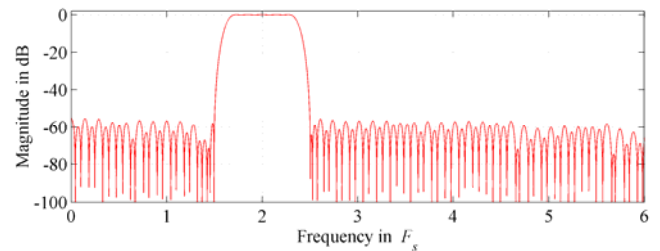


Fig. 7. The frequency response of the filter modulated by  $2F_s$ .

Optimization of Energy-efficient Computer and Communications Systems.”

REFERENCES

- [1] C.W. Farrow, “A Continuously Variable Digital Delay Element,” IEEE Int. Symp. Circ. Syst., pp. 2641–2645, Espoo, Finland, 1988.
- [2] J. Vesma and T. Saramäki, “Polynomial-based interpolation Filters - Part I: Filter synthesis,” Circuits, Systems, and Signal Processing, vol. 26, no. 2, pp. 115-146, March/April 2007.
- [3] D. Babic, T. Saramäki, M. Renfors, “Conversion between arbitrary sampling rates using polynomial-based interpolation filters,” in Proc. 2nd Int. TICSP Workshop on Spectral Methods and Multirate Signal Processing SMMSP’02, pp. 57-64, Toulouse, France, 2002.
- [4] R. Zukunft, S. Haar, T. Magesacher, “Digital interpolation in the passband domain,” Proc. IEEE Int. Conf. Acoust. Speech Signal Proc. (ICASSP), pp. 1545–1548, Orlando, USA, 2002.
- [5] H. Johansson, “Farrow-structure-based reconfigurable bandpass linear-phase FIR filters for integer sampling rate conversion,” IEEE Trans. on Circuits and Systems II: Express Briefs, (58), 1, pp. 46-50, 2011.
- [6] D. Babic, “Polynomial-based filters in bandpass interpolation and sampling rate conversion,” Proc. Int. Workshop on Spectral Methods and Multirate Signal Processing (SMMSP), pp. 31–37, Florence, Italy, 2006.
- [7] Greenberg, Michael, *Advanced Engineering Mathematics* (2nd ed.), Prentice Hall, ISBN 0-13-321431-1, 1998.
- [8] D. Babic, “Piecewise Polynomial Approximation Based on Taylor Series with Efficient Realization using Farrow Structure,” the 9th Int. Conference TELSIKS 2009, pp. 241-244, Niš, Serbia.



# Simulation of Codec for Adaptive Linear Prediction

Rumen P. Mironov<sup>1</sup>

**Abstract** – The software model of Adaptive Linear Prediction Codec for one dimensional signals is developed. The presented codec is simulated on Simulink for Matlab 6.5 environment and the obtained results for real sound signals are evaluated by the calculation of MSE and SNR for the decoded signals.

**Keywords** – Digital Signal Processing, Linear Prediction, Adaptive Data Compression, Matlab Simulation.

## I. INTRODUCTION

The basic methods for digital signal compression are lossless and lossy compression. Due to the low coefficient of information reduction, the first group of methods is used most frequently in archiving systems and didn't find widespread use in the information transmission systems. The methods for lossy compression are divided into three groups: statistical, psychoacoustic and transforms methods ([1], [2], [3] and [4]). The psychoacoustic compression methods are based on the shortcomings of the human hearing system and are suitable only when the recipient of the record signals is a man. The transformation methods provide a high degree of information reduction and work mainly on the basis of unitary transformations ([3], [4]), for which after decompression the specific block distortions are received. Statistical compression methods are based on the reduction of the information redundancy of the transmitted signals and proceeds in two stages: decorrelation of communication signals and reduction of binary digits, necessary for transmission of signals ([1], [2]). One of the simplest and most convenient for practical implementation methods for decorrelation is the method for linear prediction ([2], [5]).

An adaptive method for coding of one dimensional digital signals, based on the linear prediction and Least Mean Square (LMS) weight coefficients adaptation [6] of the prediction filter is presented. From the developed mathematical equations a general block scheme of Line Prediction Codec (LPC) is synthesized and experimental results from the simulation by Simulink for Matlab 6.5 environment for test signals are given.

## II. MATHEMATICAL DESCRIPTION

We will assume that the correlation covers  $n$  neighborhood elements of the input digital signal, represented by the stationary series  $\{x(i)\}$  with  $N > n$  values, which have zero

average component and correlation function  $R_x(r)$  for:  $r = \overline{0, n-1}$ . Then the basic equation of the linear prediction will be presented by the following way:

$$\hat{x}(i) = a_1x(i-1) + \dots + a_{n-1}x(i-n+1) = \sum_{k=1}^{n-1} a_kx(i-k), \quad (1)$$

where  $\hat{x}(i)$  is the value of the predicted element from the input signal  $x(i)$ . The prediction error is described by the equation:

$$e(i) = x(i) - \hat{x}(i), \quad (2)$$

and the quantization error - by the equation:

$$e_q(i) = Q[e(i)]. \quad (3)$$

According to the LMS algorithm of Widrow [6], the weight coefficients  $a_k$  of the prediction filter can be determined recursively:

$$a_k(i) = a_k(i-1) - \mu \nabla_k = a_k(i-1) - \mu \frac{\partial e^2(i)}{\partial a_k}, \quad (4)$$

where:  $\nabla_k$  is the gradient of the squared error by the prediction and  $\mu$  is the adaptation step for the  $i^{\text{th}}$  step of adaptation.

The derivative of the prediction error is determined by the Eqs. (1), (2), (3), and (4). The sequence for the components of weights is given by:

$$a_k(i) = a_k(i-1) + 2\mu e(i)x(i-k). \quad (5)$$

The beginning values of the weighted coefficients can be calculated after the minimization of mean square error of the prediction. From Eqs. (1) and (2) follows:

$$\overline{e^2(i)} = \frac{1}{N} \sum_{i=0}^{N-1} [x(i) - \hat{x}(i)]^2 = E \left\{ \left[ x(i) - \sum_{k=1}^{n-1} a_kx(i-k) \right]^2 \right\}, \quad (6)$$

where  $E$  is the averaging operator.

The partial derivative of  $e^2(i)$  with respect to any weight coefficient  $a_l$  can be expressed as:

$$\frac{\partial \overline{e^2(i)}}{\partial a_k} = \frac{\partial E \{ [x(i) - (a_1x(i-1) + \dots + a_{n-1}x(i-n+1))]^2 \}}{\partial a_l} = -2E \{ [x(i) - (a_1x(i-1) + \dots + a_{n-1}x(i-n+1))]x(i-l) \}$$

Once equated to zero and transformed, from the upper expression is obtained:

$$E \{ x(i)x(i-l) \} = \sum_{k=1}^{n-1} a_k E \{ x(i-k)x(i-l) \}. \quad (7)$$

The autocorrelation function of the digital signal, presented with the series  $\{x(i)\}$  is:

<sup>1</sup>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](mailto:rmironov@tu-sofia.bg)

$$R_x(r) = \frac{1}{N} \sum_{i=0}^{N-1} x(i)x(i-r) = E\{x(i)x(i-r)\}. \quad (8)$$

From the Eqs. (7) and (8) the following expression is obtained:

$$R_x(l) = \sum_{k=1}^{n-1} a_k R_x(k-l). \quad (9)$$

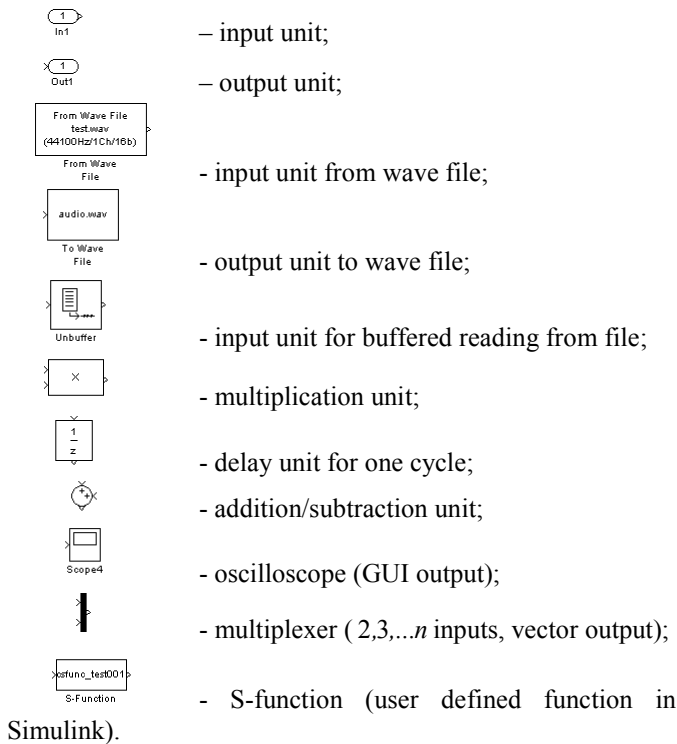
For  $l = \overline{1, n-1}$  equation (9) represent a linear system with  $n-1$  unknowns.

The following equations are used for decoding of signals:

$$\mathbf{x}'(i) = \mathbf{e}'(i) + \hat{\mathbf{x}}'(i), \quad (10)$$

$$\mathbf{e}'(i) = \mathbf{Q}^{-1}[\mathbf{e}_q(i)]. \quad (11)$$

The synthesized by the Eqs. from (1) to (11) adaptive line prediction codec (ADPCM) from 2<sup>nd</sup> order ( $k=2$  and provided that there is no quantization unit) is shown on Fig. 1. On Fig. 2 and Fig.3 the synthesized coder and decoder blocks for ADPCM are shown respectively. The presented schemes are developed through the Simulink package for Matlab 6.5 environment and included the following units:



### III. Experimental Results

The developed ADPCM codec is used for simulation on Matlab 6.5 environment of real audio signals (WAV file format, 1 channel (mono), 16 bits, 44.1 KHz sampling rate). The starting values of the weights are obtained from the correlation matrix:

$$[R_x] = \begin{bmatrix} R_x(0) & R_x(1) \\ R_x(1) & R_x(0) \end{bmatrix}$$

for autocorrelation functions  $R_x(0), R_x(1), R_x(2)$ , calculated from the Eq.(8). The starting weight vector  $\bar{A}(0) = [a_1(0), a_2(0)]^T$  is calculated from the equation:

$$[R_x]A(0) = \begin{bmatrix} R_x(1) \\ R_x(2) \end{bmatrix},$$

by the Kramer's formulas:  $a_1(0) = \frac{\Delta_1}{\Delta}$ ,  $a_2(0) = \frac{\Delta_2}{\Delta}$ .

Using Eq.(5) the values of the weight vector for the next iterations are calculated by the following:

$$\begin{bmatrix} a_1(i) \\ a_2(i) \end{bmatrix} = \begin{bmatrix} a_1(i-1) \\ a_2(i-1) \end{bmatrix} + \mu \cdot e(i) \cdot \begin{bmatrix} x(i-1) \\ x(i-2) \end{bmatrix},$$

where the errors are calculated from the Eq.(2) and the predicted values are calculated from:

$$\hat{x}(i) = a_1(i-1).x(i-1) + a_2(i-1).x(i-2).$$

The adaptation step depends on the inequality:  $0 < \mu < \frac{1}{R_x(0)}$  and the value  $\mu=0.01$  is chosen.

The mean squared error is calculated by the equation:

$$\overline{\varepsilon^2} = \frac{1}{N} \cdot \sum_{i=1}^N e^2(i).$$

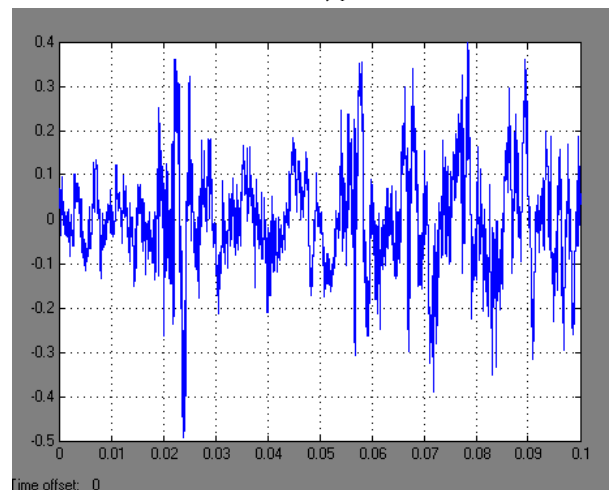


Fig. 4. Input signal  $x(i)$

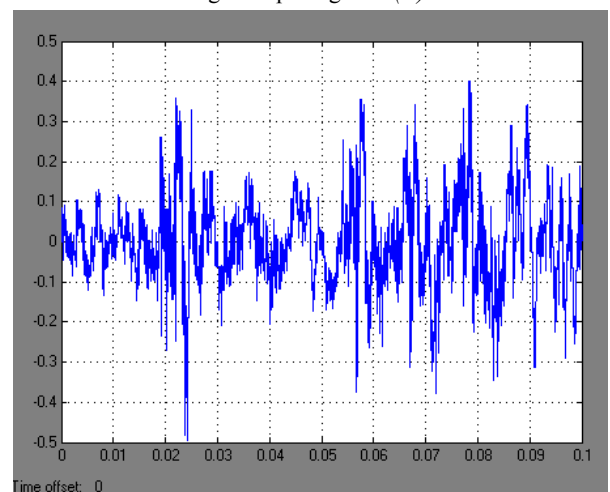


Fig. 5. Output predicted signal  $\hat{x}(i)$ .

The input test signal and output predicted signal are shown on Fig.4 and Fig.5 respectively. On Fig.6 the input and output signals are shown together (visualized from the oscilloscope Scope 4). On Fig.7 the same signals are visualized by the zooming in horizontal direction (from 0.031s to 0.0324s).

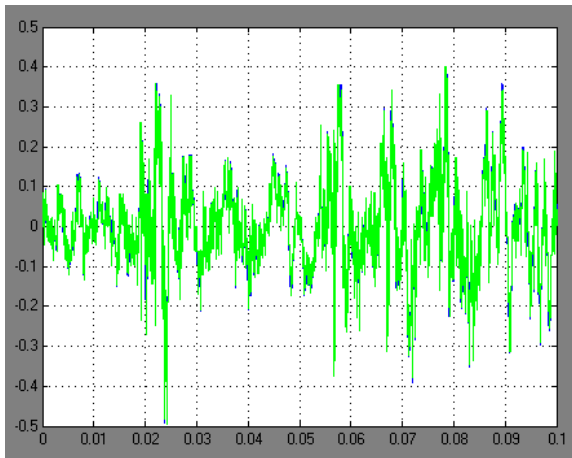


Fig. 6. Input and output predicted signal  $MUX\{x(i), \hat{x}(i)\}$ .

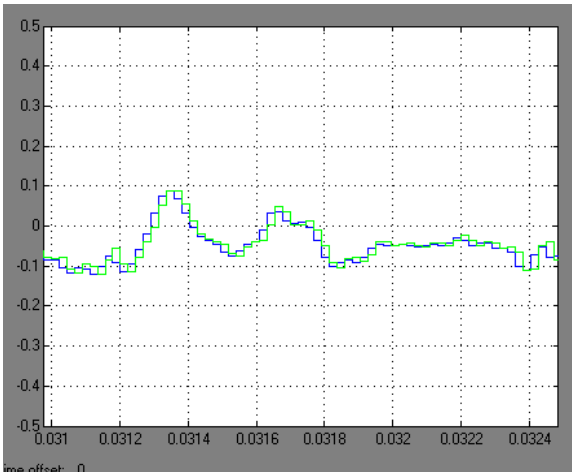


Fig. 7. Input and output signal  $MUX\{x(i), \hat{x}(i)\}$  (zoomed).

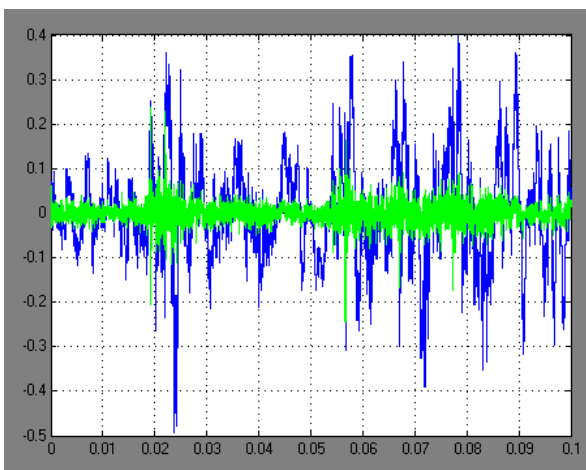


Fig. 8. Input and error signal  $MUX\{x(i), e(i)\}$ .

On Fig.8 the input signal and the error signal are visualized together.

#### IV. CONCLUSION

An adaptive method for coding of one-dimensional digital signals, based on the linear prediction and Least Mean Square (LMS) weight coefficients adaptation of the prediction filter is presented. From the developed mathematical equations an algorithm and a general block scheme of adaptive linear prediction codec is synthesized and experimental results from the simulation by Simulink for Matlab 6.5 environment for test signals in WAV format are given

The developed ADPCM codec provides minimum processing error and led to increase of PSNR with about 0.3 dB in comparison with 2 coefficients non-adaptive prediction codec.

The presented simulation model can be used in digital signal processing for spectral analysis, coding and transmission of one-dimensional signals and in distance learning by the using a Matlab Web Server.

The developed image codec is used in laboratory work on the disciplines: "Image and Signal Processing" and "Audio and Video Communication on Internet" and in the experimental work in laboratory "Electronic System for Visual Information" in Technical University of Sofia.

#### 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] D. Stranneby, W. Walker, *Digital signal processing and applications*, Elsevier, 2004.
- [2] S. Mitra, *Digital signal processing. A computer-based approach*, Mc Graw Hill, 2006.
- [3] V. Madisetti, D. Williams, Eds. *The Digital Signal Processing Handbook*, CRC Press, 1998.
- [4] A. D. Poularikas. *The Transforms and Applications Handbook*, Second Ed., CRC Press, 2000.
- [5] G. Blanchet, M. Charbit, *Digital Signal and Image Processing Using MATLAB*, ISTE, 2006.
- [6] B. Widrow, S. D. Stearns. *Adaptive Signal Processing*. Englewood Cliffs, New York, Prentice-Hall, Inc., 1985.

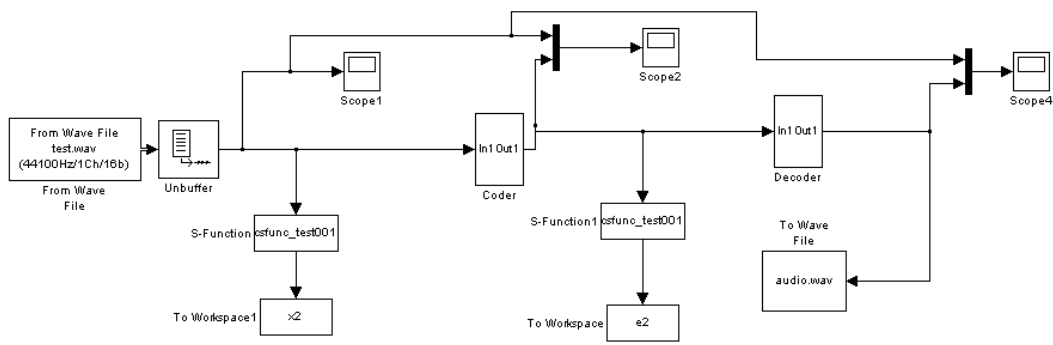


Fig. 1. ADPCM Codec.

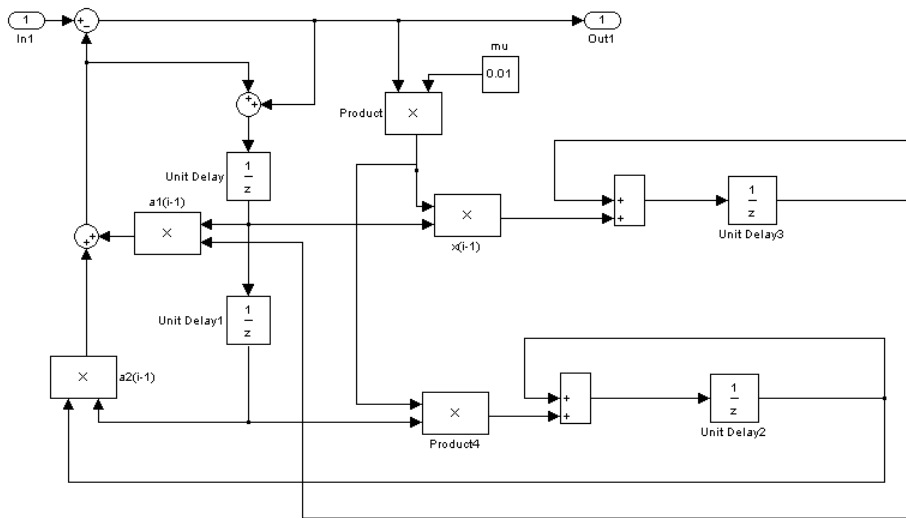


Fig. 2. ADPCM Coder.

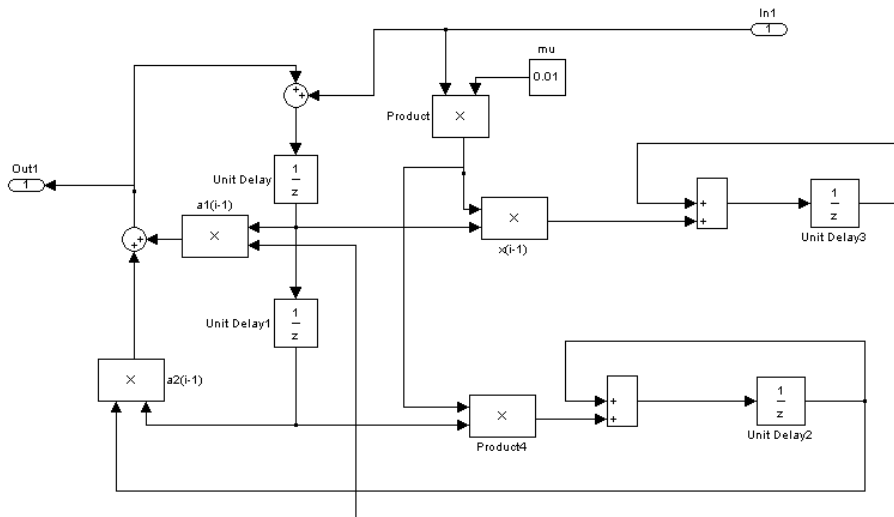


Fig. 3. ADPCM Decoder.

# Image Compression with Inverse Pyramid Decomposition over Wavelet Spectrum

Teodora G. Sechkova<sup>1</sup> and Ivo R. Draganov<sup>2</sup>

**Abstract** – In this paper a novel approach is presented for compression of digital images. It consists of finding the wavelet spectrum of an image into certain number of sub-bands after given levels of regular or irregular decomposition. Then each sub-band is decomposed with the inverse pyramid algorithm using some linear orthogonal transform such as DCT. Depending on whether a lossless or lossy compression is desired all or only some of the spectral coefficients from the inverse pyramid are preserved and then entropy coding is applied. Higher compression ratios are achieved at high image quality levels compared to some popular algorithms from the practice.

**Keywords** – Image Compression, Wavelet, IPD, DCT.

## I. INTRODUCTION

During the last three decades a large number of algorithms were designed for digital images compression. Many methods rely on spectral decorrelation combined with entropy coding [1]. Historically one of the most popular approaches that had become a standard and is still in wide use is the JPEG compression algorithm [2]. It can achieve extremely high compression ratios (over 100 times) at relatively high peak signal-to-noise ratios (*PSNR*, in some cases above 30 dB). Nevertheless some serious disadvantages could be pointed out for it such as the block effect due to the localizing property of the discrete cosine transform (DCT) and the lack of any scaling ability apart from the group change of the quantization matrix coefficients in some more flexible fashion, e.g. some multilevel or multistage processing. Extension like that could provide even more opportunities for scalable progressive transmission of image data over narrowband communication channels such as far distance satellite links for planetary observation.

Another large group of algorithms for image compression is based on the spatial-scale decomposition dividing the information for large and small objects from the image into separate bands. The wavelet transforms [3, 4] are the essence of such approaches and naturally led to the well-known JPEG 2000 standard [5] in which the block effect is absent. Here some more advanced techniques were introduced at the level of information redundancy reduction such as the Embedded Zero-Tree Wavelet (EZW) coding and the Embedded Block Coding with Optimal Truncation (EBCOT). Another advantage is the ability for scalable transmission (restoration)

of the image based on partial information transmitted in terms of bit-planes. Despite all the novelties in this approach at the level of visual objects inside the image some real multi-level scheme is thought to be proper for further enhancement.

Recently a new approach has been proposed for digital images decomposition denoted as inverse pyramid [6]. It introduces decomposing levels for a single image by dividing it in smaller blocks over which linear orthogonal transform of any kind could be applied. Then restoring of the blocks in the next level is done by using only some of the spectrum coefficients while the difference between the original and the approximated blocks is being preserved and passed to the next level. This particular decomposition is thought to be very efficient and flexible for digital images compression as well as in the field of pattern recognition for simplified object representation and speeding up algorithms. It is in the base of the proposed here algorithm along with the advantages given by the wavelet transforms.

In the next section detailed description of the new algorithm is presented, then in part three some experimental data is given and then conclusion is made in part four.

## II. ALGORITHM DESCRIPTION

The input image  $I$  is grayscale consisting of  $N$  by  $N$  pixels and intensity range from 0 to 255. It is transformed by using the Haar wavelet according to:

$$B_{2 \times 2} = T I_{2 \times 2} T^T, \quad (1)$$

where  $I_{2 \times 2}$  is a block of the image with size  $2 \times 2$  and  $T$  – the Haar transform matrix of the same size given by:

$$T = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \quad (2)$$

Going through all  $2 \times 2$  blocks of the image and rearranging the resulting spectral coefficients from the resulting spectrum  $B$  into 4 sub-bands the complete 1-level Haar decomposition is achieved.

Then each sub-band of the wavelet spectrum  $B_R$  ( $R = 1 \div 4$ ) is divided to blocks with dimensions  $2^n \times 2^n$  and each of them is presented with Inverse Pyramidal Decomposition (IPD) [6]:

$$[B_R(2^n)] = [\tilde{B}_{0R}(2^n)] + \sum_{p=1}^2 [\tilde{E}_{p-1,R}(2^n)] + [E_{2,R}(2^n)], \quad (3)$$

<sup>1</sup>Teodora G. Sechkova is with the Faculty of Telecommunications, 8 Kliment Ohridski Blvd., 1000 Sofia, Bulgaria, E-mail: teodora.sechkova@gmail.com

<sup>2</sup>Ivo R. Draganov is with the Faculty of Telecommunications, 8 Kliment Ohridski Blvd., 1000 Sofia, Bulgaria, E-mail: idraganov@tu-sofia.bg

where 3 levels are used and  $[E_{2,R}(2^n)]$  is the matrix of the residual from the decomposition. In (3) each matrix is with dimensions  $2^n \times 2^n$ . The first component  $[\tilde{B}_{0R}(2^n)]$  for the level  $p = 0$  is a rough approximation of the block  $[B_R(2^n)]$ . It is obtained by applying inverse 2D-DCT over the transformed block in correspondence with the expression:

$$[\tilde{B}_{0R}(2^n)] = [T_0(2^n)]^{-1} [\tilde{S}_{0R}(2^n)] [T_0(2^n)], \quad (4)$$

where  $[T_0(2^n)]^{-1}$  is a matrix with dimensions  $2^n \times 2^n$  for the inverse 2D-DCT.

The matrix is the transform block of the cut 2D-DCT over  $[B_R(2^n)]$ . Here  $m_0(u,v)$  are the elements of the binary matrix-mask  $[M_0(2^n)]$  with the help of which the preserved coefficients are being determined  $[\tilde{S}_{0R}(2^n)]$  in accordance to the equation:

$$m_0(u,v) = \begin{cases} 1, & \text{if } s_{0R}(u,v) \text{ is retained coefficient,} \\ 0 & \text{otherwise,} \end{cases} \quad (5)$$

for  $u, v = 0, 1, \dots, 2^n - 1$ .

The values of the elements  $m_0(u,v)$  are chosen by the condition the retained coefficients to correspond to those with the highest average energy into the transformed blocks for all the blocks to which the image has been divided. The transformed block from  $[B_R(2^n)]$  is found by the 2D-DCT:

$$[S_{0R}(2^n)] = [T_0(2^n)] [B_R(2^n)] [T_0(2^n)], \quad (6)$$

where  $[T_0(2^n)]$  is a matrix with dimensions  $2^n \times 2^n$  for level  $p = 0$  which is used for implementing the DCT.

The rest components in decomposition (3) are the approximation matrices for  $p = 1, 2$ . Each of them consists of sub-matrices with dimensions  $2^{n-p} \times 2^{n-p}$  for  $k_p = 1, 2, \dots, 4p$  obtained by its quad-tree split. On the other hand each sub-matrix is calculated by:

$$[\tilde{E}_{p-1,R}^{k_p}(2^{n-p})] = [T_p(2^{n-p})]^{-1} [\tilde{S}_{pR}^{k_p}(2^{n-p})] [T_p(2^{n-p})]^{-1} \quad (7)$$

for  $k_p = 1, 2, \dots, 4p$ . Here  $4p$  is the number of the branches of the quad-tree in level  $p$  of the decomposition;  $[T_p(2^{n-p})]^{-1}$  - matrix for inverse 2D-DCT;  $[\tilde{S}_{pR}^{k_p}(2^{n-p})]$  - the transformed block of the cut 2D-DCT of the difference matrix  $[E_{p-1,R}^{k_p}(2^{n-p})]$ . The elements  $\tilde{s}_{pR}^{k_p}(u,v) = m_p(u,v) \cdot s_{pR}^{k_p}(u,v)$  of the matrix  $[\tilde{S}_{pR}^{k_p}(2^{n-p})]$  depend on the elements  $m_p(u,v)$  of the binary mask  $[M_p(2^{n-p})]$  for  $u, v = 0, 1, \dots, 2^{n-p} - 1$  according:

$$m_p(u,v) = \begin{cases} 1, & \text{if } s_{pR}^{k_p}(u,v) \text{ - retained coefficient,} \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

Here  $s_{pR}^{k_p}(u,v)$  are elements of the transformed block  $[S_{pR}^{k_p}(2^{n-p})]$  which is obtained also by the 2D-DCT:

$$[S_{pR}^{k_p}(2^{n-p})] = [T_p(2^{n-p})] [E_{p-1,R}^{k_p}(2^{n-p})] [T_p(2^{n-p})], \quad (9)$$

where  $[T_p(2^{n-p})]$  is a matrix with dimensions  $2^{n-p} \times 2^{n-p}$  for level  $p = 0$  by which DCT is applied.

It is possible to represent each group of four neighbouring elements  $\tilde{s}_{pR}^{k_p}(u,v)$  for one and the same  $u$  and  $v$  in the way defined by (10) which allows to gain even higher correlation between the spectral coefficients since the last three ones for positions (0,1), (1,0) and (1,1) form differences two by two and these differences often are zero valued because neighboring blocks contain almost identical content.:

$$\begin{bmatrix} \tilde{d}_{pR}^{k_p}(0,0) \\ \tilde{d}_{pR}^{k_p}(0,1) \\ \tilde{d}_{pR}^{k_p}(1,0) \\ \tilde{d}_{pR}^{k_p}(1,1) \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 4 & 0 & -4 \\ -4 & 0 & 4 & 0 \\ 0 & 0 & -4 & 4 \end{bmatrix} \begin{bmatrix} \tilde{s}_{pR}^{k_p}(0,0) \\ \tilde{s}_{pR}^{k_p}(0,1) \\ \tilde{s}_{pR}^{k_p}(1,0) \\ \tilde{s}_{pR}^{k_p}(1,1) \end{bmatrix} \quad (10)$$

The inverse transform which leads to full restoration of  $\tilde{s}_{pR}^{k_p}(u,v)$  is given by:

$$\begin{bmatrix} \tilde{s}_{pR}^{k_p}(0,0) \\ \tilde{s}_{pR}^{k_p}(0,1) \\ \tilde{s}_{pR}^{k_p}(1,0) \\ \tilde{s}_{pR}^{k_p}(1,1) \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 4 & -1 & -3 & -2 \\ 4 & 3 & 1 & 2 \\ 4 & -1 & 1 & -2 \\ 4 & -1 & 1 & 2 \end{bmatrix} \begin{bmatrix} \tilde{d}_{pR}^{k_p}(0,0) \\ \tilde{d}_{pR}^{k_p}(0,1) \\ \tilde{d}_{pR}^{k_p}(1,0) \\ \tilde{d}_{pR}^{k_p}(1,1) \end{bmatrix}. \quad (11)$$

The difference matrix  $[E_{p-1,R}(2^{n-p})]$  for level  $p$  containing the sub-matrices  $[E_{p-1,R}^{k_p}(2^{n-p})]$  is determined by the following equation:

$$[E_{p-1,R}(2^{n-p})] = \begin{cases} [B_R(2^n)] - [\tilde{B}_{0R}(2^n)] & \text{- for } p = 1; \\ [E_{p-2,R}(2^{n-p})] - [\tilde{E}_{p-2,R}(2^{n-p})] & \text{- for } p = 2. \end{cases} \quad (12)$$

Over the coefficients  $\tilde{s}_{pR}^{k_p}(u,v)$  for all the levels of the pyramid for each sub-band a lossless entropy coding is applied which includes run-length encoding, Huffman coding and arithmetic coding. The resulting values could be stored in a file which volume determines the compressed image size.

A decoder for restoring the compressed images with the proposed approach should consist of all the opposite operations to those described by (1)-(12).

### III. EXPERIMENTAL RESULTS

Two test images were used with the proposed approach called Flower and Gargoyl. They are 8 bpp (bits per pixel) grayscale images with size 512x512 pixels shown in Fig. 1. Lossy compression was applied over them with 3 levels of the IPD after the Haar transform with 1 level of decomposition into 4 sub-bands –from LL to HH. On each level of the pyramid the 4 lowest frequencies in the DCT spectrum were retained when working with  $n = 4, 3$  and  $2$  respectively, that is DCT transform matrix of  $16 \times 16$  for 0<sup>th</sup> level,  $8 \times 8$  – for the 1<sup>st</sup> and  $4 \times 4$  – for the last one.

To compare the obtained results in terms of quality and compression levels LuraWave SmartCompress 3 [7] application was used to compress the test images using the JPEG2000 standard (in JP2 format). The results are given in Table 1. Lossless compression is also possible by preserving the residual after level 2 in the pyramid but which here is omitted.

In the range between 0,5 and 2 bpp for the compression ratio (CR) there is tangible overweight in the PSNR achieved by the proposed approach over the JPEG2000 algorithm from about 4 dB for the higher ratios to around 0,05 dB. Obviously going towards lossless compression both algorithms tend to flatten the results they achieve. And the opposite, with the reduction of the CR the inverse pyramid decomposition over the wavelet spectrum of the image produces higher quality restored images at equal compression levels.

TABLE I  
COMPRESSION RESULTS

CR, bpp	Flower Test Image		Gargoyl Test Image	
	PSNR, dB		PSNR, dB	
	IPD	JPEG2000	IPD	JPEG2000
0,5	34,772	30,876	28,689	27,594
1	35,698	34,952	29,075	28,632
2	36,615	36,560	29,351	28,691

The visual analysis of the restored images from both algorithms (Fig. 1) reveals the same tendency – for low CR levels most of the smaller details are preserved in almost identical way for IPD and JPEG2000. At the same time large homogenous areas are clean from any additional artifacts introduced by the compression. But when CR begins to decrease considerably JPEG2000 blurs most of the smaller details in the image mainly because of the quantization of the wavelet spectral coefficients and specially of the low-level HH components. In the new approach where IPD is used no quantization of the wavelet spectrum is applied directly. The second transform in the pyramid (DCT) produces highly efficient decorrelation of these coefficients and even then they are being quantized for the given level retaining the error for the next level. In such a way much smoother transition is guaranteed from level to level not omitting the details from the image in intolerable degree. Another major advantage is the almost full absence of artifact distortions in the even areas because of the properties of the wavelet spectrum which is used in a first place but not the original image as it is with the

classical JPEG approach. Here the block effect is much more suppressed.

### IV. CONCLUSION

In this paper a new approach is presented for highly efficient image compression using wavelet and inverse pyramid decomposition. The quality and the compression level of the images could be smoothly controlled by choosing different levels of both the decompositions and the quantization mask for the pyramid at given level. Progressive image transmission from level to level is possible which proves to be very useful in the case of narrow-band communication channels.

Incorporating the advantages of the wavelet decomposition which minimizes the block effects due to the absence of localizing properties with the strong decorrelating properties of the orthogonal transforms such as DCT it is possible now to have strongly compressed digital images which when restored have low presence of artifacts. As it is obvious from experimental results higher compression ratios are achieved at equal image quality levels compared to some popular algorithms from the practice such as JPEG2000. This makes the proposed approach a suitable candidate for further improvement and establishing a new codec design for highly efficient image compression.

### ACKNOWLEDGEMENT

This paper was supported by the National Science Fund of the Bulgarian Ministry of Education, Youth and Science (Contract – DDUVU 02/13 – “Public and Private Multimedia Network Throughput Increase by Creating Methods for Assessment, Control and Traffic Optimization”).

### REFERENCES

- [1] K. R. Rao and P. Yip, *The Transform and Data Compression Handbook*, CRC Press, Boca Raton, USA, 2001.
- [2] G. Wallace, “The JPEG Still Picture Compression Standard”, In *IEEE Transactions on Consumer Electronics*, Vol. 38, Issue 1, pp. 18-34, 1992.
- [3] J. Walker and T. Nguyen, “Wavelet-Based Image Compression”, *Handbook of Transforms and Data Compression*, Ch. 6, pp. 267 - 312, CRC Press, Boca Raton, 2000.
- [4] S. Jacob and A. Cheeran, “Wavelet Based Image Compression”, In *Proceedings of the International Conference and Workshop on Emerging Trends in Technology ICWET’10*, Mumbai, India, p. 999, 2010.
- [5] A. Skodras, C. Christopoulos, and T. Ebrahimi, “The JPEG 2000 Still Image Compression Standard”, *IEEE Signal Processing Magazine*, Vol. 18, Issue 5, pp. 36-58, 2001.
- [6] R. Kountchev and R. Kountcheva “Image Representation with Reduced Spectrum Pyramid”. Book chapter in: “New Directions in Intelligent Interactive Multimedia”, Eds. G. Tsihrintzis, M. Virvou, R. Howlett, L. Jain, Springer, Berlin, pp. 275-284, 2008.
- [7] <http://www.luratech.com>

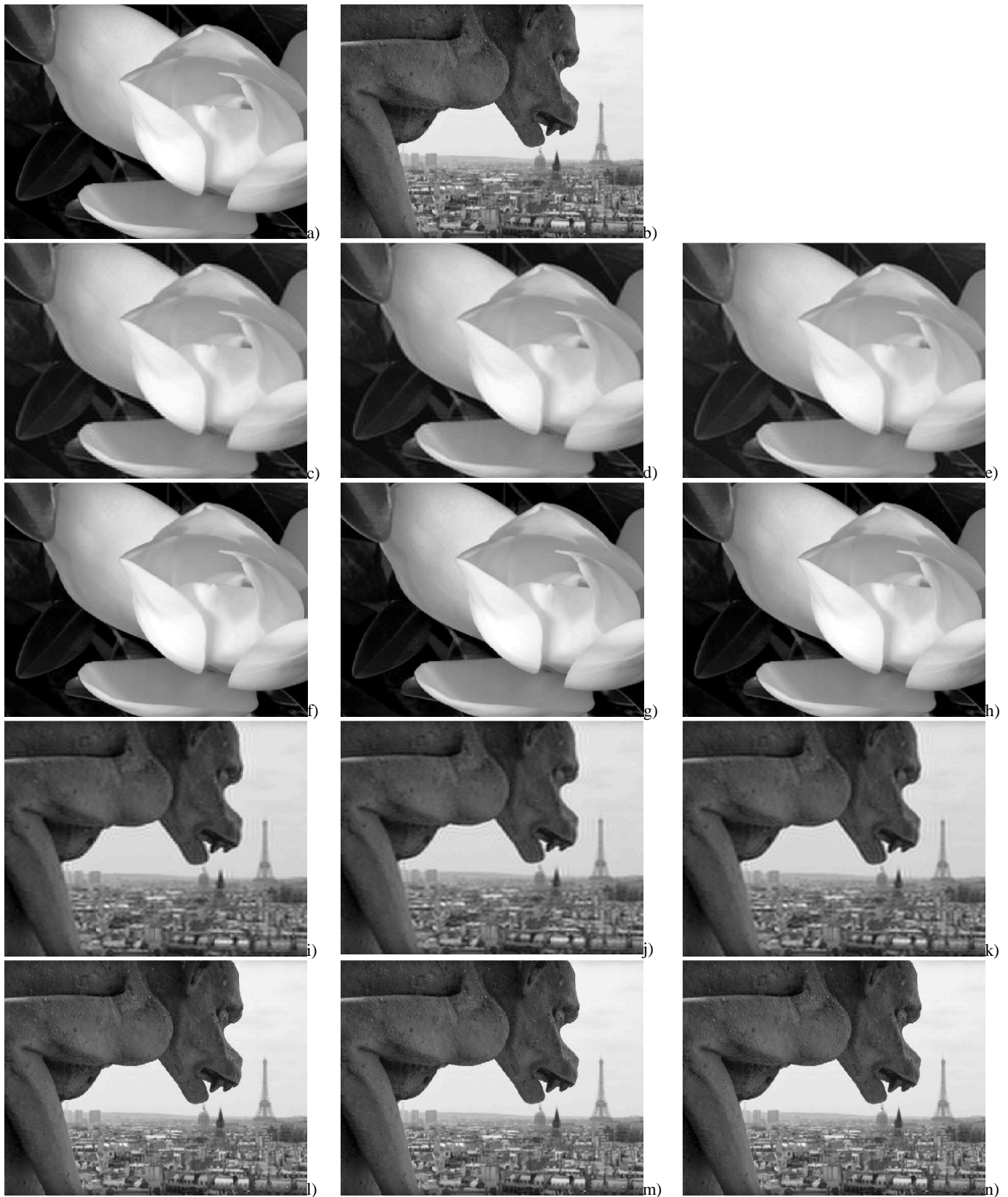


Fig. 1. Visual test results: a) original Flower image; b) original Gargoyle image; c) IPD 0,5 bpp; d) IPD 1 bpp; e) IPD 2 bpp; f) JP2 0,5 bpp; g) JP2 1 bpp; h) JP2 2 bpp; i) IPD 0,5 bpp; j) IPD 1 bpp; k) IPD 2 bpp; l) JP2 0,5 bpp; m) JP2 1 bpp; n) JP2 2 bpp.



# Efficient Adaptive Local Binarization Algorithm for Text Extraction from Image with Complex Background

Antoaneta Popova<sup>1</sup>

**Abstract** – This paper describes a fast adaptive local thresholding algorithm with the same quality as the Sauvola method and as fast as the global thresholding methods. The algorithm is independent of the operator window size and combines the advantages of the Wiener filter preprocessing, the adaptive local threshold calculation and the integral images.

**Keywords** – Binarization, local thresholding, integral images, Wiener filter.

## I. INTRODUCTION

The binarization converts the input grayscale or color image into a bi-level representation. Most document analysis systems use this approach. The next steps in document analysis like Optical Character Recognition (OCR) heavily depend on the result of binarization algorithm. Several different methods for image thresholding have been proposed the last decade.

In this paper we focus on the binarization of color or grayscale images with superimposed text. There are two groups of binarization methods: global binarization and local binarization. Global binarization methods like that of Otsu, IsoData, K-Means define a single threshold value for the whole image and are very fast [1]. They give good results for typical scanned documents, but are not appropriate for our goal to develop an efficient and fast thresholding for text extraction from image with complex background: for example Web pages or camera-captured documents; images with not uniform illumination; noised / degraded images; images with significant intensity changes.

Local binarization methods [2, 3, and 4] try to overcome problems of the complex background by computing thresholds individually for each pixel using information from the local neighborhood of the pixel. The achieved results are good, but they are often slow because the local neighborhood is computed for each pixel.

This paper presents a modified approach with a good performance for computing of adaptive local thresholds with speed close to the global thresholding methods. The suggested algorithm has considerably lower computational complexity. The proposed adaptive threshold level computation gives smooth, faster image binarization and often it has better noise robustness. The “coarse to fine” approach combining Wiener filter [5], adaptive local thresholding and integral images [6, 7] is applied.

<sup>1</sup>Antoaneta Popova is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: antoaneta.popova@tu-sofia.bg.

The following section (Sec. 2) describes the modified approach for combining integral images with the local adaptive thresholding techniques. The evaluation of the suggested algorithm is described in Sec. 3, followed by conclusion in Sec. 4.

## II. ADAPTIVE LOCAL BINARISATION ALGORITHM AND BLOCK DIAGRAM

Integral image-based representation is applied in the modified algorithm, because it allows very fast multi-scale image processing.

An integral image  $I(x,y)$  of an input grayscale image  $I_g(x,y)$  is defined as the image in which the intensity at a pixel position is equal to the sum of the intensities of all the pixels above and to the left of that position in the original image. So the intensity at position  $(x, y)$  can be written as:

$$I(x, y) = \sum_{i=0}^x \sum_{j=0}^y I_g(i, j). \quad (1)$$

The integral image is computed in a single pass. A cumulative row sum  $S(x, y)$  is used to calculate the integral image  $I(x,y)$ :

$$\begin{aligned} S(x, y) &= S(x, y-1) + I_g(x, y) \\ I(x, y) &= I(x-1, y) + S(x, y). \end{aligned} \quad (2)$$

In Fig. 1 a region A of the integral image can be computed using the following 4 array references:

$$A = (A_1 + A_4) - (A_2 + A_3). \quad (3)$$

If the integral image is the sum of pixels in a given area and it can be represented as two addition and one subtraction operations.

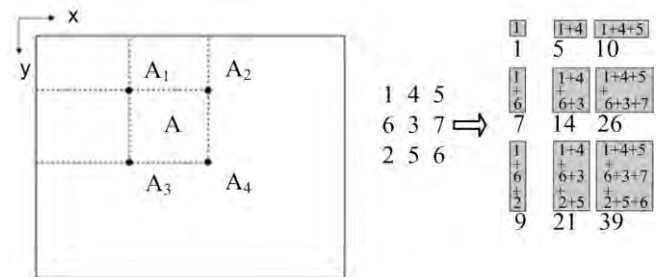


Fig. 1. Integral image feature computation

For example the sum of the 4 right-down pixels (3, 7, 5, 6) of the input grayscale image block is calculated faster from the elements of the integral image (right block in Fig. 1) without using the top row and the left column of the input image:

$$3 + 7 + 5 + 6 = (1 + 39) - (9 + 10). \quad (4)$$

Then the local mean  $m(x,y)$  for any window size  $w$  is computed by using two addition and one subtraction operations instead of the summation over all pixel values in the window:

$$m(x,y) = \frac{\left( \begin{matrix} (I(x + \frac{w}{2}, y + \frac{w}{2}) + I(x - \frac{w}{2}, y - \frac{w}{2})) \\ - (I(x + \frac{w}{2}, y - \frac{w}{2}) - I(x - \frac{w}{2}, y + \frac{w}{2})) \end{matrix} \right)}{w^2} \quad (5)$$

Similarly, the local variance  $\delta$  is computed very efficiently, independent of the local window size  $w$ :

$$\delta^2(x,y) = \frac{1}{w^2} \sum_{i=x-w/2}^{x+w/2} \sum_{j=y-w/2}^{y+w/2} I_g^2(i,j) - m^2(x,y). \quad (6)$$

An adaptive local binarization algorithm block diagram is presented in Fig. 2.

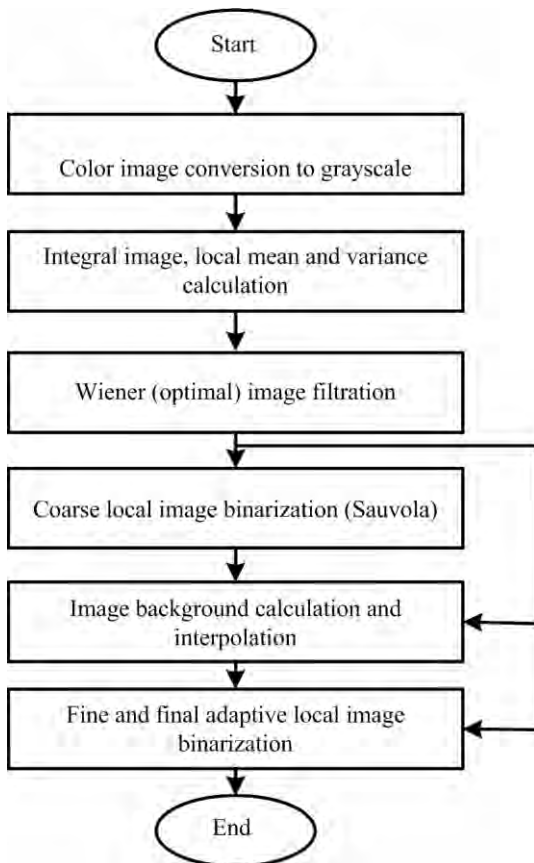


Fig. 2. Local binarization algorithm block diagram

The main algorithm steps are the following:

**1 step:** A color image conversion to grayscale image  $I_g$  is done, allowing color depth 2/24/32 bits:

$$I_g = 0,3.R + 0,59.G + 0,11.B. \quad (7)$$

**2 step:** Integral image, local mean and variance calculation is applied according to the above description.

**3 step:** Wiener (optimal) image filtration is selected for a preprocessing stage as more effective and faster, with low complexity and giving high accuracy of the OCR results. The Wiener filtered image is calculated:

$$I_{wf}(x,y) = \frac{m + (\sigma^2 - \sigma_n^2) \cdot (I_g(x,y) - m)}{\delta^2}, \quad (8)$$

where  $\delta_n$  is a standard noise variance, defined from the whole image.

**4 step:** Coarse local image binarization (Sauvola) is applied first, calculating a threshold for each pixel:

$$\theta(x,y) = m(x,y) \left[ 1 + k \left( \frac{\delta(x,y)}{R} - 1 \right) \right], \quad (9)$$

where  $k$  parameter has a value 0,2 and  $R$  is the defined in advance dynamic range of the standard variance  $\delta$ . This algorithm uses the mean and the variance around the pixel in a local area, but threshold is better adaptively calculated for the background with changed intensity.

A pixel  $S(x,y) = 1$  is accepted as part of the text object if its grayscale intensity  $I_g$  is less than the threshold  $\theta$ :

$$S(x,y) = 1 \quad \text{if} \quad I_g(x,y) < \theta(x,y) \\ S(x,y) = 0, \quad \text{else} \quad (10)$$

else the pixel is part of the background  $S(x,y) = 0$ .

**5 step:** Image background  $B_g$  calculation and interpolation is performed to obtain an image with background only – the previously selected text pixels now are interpolated with the closest intensity of the background pixels. The pixels in the text object are only changed in this step. The purpose is to achieve an equalized background for a more successful following finer binarization:

$$B_g(x,y) = \begin{cases} I(x,y) & \text{if } S(x,y) = 0 \\ \frac{\sum_{i=x-dx}^{x+dx} \sum_{j=y-dy}^{y+dy} (I(i,j) \cdot (1 - S(i,j)))}{\sum_{i=x-dx}^{x+dx} \sum_{j=y-dy}^{y+dy} (1 - S(i,j))} & \text{if } S(x,y) = 1 \end{cases}, \quad (11)$$

where  $2dx$  and  $2dy$  are the size of the operator window (typically  $40 \times 40$  px. to cover minimum two characters). If the pixel is part of the background  $S(x,y) = 0$  it has the same unchanged intensity  $I(x,y)$ .

**6 step:** Fine and final adaptive local binarization over the grayscale input image is applied, taking in account the calculated background image. The subtraction between the background and the input image ( $B_g(x,y) - I(x,y)$ ) gives an image with the replaced position parts of the text and background (darker) and the image can be assumed to be with a constant background.

For the background the pixels  $B_g(x,y)$  and  $I(x,y)$  are equal and the background pixel intensities are not considered. The equation for achieving the fine and final binarized image is:

$$I_f(x,y) = \begin{cases} 1 & \text{if } (B_g(x,y) - I(x,y)) > d(B_g(x,y)) \\ 0 & \text{else} \end{cases}, \quad (12)$$

where  $d(B_g(x,y))$  is a local threshold for pixel  $(x,y)$ .

For images with a constant background the value of the minimal threshold between text and background is:

$$d = q \cdot \delta_f, \quad (13)$$

where  $q$  is a weight coefficient, and  $\delta_f$  is a mean distance between the background and the text and it is calculated as follows:

$$\delta_f = \frac{\sum_x \sum_y (B_g(x,y) - I(x,y))}{\sum_x \sum_y S(x,y)}. \quad (14)$$

For achieving different adaptive thresholds for different contrasts between the text and background, the mean value  $b$  of the intensity of the grayscale pixels  $B_g(x,y)$ , belonging to the coarse binarized text, is calculated using the equation:

$$b = \frac{\sum_x \sum_y (B_g(x,y)(1 - S(x,y)))}{\sum_x \sum_y (1 - S(x,y))}. \quad (15)$$

The threshold for the given pixel  $(x,y)$  is equal to  $q \cdot \delta_f$ , if the intensity of the image  $B_g(x,y)$  is bigger than  $p1 \cdot b$ . Else the threshold is equal to  $p2 \cdot q \cdot \delta_f$ . The coefficients  $p1$  and  $p2$  are in the range  $[0; 1]$ . The above is achieved using the function:

$$d(B_g(x,y)) = q \cdot \delta_f \left( \frac{1 - p2}{1 + \exp\left(\frac{4B_g(x,y)}{b(1 - p1)} + \frac{2(1 + p1)}{1 - p1}\right)} + p2 \right). \quad (16)$$

### III. EXPERIMENTS AND RESULTS

The conducted tests include images with text over the picture, with darker background parts, and different background artifacts.

The algorithm implementation is done using program language C. During the experimental testing the selection of

the window size  $w$  is done in Wiener filter (3 px.) for preserving the character contours, in the local area for Sauvola (15 px.) and in the local area for background calculation (40 px.).

In order to demonstrate the advantages (speed and ORC accuracy after text image binarization) of our proposed algorithm, 5 existing binarization algorithms (IsoData, Otsu, Local Mean, Niblack, Sauvola) were tested.

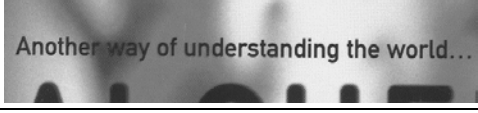
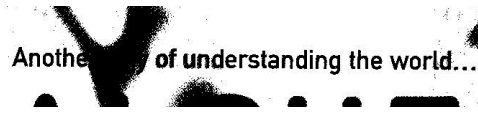

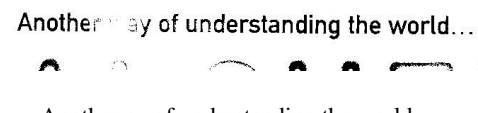
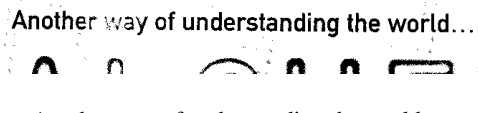
Used algorithm	Binarized image and Recognized text
Original	
Otsu	 Another of understanding the world... Am 5. of understanding the world...
Niblack	 idini FI; - .__ ii: F
Sauvola	 Another ay of understanding the world...
Suggested algorithm	 Another way of understanding the world...

Fig. 3. Original image, binarized by Otsu, Niblack, Sauvola and the suggested algorithm with the corresponding recognized texts

In Fig. 3 is presented the original image with superimposed text on a complex background, the binarized images comparing the 4 methods - Otsu, Niblack, Sauvola and the suggested algorithm in this paper, and the recognized characters from OCR system (shown under the binarized images). After the Otsu global binarization 9 text symbols cannot be recognized. Using the Niblack binarization many black pixels pass from the background to the text object. This is the worst case and no one of 36 characters/ symbols are recognized. As it is shown in the Sauvola algorithm, binarization results in the threshold being higher than the optimal and 3 characters are almost unreadable in the binary image, but the OCR system helps and only one symbol is not recognized. In the suggested by us adaptive local binarization

algorithm, the text is maximally readable and all 36 characters are recognized.

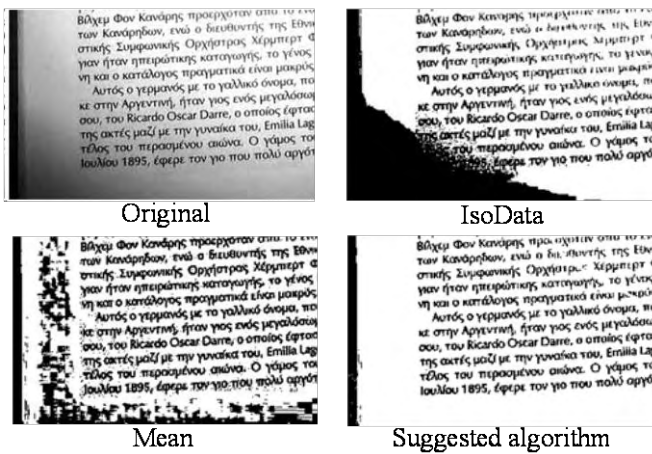


Fig. 4. Original image, binarized by IsoData, Mean and the suggested algorithm

The application of the proposed algorithm, as well as IsoData and Mean algorithms, on test images with dark areas due to bad scanning shows that the suggested approach better cleans such backgrounds (Fig. 4).

The typical binarization algorithm calculates local image areas sequentially around each pixel. The calculation complexity presented as a number of operations  $O$  is:

$$C = O(N^2 \cdot W^2), \quad (17)$$

where  $N^2$  is the area of the whole image in pixels,  $W^2$  is the area of the operator window processing each pixel. We allow maximal image  $4000 \times 3000 \text{ px.}$ , an operator window  $W^2 = 20 \times 20 \text{ px.}$  In this case the number of the arithmetical operation is  $4.9e+9$ . This speed is equal to 5 sec. for CPU with 20000 MIPS.

Using the integral image approach in the proposed algorithm the arithmetical operations are reduced to

$$C_I = 3(N^2). \quad (18)$$

The achieved speed of the suggested algorithm with the applying integral images is 3 sec. for the same CPU, that means 1,7 times faster only for one algorithm step.

#### IV. CONCLUSION

An adaptive local binarization algorithm was proposed for text images with complex background. This algorithm is appropriate for document images that are difficult to be recognized correctly directly as grayscale or after using a global thresholding like Otsu. After the applying of different binarization algorithms to the real images with text and

complex background, it is easy to evaluate the results by comparing the recognized texts in images.

In this paper we presented a modified algorithm of the adaptive computing thresholds for local image binarization.

The new approach in this paper is the application of integral images in the Wiener filter in order to compute the mean and the variance in the local processing window. This results in increasing the speed of the thresholding algorithm regardless of the local processing window size. The second new approach in the presented algorithm is the computation of the Sauvola threshold function with integral images in the stage of coarse binarization in combination with the Wiener filter. The proposed algorithm includes a fine adaptive binarization technique that achieves better thresholding compared to Sauvola, resulting in a higher recognition accuracy of the image presented in Fig. 3. As a result all text characters are recognized correctly after applying the described binarization algorithm. An additional advantage of the developed algorithm is the reduced time of the binarization process, compared to the other local thresholding algorithms. The decreased time is close to the fast global binarization schemes like Otsu.

#### ACKNOWLEDGEMENT

This paper was supported by the National Science Fund of the Bulgarian Ministry of Education, Youth and Science (Contract – DDUVU 02/13 – “Public and Private Multimedia Network Throughput Increase by Creating Methods for Assessment, Control and Traffic Optimization”).

#### REFERENCES

- [1] Liang X., Image Binarization using Otsu Method, NLPR-PAL Group, CASIA, 2009.
- [2] V. Vonikakis, I. Andreadis, N. Papamarkos and A. Gasteratos, "Adaptive Document Binarization: A human vision approach", 2nd International Conference on Computer Vision Theory and Applications (VISAPP), Barcelona, 2007.
- [3] J. Sauvola and M. Pietikainen, "Adaptive Document Image Binarization", Pattern Recognition, vol. 33 (2), pp. 225-236, 2000.
- [4] S. Pavlos, E. Kavallieratou and N. Papamarkos, "An Evaluation Technique for Binarization Algorithms", Journal of Universal Computer Science, vol. 14, no. 18, pp. 3011-3030, 2008.
- [5] E. Smith, L. Likforman-Sulem and J. Darbon, "Effect of pre-processing on binarization", Proceedings SPIE Electronic Imaging Document Recognition and Retrieval, vol. 7534, pp. 75340H-75340H-8, 2010.
- [6] K. Derpanis, "Integral image-based representations", Department of Computer Science and Engineering, York University, 2007.
- [7] F. Shafait, D. Keysers and T. M. Breuel, "Efficient Implementation of Local Adaptive Thresholding Techniques Using Integral Images", Proceedings of 15th International Conference on Document Recognition and Retrieval, vol. 6815, pp. 81510, , San Jose, CA, USA, 2008.

# Text Skew Detection using Log-polar Transformation

Darko Brodić<sup>1</sup>, Zoran N. Milivojević<sup>2</sup>, Dragan R. Milivojević<sup>3</sup>

**Abstract** – The paper proposes the method for text skew detection based on log-polar transformation and cross-correlation. The text image is transformed into log-polar domain as well as the control ellipse. Theirs cross-correlation established the cost function. The extraction of the cost function maximum represents the text skew value in the region. The method is characterized by the accuracy and computational time inexpressiveness.

**Keywords** – Document image processing, Log-polar transformation, Text skew.

## I. INTRODUCTION

The printed text is a strongly formed text type with a articulated regularity in shape [1]. Accordingly, the letters are of the similar size and the distance between text lines is generally sufficient. Hence, the spacing between text lines is decent. The orientation of the text lines is similar, which leads to the uniform text skew. These attributes represent the relatively predicted characteristics, which simplify the printed skew identification.

However, the text skew extraction represents a severe problem. It is a consequence of the digitization process. Hence, the text skew occurrence is simply unavoidable. The existence of this phenomena could cause the optical character recognition system failing. Hence, its identification represents one of the crucial steps [2].

Existing methods for the text skew identification can be grouped as follows [3]: projection profiles method, k-nearest neighbor clustering method, Hough transforms method, Radon transforms method, Fourier transformation method, cross-correlation method, and other methods.

This paper recommends a new algorithm based on the interaction of the log-polar transformation and cross-correlation. Firstly, it converts images into log-polar space. Furthermore, two images are cross-correlated in log-polar domain to extract their similarity. As a consequence, cross-correlation function called cost function has been obtained. Its maximum values represent the angle of the text skew estimation. The result gives the fractured line with two text skew values: left and right. These two values represent the new elements in the skew estimation compared to the previously mentioned methods like Hough transform, Radon

transform, vertical projection profiles, etc. The proposed algorithm shows a good skew estimation in the standard resolution. Hence, it contributes to their robustness.

Organization of this paper is as follows. Section 2 describes the proposed algorithms for the estimation of the text skew. Section 3 defines text experiments. Section 4 compares and discusses obtained results. Section 5 makes conclusions.

## II. ALGORITHM

### A. Document Image

Document text image is a product of the image scanning. It is a digital gray-level image, which is represented by matrix  $\mathbf{D}$ . It consists of  $M$  rows,  $N$  columns, and contains the elements which intensity has  $L$  discrete levels of gray.  $L$  is the integer from  $\{0, \dots, 255\}$ ,  $D(i, j) \in \{0, \dots, 255\}$ , where  $i = 1, \dots, M$  and  $j = 1, \dots, N$ . After performing the binarization procedure the image represented by matrix  $\mathbf{D}$  is transformed into binary image  $B(i, j)$ . Its elements are equal to 1 if  $D(i, j) \geq D_{th}(i, j)$ , or to 0 if  $D(i, j) < D_{th}(i, j)$ , where  $D_{th}$  is given by any local binarization method [4]–[5].  $D_{th}$  represents local threshold sensitivity decision value. Currently, document image is given as binary matrix  $\mathbf{B}$  featuring  $M$  rows and  $N$  columns.

### B. Log-polar Transformation

The log-polar transformation is a nonlinear and non-uniform sampling of the spatial domain. Nonlinearity is introduced by polar mapping, while non-uniform sampling is the result of logarithmic scaling [6]. Consider the log-polar coordinate system, where denotes radial distance from the center and denotes angle. For the input binary image  $B(i, j)$ , the center point has been extracted as  $B(m, n)$ . The radius, which ensures the maximum number of pixels to be included within reference circle of the conversion is assigned as  $R$ . Center of the circle is given as  $m = M/2$ , and  $n = N/2$  [6]. Furthermore, the image is converted into polar coordinate system. This way, the input binary image  $B(i, j)$  has been transformed into polar domain  $(r, \theta)$  where [6]:

$$r = \sqrt{(i-m)^2 + (j-n)^2}, 0 \leq r \leq R, \quad (1)$$

and

$$q = \arctan\left(\frac{j-n}{i-m}\right), 0^\circ \leq q \leq 360^\circ. \quad (2)$$

Furthermore, log-polar transform is given as  $(\rho, \theta)$  where:

$$r = \ln r. \quad (3)$$

<sup>1</sup>Darko Brodić is with the University of Belgrade, Technical Faculty in Bor, V.J. 12, 19210 Bor, Serbia, E-mail: dbrodic@tf.bor.ac.rs.

<sup>2</sup>Zoran N. Milivojević is with Technical College Niš, Aleksandra Medvedeva 20, 18000 Niš, Serbia, E-mail: zoran.milivojevic@jotel.co.rs.

<sup>3</sup>Dragan R. Milivojević is with Institute for Mining and Metallurgy, Zelene bulevar bb, 19210 Bor, Serbia.

Applying a polar coordinate transformation to an image maps radial lines in Cartesian space to horizontal lines in the polar coordinate space.

### C. Cross-correlation

Cross-correlation is a measure of similarity of two images. In the discrete form, it is given as [7]:

$$cc(i, j) = B(i, j) \bullet E(i, j) = \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} B(k, l) E(i+k, j+l). \quad (4)$$

However, in our case cross-correlation have to be made between the images in log-polar domain. Hence, eq. (4) should be transformed adequately. Hence, suppose that in the log-polar domain matrices of text image **B** and referent object **E** are marked as **BC** and **EC**. Furthermore,  $cc$  (in log-polar domain) that represents the cross-correlation function can be defined as:

$$cc(q) = C_{coeff} (BC, circshift(EC, q)), \quad (5)$$

where **ECS** is  $circshift(EC, \theta)$  and  $C_{coeff} (BC, circshift(EC, \theta))$  is given as [7]:

$$C_{coeff} = \frac{\sum_r \sum_q (BC_{rq} - \overline{BC})(ECS_{rq} - \overline{ECS})}{\sqrt{\left( \sum_r \sum_q (BC_{rq} - \overline{BC})^2 \right) \left( \sum_r \sum_q (ECS_{rq} - \overline{ECS})^2 \right)}}, \quad (6)$$

If the images are more alike, then the cross-correlation function  $cc(\theta)$  will tend to approach 1.

The identification of the rotation in the spatial domain, i.e image space is a complex task. However, the rotation in the log-polar space is mapped into translation. The translation in the direction of one axis is an easy task to solve. Suppose that a referent object is rotated in the space domain. If it is cross-correlated with the text image for the different angles, then it will be readout as the translation in the log-polar space. The objective is the selection of the referent object. In this paper, the ellipse is selected as a referent object. It is a suitable object because it can overlap text efficiently. However, the ellipse has to be normalized according to the text image dimension. Furthermore, the ellipse is split into left and right half part from the center point of the transformation. This way, those parts of the ellipse are matching with the original image by the cross-correlation. Hence, they establish the left and right skew estimation. Unlike the other methods, the log polar transformation identifies two skews: left and right one. This fact is the advantage of the proposed method.

### D. Algorithm's Steps

The algorithm for the estimation of the text skew based on log-polar transformation is as follows:

1. Text image extraction by the bounding box (text image).

2. Identification of the center point needed for the log-polar transformation.
3. Creation of the binary image with normalized ellipse (ellipse image).
4. Log-polar transformation of the text image.
5. Log-polar transformation of the ellipse image.
6. Cross-correlation of the text image with ellipse image in the log-polar domain.
7. Extraction of the maximum values from the cross-correlation function.
8. Identification of the left and right side skew angle from the center transformation point.

#### Step 1.

The original binary text image **B** is shown in Fig. 1.

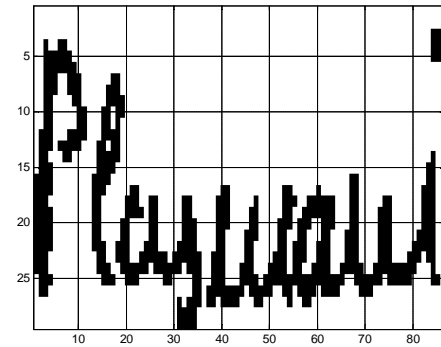


Fig. 1. The original binary text image.

The text is extracted by the bounding box.

#### Step 2.

Typical center point  $B(m, n)$  needed for transformation is extracted according to the pixel density in the center of the bounding box.

#### Step 3.

Furthermore, the binary image with ellipse is created. The size of the ellipse depends on the original text image. It can be said that ellipse is normalized according to the text image size. The binary image with ellipse is shown in Fig. 2.

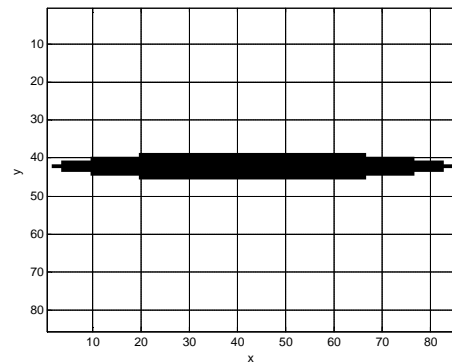


Fig. 2. Ellipse image (ellipse has been normalized according to the text object).

According to the eq. (1)-(3) log-polar transformation of the text and ellipse image is achieved.

**Step 4.**

In Fig. 3, the log-polar transformation of the text image is shown.

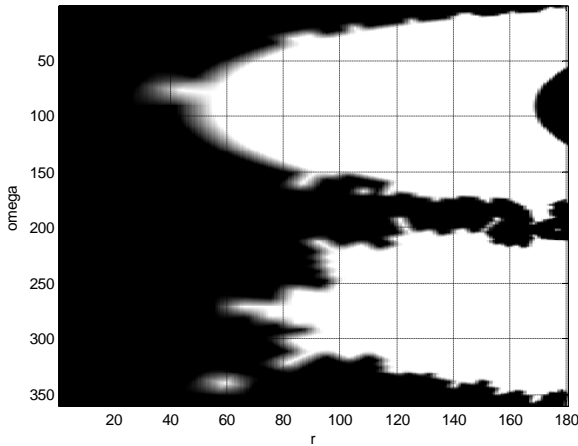


Fig. 3. Log-polar transformation of the original text image.

**Step 5.**

In Fig. 4, the log-polar transformation of the ellipse image is shown.

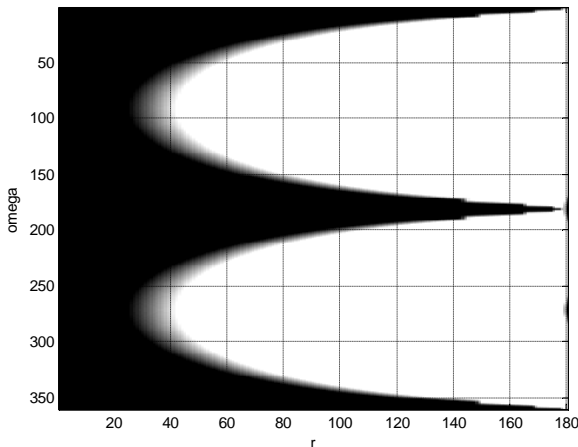


Fig. 4. Log-polar transformation of the ellipse.

**Step 6.**

As a result of the cross-correlation of the text and ellipse image in log-polar domain, a cross-correlation function is obtained. It is so-called cost function. This function is shown in Fig. 5.

**Step 7.**

From Fig. 5 the cost function has two maximums. These maximums represent the two angles of the text skew rate given from the central point of the transformation. This information has been return from the log-polar domain to the spatial image domain.

**Step 8.**

As a result the left and right angle skew line in drawn in the image. This is shown in the Fig. 6.

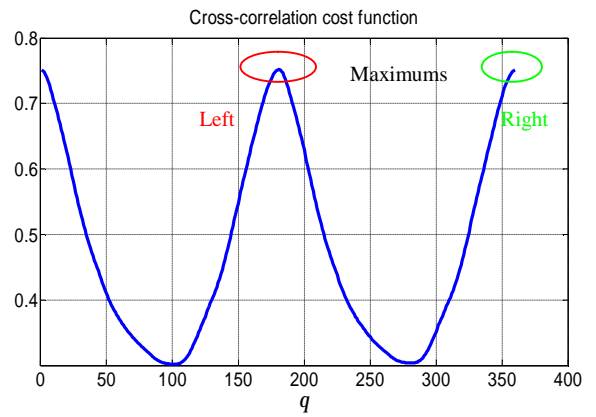


Fig. 5. Cross-correlation of the original text image and ellipse in log polar domain.

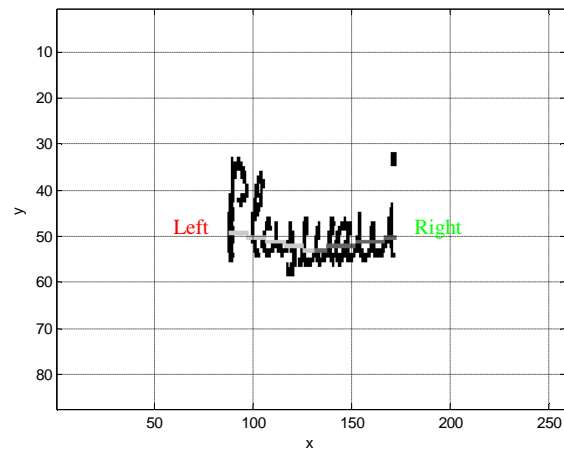


Fig. 6. Enlarged original image with the skew line obtained from the log-polar cross-correlation function.

**III. EXPERIMENTS**

The main goal of the experiments is the evaluation of the algorithm for the text skew estimation. It evaluates the algorithm's performance in the skew tracking domain.

Experiments were performed mostly on the synthetic datasets, which represents the single line of the printed text sample [8].

The test consists of the single line printed text rotated for the angle  $\beta$  from  $0^\circ$  to  $60^\circ$  by the  $5^\circ$  steps around  $x$ -axis [8]. Text sample is shown in Fig 7.

Furthermore, all text samples are given in the standard resolution of 300 dpi. The results are evaluated by the absolute deviation, i.e. error. It is given as:

$$\Delta q_A = |q_A - q_{REF}|, \tag{7}$$

where  $\theta_{REF}$  is the referent skew of the input text sample and  $\theta$  i.e.  $\theta_A$  is the skew of the text sample obtained with a tested algorithms. Furthermore, a relative error (RE) [9] is important for the algorithm evaluation as well. It is given as:

$$RE(q_A) = \frac{\Delta q_A}{q_{REF}} = \frac{|q_A - q_{REF}|}{q_{REF}} \quad (8)$$

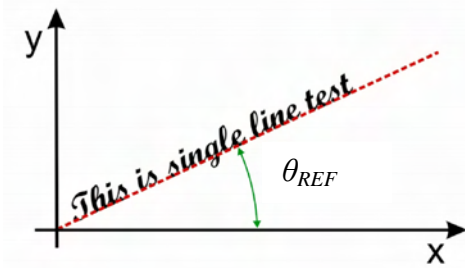


Fig. 7. Enlarged original image with the skew line obtained from the log-polar cross-correlation function.

#### IV. RESULTS AND DISCUSSION

The result of testing is given in Table 1.

TABLE I  
TESTING RESULTS

$\theta_{REF} (^{\circ})$	$\theta_A (^{\circ})$	$\Delta\theta_A (^{\circ})$	$RE(\theta_A)$
0	0	0	-
1	1	0	0.000
2	2	0	0.000
3	3	0	0.000
4	4	0	0.000
5	5	0	0.000
10	10	0	0.000
15	15	0	0.000
20	20	0	0.000
25	25	0	0.000
30	30	0	0.000
35	35	0	0.000
40	40	0	0.000
45	45	0	0.000
50	49	1	0.020
55	56	1	0.018
60	61	1	0.016

The investigated algorithm shows good results in the whole testing angle range. Hence, the presented method is promising in the domain of the accuracy. Furthermore, it is computer time non-intensive.

#### V. CONCLUSION

This paper gives the analysis of the text skew estimation techniques based on the log-polar transformation method. This method estimates the similarity of the text image and ellipse in the log-polar domain. As a result, the cross correlation cost function is obtained. Given method shows good results for the skew estimation of the printed and hand printed text. Hence, it proves its accuracy in the standard

resolution of text images. Furthermore, the method is computer time inexpensive.

The future investigation will be toward the estimation of the handwritten text skew in the unconstrained text.

#### ACKNOWLEDGEMENT

This work is partly supported by the project funded by the Ministry of Education and Science of the Republic of Serbia, No. TR33037.

#### REFERENCES

- [1] D. Brodic, "The Evaluation of the Initial Skew Rate for Printed Text," *Journal of Electrical Engineering - Elektrotechnický časopis*, Vol. 62, No. 3, pp. 142-148, 2011.
- [2] P. Shivakumara, G.H. Kumar, D.S. Guru, P. Nagabhushan, "A Novel Technique for Estimation of Skew in Binary Text Document Images based on Linear Regression Analysis," *Sādhana*, Vol.30, No.1, pp. 69-86, 2005.
- [3] A. Amin, S. Wu, "Robust Skew Detection in Mixed Text/Graphics Documents," *Proceedings of the 8th ICDAR '05*, Seoul, Korea, Vol.1, pp. 247-251, 2005.
- [4] Sauvola L.; Pietikainen M., "Adaptive Document Image Binarization," *Pattern Recognition*, Vol.33, No.2, pp. 225-236, 2000.
- [5] Khasman A., Sekeroglu B., "Document Image Binarisation Using a Supervised Neural Network," *International Journal of Neural Systems*, Vol.18, No.5, pp. 405-418, 2008.
- [6] Mrinal Kanti Bhowmik, Debotosh Bhattacharjee, Mita Nasipuri, Mahantapas Kundu, Dipak Kumar Basu, "Classification of Log-Polar-Visual Eigenfaces using Multilayer Perceptron," *International Journal of Image Processing (IJIP)*, Vol. 4, No. 1, pp. 12-23, 2010.
- [7] Gonzalez R. C., Woods R. E. *Digital Image Processing*, 2nd edn. - Prentice-Hall, 2002.
- [8] D. Brodić, D.R. Milivojević, Z. Milivojević, "Basic Test Framework for the Evaluation of Text Line Segmentation and Text Parameter Extraction," *Sensors*, vol.10, no.5, pp. 5263-5279, 2010.
- [9] V. S. Popov, "Principle of Symmetry and Relative Errors of Instrumentation and Transducers," *Automation and Remote Control*, Vol. 62, No.5, pp. 183-189, 2001.



# Directional Transforms Applicability in Image Coding

Ivo R. Draganov<sup>1</sup>

**Abstract** – In this paper an attempt for review of the most popular directional transforms is made in relation to their applicability in image coding along with some directions for future improvement of their incorporation into complete compression algorithms. Some comparison considerations are given for the DDCT, DWHT, DDWT and others based on image quality, execution time and memory efficiency when applied for image compression. These transforms are considered even more effective if used in combination with some spectral coefficients rearrangement as suggested for further elaboration.

**Keywords** – Image Coding, Directional Transform, DDCT, DWHT, DDWT.

## I. INTRODUCTION

One of the most popular and widely used idea in image coding (compression) is the decorrelation of the intensity (color) values and possibly removing non-significant portions of the transformed data for which the human eye is less sensitive. Such classical techniques are the Discrete Cosine Transform (DCT) used in JPEG compressors and Discrete Wavelet Transform (DWT) in JPEG2000 algorithm [1]. Due to their separability along horizontal and vertical direction it is possible to apply them by using 1D masks reducing the amount of memory needed in one computational pass and reducing the complexity of the code structure, e.g. by using simpler look-up tables, etc.

Nevertheless of these enhancements there are large amount of cases where inside the coded image exist periodic structures with one and the same pattern alternating along an arbitrary direction different from horizontals and verticals, e.g. stripes, slopes, etc. In these cases it is appropriate to apply a transform in the same direction in which the dominant pattern spreads, that is a directional transform.

Two large groups can be defined for the existing directional transforms – the group of the linear orthogonal transforms [7-15, 17-22] and that of the directional wavelet transforms [2-6, 16]. Here a brief review is made for both groups revealing their basic principle, advantages and range of applicability in image coding.

The rest of this paper is organized as the following – in section 2 a description of some of the most popular directional linear orthogonal transforms is given followed by directional wavelet transforms in section 3 and then in section 4 – conclusion is made.

## II. DIRECTIONAL LINEAR ORTHOGONAL TRANSFORMS

One of the simplest 2D-transforms is considered to be the Hadamard transform which is generalized in [10] by using the following jacket matrix:

$$K = \begin{bmatrix} 1 & 1 & \dots & 1 & 1 \\ 1 & * & \dots & * & \pm 1 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & * & \dots & * & \pm 1 \\ 1 & \pm 1 & \dots & \pm 1 & \pm 1 \end{bmatrix}, \quad (1)$$

which later is used by Monadjemi and Moallem [17] for texture classification where the case of sequency-ordered matrix of rank=3 (8x8 size) is applied to extract features from the image. They achieved classification accuracy of 90.5 % against 90.0 % for approach using the Gabor filter and only 77.5 % for the ordinary Walsh-Hadamard transform-like features. More than that, the execution time for the Directional Walsh-Hadamard Transform (DWHT) is more than 10 times less than for the Gabor filter which is considerable difference.

The more advanced transform – DCT – has been extended to its directional form by Zeng and Fu [8, 15] by modifying the weighting factors according to:

$$\hat{\alpha}(i) = \begin{cases} 1/N_k, & i=0 \\ \sqrt{2}/N_k, & i \neq 0 \end{cases}, k=0,1,\dots,2N-2, \quad (2)$$

where  $N$  is the size of the transform vector and  $k$  – the current number of the direction the transform is being applied. They also introduce a DC correction given in [15] by:

$$\Delta DC = \sum_{k=0}^{2N-2} \sqrt{N_k} B^Q(0,k) / \sum_{k=0}^{2N-2} \sqrt{N_k}, \quad (3)$$

where  $B$  denotes a column-vector with the spectral coefficients along the current direction.

Depending on the direction type, e.g. diagonal down-left, diagonal down-right, etc. a certain amount of modes is defined, usually 8. In Fig. 1 some of the possible modes are given by their basic direction. Combining the proper ones inside different images Zeng and Fu [15] obtain difference for the Peak Signal-to-Noise Ratio (PSNR) from 0.5 to 1.5 dB compared to that produce by the JPEG coder. This is true for static images compressed inside the range from 0.1 to 2 bpp where the actual PSNR changes between 31 and 42 dB and the transform block size is fixed to 8x8 pixels. Similar experiment is done for motion pictures where the comparison is made between the H.263 codec and its modified version using the DDCT. The compression ratios achieved are between 0.1 and 6 bpp for the different videos where PSNR

<sup>1</sup>Ivo R. Draganov is with the Faculty of Telecommunications, 8 Kliment Ohridski Blvd., 1000 Sofia, Bulgaria, E-mail: idraganov@tu-sofia.bg

varies between 28 and 45 dB. The advantage of using DDCT is obvious along all the selected range where the PSNR dominance is from 0.1 to 1.5 dB.

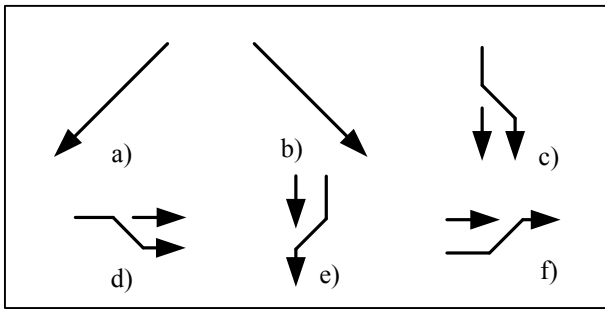


Fig. 1. Different DDCT modes: a) Diagonal Down-Left, b) Diagonal Down-Right, c) Vertical-Right, d) Horizontal-Down, e) Vertical-Left, f) Horizontal-Up

Some simplified form of the DDCT using only diagonal directions is fully described in [7] showing that it is actually optimal orthonormal transform because of the minimized object function maximizing smoothness. Also it is proven that no separability can be achieved here and thus it has no explicit functional form and any fast algorithm.

More productive attempts to speed up the DDCT are developed using lighting-based schemes. In [11] Xu et al. use factorization of 8-point DCT into 35 primary operations formalized in:

$$\vec{Y} = DCT(\vec{X}) = O_{35} \circ O_{34} \circ \dots \circ O_2 \circ O_1(\vec{X}), \quad (4)$$

where  $\vec{X}$  is the input vector and  $\vec{Y}$  - the transformed one. There are two types primary operations – the first represents direct connection between two pixels along a direction of processing and is defined according [14] by:

$$O(X[n_i], X[n_j], \alpha) =$$

$$= \left\{ \vec{X} \leftarrow \begin{bmatrix} 1 & & & & & & & & \\ & 1 & & & & & & & \\ & & 1 & & & & & & \\ & & & \beta & & & & & \\ & & & & 1 & & & & \\ & \gamma & & & & 1 & & & \\ & & & & & & 1 & & \\ & & & & & & & 1 & \\ & & & & & & & & 1 \end{bmatrix} \vec{X} \right\}, \quad (5)$$

where  $\beta = 1$  and  $\gamma = \alpha$ ; the second type is when direct path could not be selected and scale parameter  $\alpha$  is equal to  $\beta$  while  $\gamma = 0$  representing the crossings over mid-lying pixels (Fig. 2).

Experimental results based on that approach are presented in [11, 12, 14] concerning the quality comparison with the JPEG algorithm in the range from 0.5 to 2.5 bpp for lossy compression. The PSNR for static images changes generally from 28 to 42 dB. In virtually all the cases DDCT produces PSNR positive difference from 0.2 to 2.0 dB.

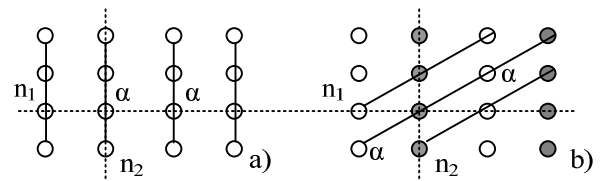


Fig. 2. Two types of pixel connections determining the primary type operations in the Fast DDCT: a) direct and b) non-direct

Dremeau et al. propose in [13] a new compression algorithm based on extended DDCT to rectangular bases and then use bintree segmentation along with dynamic programming for optimal bases selection according to a rate-distortion criterion. They compare their approach with the JPEG and JPEG2000 coders. The test images compression ratios achieved are between 0.01 and 2 bpp while the PSNR is changing between 15 and 44 dB. All along this range their approach proves to be better than both the other with between 1 and 5 dB.

Another approach employing the directional linear transforms is described in [9] for predicting visual residuals in a large number of cases – moving vehicles, people and other objects in city environment, landscapes, etc. The approach is tested with videos for both intra and inter coding and the results show enhancement as for the total PSNR with 1 to 7 dB in comparison to AVC coder. The direction in which the transform is going to be applied is found adaptively in each frame.

Some considerable expanding of the directional transforms is done when considering the challenges met in high resolution image coders design. In coders such as HD Photo overlapping is introduced. In [18] similar technique is undertaken in combination with DDCT and the achieved results are very promising. In a range from 0.1 to 7 bpp the PSNR benefit here is from 1 to 20 dB over the HD Photo itself which is remarkable result.

In [19] a hierarchical class structure is introduced for the I- and B-frames from video processed by 1D directional unified transform along with bidirectional intra prediction. There are 5 classes – from A to E. The total rate gain only for the bidirectional prediction is 3.72 %, for the directional unified transform – it is 5.64 % and the cumulative effect for both is 8.76 %. This scheme is currently used by the Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T.

Some further development of the spatial prediction inside a video frame is done in [20] where such an optimal prediction is looked for along with an adaptive transform by Han et al. Here a hybrid transform is constructed alternating between sinusoidal transform regarding the frequencies and phases of the harmonics which are precisely defined by the boundary effects between adjacent blocks inside the frame. Inter-block correlations are exploited in an effective manner which is proven by the fact that PSNR is higher by 1 to 5 dB over the classical approaches using DCT for video in a range of 0.3 to 1.8 bpp compression ratios.

Some comparable results are presented in [21] and in [22] where sparse orthonormal transforms and direction-adaptive partitioned block transform are introduced.

III. DIRECTIONAL WAVELET TRANSFORMS

The second large group of directional transforms includes the wavelet ones. All of them are based on the classical wavelet transform realized by the lifting implementation (Fig. 3) [6].

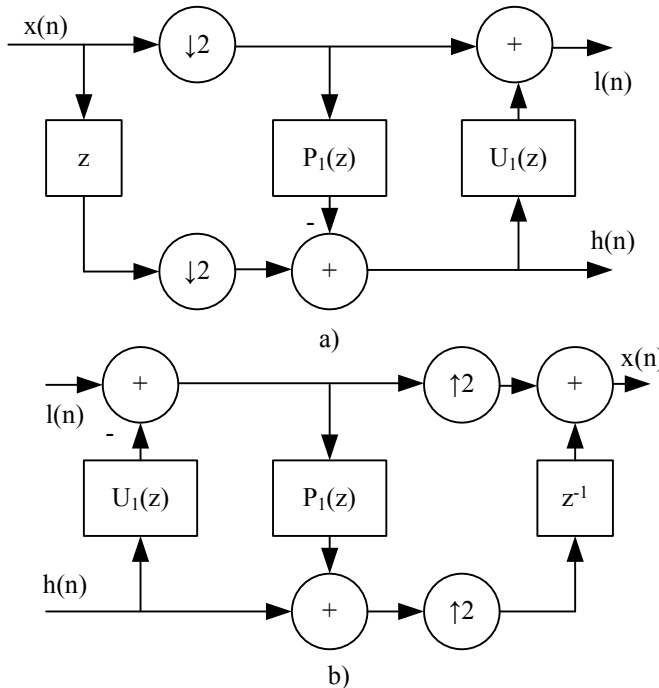


Fig. 3. One-pass link for the lifting scheme of the wavelet transform: a) analysis filter bank and b) synthesis filter bank

The prediction  $P(z)$  and update  $U(z)$  filters are at the base of the analysis and synthesis filter banks of the DWT. One of the most often filters used in this process are the 9/7 biorthogonal wavelet filters described in [6] by:

$$\begin{cases} P_1 = +1.58613(1 + z^{-1}) \\ U_1(z) = -0.05298(1 + z^{+1}) \\ P_2 = -0.88291(1 + z^{-1}) \\ U_2(z) = +0.44350(1 + z^{+1}) \\ s_1 = 1.23017 \quad s_2 = 1/s_1 \end{cases} \quad (6)$$

Here the odd samples are predicted by two neighbouring even pixels which are first averaged and then the result is scaled while the even ones are predicted by averaging of two neighbouring odd pixels of the prediction residual.

In [2] an efficient embedded coding is suggested for medical image compression using the contourlet transform. It is an extensive scheme which incorporates discrete contourlet transform, laplacian pyramid, directional filter bank, some post noise removal steps, optionally extraction of region of

interest (ROI), modified fuzzy C means segmentation, ROI based modified EZW algorithm ending with modified Huffman coding. The authors managed to raise the compression ratio from almost 2 times at relatively constant PSNR around 33 dB for the test MRI images in comparison to the classic EZW approach and the SPIHT algorithm.

Another approach is the robust adaptive directional lifting wavelet transform used for image denoising in [3] by Wang et al. The adaptive directional lifting is at the base of this method combining the directional spatial prediction and the conventional lifting scheme which removes the spatial redundancy leaved by the directional attributes. Additional novelty here is the classification at pixel level and the inter-scale correlation which assure more robustness of the orientation estimation algorithm. The transform itself is applied at pixel level and affecting only those pixels which belong to texture regions of interest. The PSNR for filtered images is increased by 7 dB for some typical for the practice cases.

In [4] some optimization is done for the directional lifting with reduced complexity. While the major disadvantage of the direction-adaptive discrete wavelet transform is the need of exhaustive search for the optimal prediction direction which makes it too complex in contrast to the classical DWT, here lowering of this complexity is aimed. Prediction of the optimal direction is done using gradient-based technique over a formal model of the prediction errors generated by the directional lifting of input wedge image. Stevens et al. [4] proved practically that the prediction step remains very simple and fast and the total complexity reduction has a factor of 11/4 preserving the prediction accuracy. The difference with the original test images and the coded ones with exhaustive search and with the optimized algorithm is about or less than 0.5 dB.

Similar algorithm to that described in [3] is presented in [5] by Chang and Girod. They use local adaptation of the filtering directions to the image content based on directional lifting. The advantage is that energy compaction is more for sharp image features. Additionally anisotropic statistical image model is created for quantifying the gain achieved by adapting the filtering directions. In such a way the authors claim that this algorithm is even more effective than similar ones developed earlier and gain of up to 2.5 dB for the PSNR is achieved. No loss of image structure is reported in the processed pictures.

Kamisli and Lim propose in [6] the directional wavelet transform to be used for prediction residuals in the video coding process. They clearly distinguish the coding of prediction residuals of frame intensities such as the motion compensation residual and the resolution enhancement residual. Special attention is dedicated to the specific characteristics of the different prediction residuals and how they differ from those of the image (frame) itself. Adapting the model for the directional transforms used then produce better results than to unify one and the same algorithm for both. The experiments carried out by Kamisli and Lim indicate that coefficient savings over the classical DWT are between 1 and 40 % with an average close to 30% and considerably more in some cases when DDWT model is adapted to the specific type of residual being processed.

## IV. CONCLUSION

In this paper a brief review is presented of the directional transforms used in image coding. They can be divided in two large groups consisting of linear orthogonal transforms and wavelet approaches respectively. A lot of combinations exist with other popular in practice techniques such as local and holistic decompositions, lifting schemes, groupings based on spatial and time correlation, etc. Depending on the application being developed all of them have their place in practice revealing even broader opportunities for future study and addendum. Some especially perspective approach is considered the hierarchical spectral multistage decomposition of the wavelet spectrum of an image where classical linear orthogonal and wavelet transforms are combined together. Substituting transforms represented of fixed matrix coefficients with adaptive ones such as Karhunen-Loev transform will produce even more efficient algorithms as for the quality and compression ratios of the coded images.

## ACKNOWLEDGEMENT

This paper was supported by the National Science Fund of the Bulgarian Ministry of Education, Youth and Science (Contract – DDVU 02/13 – “Public and Private Multimedia Network Throughput Increase by Creating Methods for Assessment, Control and Traffic Optimization”).

## REFERENCES

- [1] K. R. Rao and P. Yip, *The Transform and Data Compression Handbook*, CRC Press, Boca Raton, USA, 2001.
- [2] M. Tamilarasi and V. Palanisamy, “An Efficient Embedded Coding For Medical Image Compression Using Contourlet Transform”, *European Journal of Scientific Research*, Vol. 49, No. 3, pp. 442-454, 2011.
- [3] X. Wang, G. Shi, Y. Niu, and L. Zhang, “Robust adaptive directional lifting wavelet transform for image denoising”, *IET Image Processing*, Vol. 5, Iss. 3, pp. 249-260, 2011.
- [4] R. Stevens, A. Munteanu, J. Cornelis, and P. Schelkens, “Optimized Directional Lifting with Reduced Complexity”, In *Proceedings of the 16<sup>th</sup> European Signal Processing Conference (EUSIPCO 2008)*, Lausanne, Switzerland, pp. 1-5, August 25-29, 2008.
- [5] C.-L. Chang, and B. Girod, “Direction-Adaptive Discrete Wavelet Transform for Image Compression”, In *IEEE Transactions on Image Processing*, Vol. 16, No. 5, pp. 1289-1302, May 2007.
- [6] F. Kamisli, and J. Lim. “Directional Wavelet Transforms for Prediction Residuals in Video Coding”, In *Proceedings of the 16<sup>th</sup> IEEE International Conference on Image Processing (ICIP)*, Cairo, Egypt, pp. 613-616, November 7-10, 2009.
- [7] I. Selesnick and O. Guleryuz, “A Diagonally-Oriented DCT-Like 2D Block Transform”, *Proc. SPIE*, Vol. 8138, 2011.
- [8] J. Fu and B. Zeng, “Directional Discrete Cosine Transforms: A Theoretical Analysis”, In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2007)*, Honolulu, HI, USA, pp. I-1105 – I-1108, April 15-20, 2007.
- [9] R. Cohen, S. Klomp, A. Vetro, H. Sun, “Direction-Adaptive Transforms for Coding Prediction Residuals”, In *Proceedings of the 17<sup>th</sup> IEEE International Conference on Image Processing*, Hong Kong, PRC, pp. 185-188, Sept. 26-29, 2010.
- [10] K. Horadam, “A Generalised Hadamard Transform”, In *Proceedings of the IEEE International Symposium on Information Theory*, Adelaide, Australia, pp. 1006-1008, Sept. 4-9, 2005.
- [11] H. Xu, J. Xu, and F. Wu, “Lifting-Based Directional Dct-Like Transform for Image Coding”, In *Proceedings of the IEEE International Conference on Image Processing (ICIP 2007)*, Vol. 3, San Antonio, TX, USA, pp. III-185 – III-188, Oct. 16-19, 2007.
- [12] B. Chen, H. Wang, and L. Cheng, “Fast Directional Discrete Cosine Transform for Image Compression”, *Optical Engineering*, Vol. 49, No. 2, pp. 020501-1 – 020501-3, Feb. 2010.
- [13] A. Dr’emeau, C. Herzet, C. Guillemot, and J.-J. Fuchs, “Sparse Optimization with Directional DCT Bases for Image Compression”, In *Proceedings of the IEEE Int’l Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Dallas, TX, USA, pp. 1290-1293, March 14-19, 2010.
- [14] H. Xu, J. Xu, and F. Wu, “Lifting-Based Directional DCT-Like Transform for Image Coding”, In *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 17, No. 10, pp. 1325-1335, October 2007.
- [15] B. Zeng and J. Fu, “Directional Discrete Cosine Transforms for Image Coding”, In *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME 2006)*, Toronto, Ontario, Canada, pp. 721-724, July 9-12 2006.
- [16] A. Monadjemi, *Towards Efficient Texture Classification and Abnormality Detection*, PhD Thesis, University of Bristol, UK, Oct. 2004.
- [17] S. Monadjemi and P. Moallem, “Texture Classification Using a Novel Walsh/Hadamard Transform”, In *Proceedings of the 10<sup>th</sup> WSEAS International Conference on COMPUTERS*, Vouliagmeni, Athens, Greece, pp. 1002-1007, July 13-15, 2006.
- [18] J. Xu, F. Wu, J. Liang, and W. Zhang, “Directional Lapped Transforms for Image Coding”, *IEEE Transactions on Image Processing*, Vol. 19, No. 1, pp. 85-97, January 2010.
- [19] A. Tanizawa, J. Yamaguchi, T. Shiodera, T. Chujoh, and T. Yamakage, “Improvement of Intra Coding by Bidirectional Intra Prediction and 1 Dimensional Directional Unified Transform”, *Input Document to JCT-VC, JCTVC-B042, 2<sup>nd</sup> Meeting: Geneva, CH, July 21-28, 2010.*
- [20] J. Han, A. Saxena, and K. Rose, “Towards Jointly Optimal Spatial Prediction and Adaptive Transform in Video/Image Coding”, In *Proceedings of the IEEE Int’l Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Dallas, TX, USA, pp. 726-729, March 14-19, 2010.
- [21] O. Sezer, O. Harmancıy, and O. Guleryuz, “Sparse Orthonormal Transforms for Image Compression”, In *Proceedings of the 15<sup>th</sup> IEEE International Conference on Image Processing (ICIP 2008)*, San Diego, CA, USA, pp. 149-152, Oct. 12-15, 2008.
- [22] C.-L. Chang, M. Makar, S. Tsai, and B. Girod, “Direction-Adaptive Partitioned Block Transform for Color Image Coding”, In *IEEE Transactions on Image Processing*, Vol. 19, No. 7, pp. 1740-1755, July 2010.

# An algorithm and a program module for calculating the border height of the mass centre of a vessel

Emiliya Koleva<sup>1</sup>, Mariya Nikolova<sup>2</sup>, Mariya Eremieva<sup>3</sup>, Viktoriya Sabeva<sup>4</sup>

*Abstract – This paper presents an algorithm and a program module based on “Method for calculating the stability at moderate and big heeling angles of a vessel”. The program is built to help the command staff of the ship by considerably simplifying the calculations connected to designing the cargo plan and respectively the stability of the vessel. The program is realized on Matlab.*

**Keywords – Static Stability Curve, Dynamic Height of the Mass Centre, Metacentric height, Matlab**

## I. INTRODUCTION

The researches related to the normalization of the stability of a ship meet significant difficulties caused by the ignorance of its hydrodynamic interaction with the water. The empirical-statistic method widely covers in the introduction of certain norms for stability in the design and exploitation of the vessels. The research of the stability could be most clearly made with graphical interpretation of the so called Static Stability Curve (SSC). The direct building of the curve is too hard for the command staff because of the many calculations. Therefore, a new method is presented in “Method for calculating the stability at moderate and big heeling angles of a vessel”. It simplifies the graphic work on creating SSC and initiating a diagram of the dynamic height of the mass centre (DHMC). An algorithm and a program module are created using this method. The program module is realized on Matlab [3].

## II. ALGORITHM

The sequence of the algorithm based on the new method for obtaining the border graphical dependence between the mass centre and the displacement of the ship  $\nabla$  is the following:

1. The values  $(KN_{\theta})_j$  (Fig. 1) are determined from the  $KN$ -curves for each displacement  $\nabla_i$  from  $\nabla_1$  = "empty" to  $\nabla_n$  = "full" ship in  $m^3$  at intervals of  $\delta\nabla = 500m^3$ . (Fig. 2)

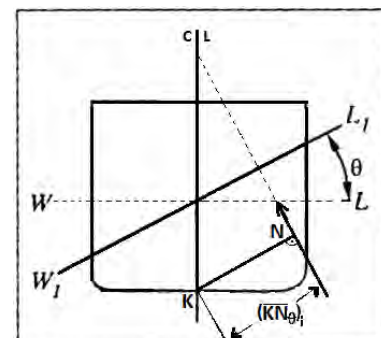


Fig. 1 – Arm of the form

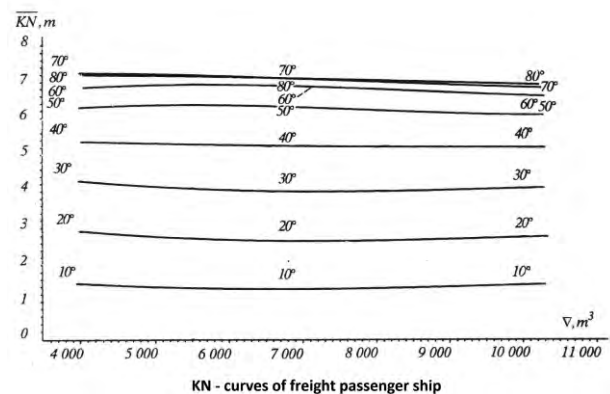


Fig. 2 –  $KN$ -curves of freight passenger ship [1]

2. For each  $\nabla_i$  are taken heights of the mass centre  $\overline{KG}_j$  in realistic exploitation borders for the height of the mass centre  $G$  throughout  $0.1m$ .

3. The arm of the height  $(l_G)_{\theta}$  is calculated for each  $\overline{KG}_j$ , heeling angle  $\theta$  and  $\nabla_i$  by the formula [1,3]

$$[(l_G)_{\theta}]_j = \overline{KG}_j \sin \theta, \tag{1}$$

where  $\theta$  is from  $10^0$  to  $40^0$  throughout  $10^0$ . [3](Fig. 3)

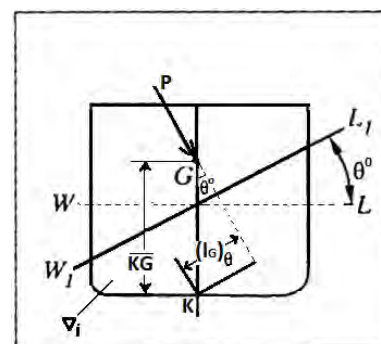


Fig. 3 – Arm of the height

<sup>1</sup>Emiliya Koleva, <sup>2</sup>Mariya Nikolova, <sup>3</sup>Mariya Eremieva, <sup>4</sup>Viktoriya Sabeva are with the Engineering Faculty at Naval Academy “Nikola Yonkov Vaptsarov”, 73 Vasil Drumev str, Varna 9026, Bulgaria, e-mails respectively: [emilia\\_f@abv.bg](mailto:emilia_f@abv.bg), [mnp@abv.bg](mailto:mnp@abv.bg), [viktoriya\\_sabeva@yahoo.com](mailto:viktoriya_sabeva@yahoo.com), [eremirvam@abv.bg](mailto:eremirvam@abv.bg)

4. The arm of the righting moment  $(\overline{GZ})_{i,j}$  is evaluated for each  $\nabla_i$  and each  $\theta = 10^0, 20^0, 30^0, 40^0$ . (Fig. 4) [2]

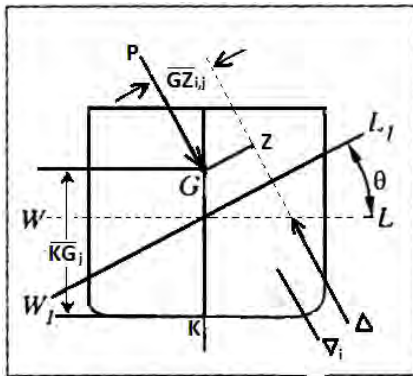


Fig. 4 – Arm of the righting moment

5. The height of the transverse metacentre  $\overline{KM}_i$  is determined depending on the kind of the vessel. (Fig. 5)

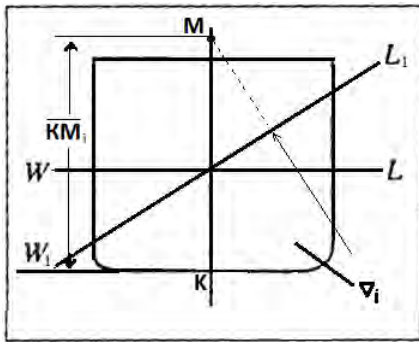


Fig. 5 – Initial transverse metacentre

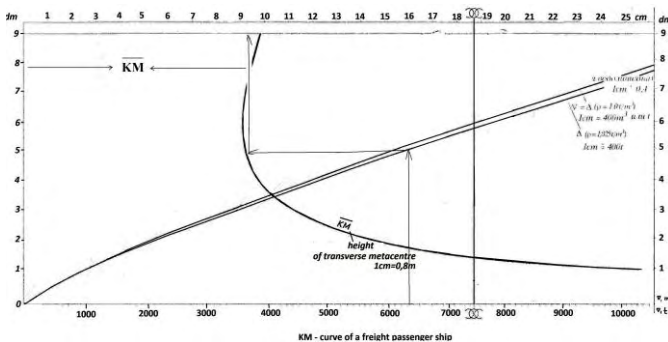


Fig. 6 – Determining the height of the transverse metacentre [1]

6. The transverse metacentric heights  $\overline{G_jM}_i$  are determined as

$$\overline{G_jM}_i = (\overline{KM})_i - \overline{KG}_j \quad (2)$$

for each  $\nabla_i$ .

7. The Static Stability Curve (SSC)

$$(\overline{GZ})_{i,j} = f(\theta). \quad (3)$$

For each  $\nabla_i$  are built  $j$  in number diagrams respectively for each  $l_{j,i}$  i.e. same  $\overline{KN}$  for each angle  $\theta_i$  [2].

To form the initial area from  $0^0 - 10^0$  is necessary to lay vertically the initial metacentric height  $\overline{G_jM}_i$  at  $1rad = 57.3^0$ . The vertex  $\overline{G_jM}_i$  is connected to the origin of the coordinate system. At this area the curve must be tangent of the connection between the vertex and the origin. (Fig. 7)

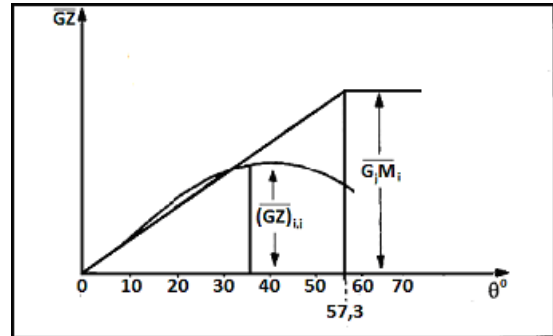


Fig. 7 Static Stability Curve

8. Next step is to calculate the areas  $S_{0^0-30^0}$ ,  $S_{0^0-40^0}$  and  $S_{30^0-40^0}$  and compare them with the ones required in IMO. If the three areas are the same as the required or one of them is the same and the others are bigger  $\overline{KG}_j$  is considered as a solution.[6]

9. The obtained  $\overline{KG}_j$  responds to a specific displacement  $\nabla_i$ . The graph of  $\overline{KG}_j = f(\nabla_i)$  is built for the whole diapason  $\nabla_1 \div \nabla_n$ . The area over the graph is impermissible for the cargo height[1,4].

### III. PROGRAM MODULE BASED ON THE ALGORITHM

For the calculations in this report are used values from the Naval Academy's training ship's documents.

The displacements  $\nabla_i$  are from  $\nabla_1 = 4000m^3$  to  $\nabla_n = 10000m^3$  at intervals of  $\delta\nabla = 500m^3$ . [6]

The heights of the mass centre  $\overline{KG}_j$  from  $\overline{KG}_1 = 6m$  to  $\overline{KG}_m = 9m$  at intervals of  $0.1m$ . [6]

The heights of the transverse metacentre are taken from the  $\overline{KM}$ -curve of a freight passenger ship for each displacement  $\nabla_i$  [6].

The program is created following the steps shown in the flowchart. (Fig. 8)

Finally an acceptable zone for dynamic stability is determined. (Fig.9)

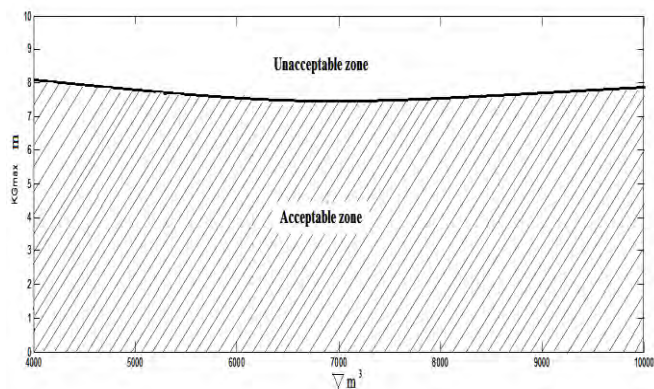


Fig. 9 - Diagram of the dynamic height of the mass centre

## CONCLUSIONS

- 1) Simplifies the evaluation of the stability at moderate and big heeling angles of the ship.
- 2) The program module can be applied to all kinds of ships, for which is necessary to build SSC.
- 3) The graph of the DHMC could be introduced in the ship's documents as reliable evaluation of stability.

## REFERENCES

- [1] M. Zhelyazkov, P. Kaloyanchev, "Book of problems based on ships theory", Military publisher, 2001
- [2] Matlab help
- [3] P. Petkov, V. Petkova, I. Draganov, "Theory of the ships and ships construction", Steno, 2008
- [4] STCW Module 7-Ship Construction, Stability and Damage Control.
- [5] STCW Module 17-Introduction to Ships.
- [6] Rezolution A 749 (18) of IMO-requirements for SSC.

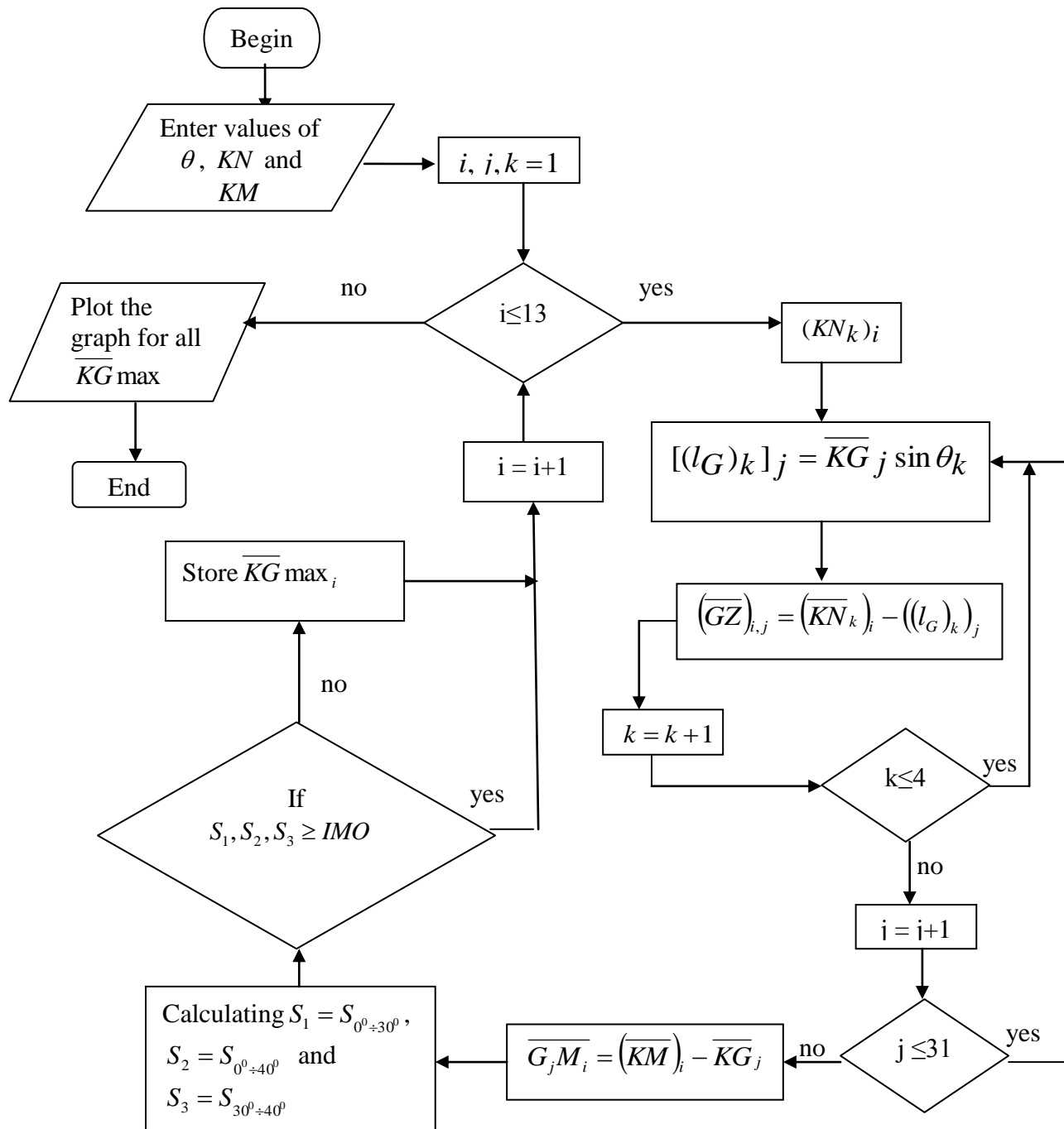


Fig. 8 – Flowchart of the algorithm



# Power consumption analysis of fault tolerant real-time systems

Sandra Djosic<sup>1</sup>, Milun Jevtic<sup>2</sup> and Milunka Damnjanovic<sup>3</sup>

**Abstract** – In this paper we analyze the power consumption of fault tolerant real-time systems. We consider real-time systems where the fault tolerance is achieved running the task affected by a transient fault again using time redundancy. The power consumption analysis is done using one heuristic-based dynamic voltage and frequency scaling algorithm which we designed. The simulation results show that our proposed algorithm can be successfully used for power consumption according to fault tolerance analysis.

**Keywords** – Dynamic voltage and frequency scaling, Fault tolerance, Real-time systems.

## I. INTRODUCTION

Real-time systems (RTSs) have been designed in order to be safe and extremely reliable. They are usually realized as real time systems with the ability of tolerating some faults. A fault-tolerant RTS has to ensure that faults in the system do not lead to a failure. Faults could be transient, permanent or intermittent faults. Among these, the transient faults are much more common than faults of other two types. Transient faults have the feature that they occur and then disappear so fault tolerance can be achieved running the task affected by a transient fault again (i.e. re-executing the task). It means that time redundancy can be used as fault-tolerance techniques by using free slack time in the system schedule to perform recovery executions, [1], [2].

Energy efficiency is also crucial to many real-time systems. Dynamic Voltage and Frequency Scaling (DVFS) is the most popular and widely deployed technique for reducing power consumption of processors [3], [4], [5]. Nowadays, DVFS is a commonly used technique for energy management and is supported by many commercial processors [6].

Fault tolerances through time redundancy as well as energy management through frequency and voltage scaling have been well studied in the context of real-time systems. But simply applying fault recovery techniques and energy minimization techniques one after the other results in inferior quality. These techniques use free slack time and since free slack time is a limited resources, it is obvious that more slack time for DVFS

<sup>1</sup>Sandra Djosic is with the University of Nis, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: [sandra.djosic@elfak.ni.ac.rs](mailto:sandra.djosic@elfak.ni.ac.rs).

<sup>2</sup>Milun Jevtic is with the University of Nis, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: [milun.jevtic@elfak.ni.ac.rs](mailto:milun.jevtic@elfak.ni.ac.rs).

<sup>3</sup>Milunka Damnjanovic is with the University of Nis, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: [milunka.damnjanovic@elfak.ni.ac.rs](mailto:milunka.damnjanovic@elfak.ni.ac.rs).

technique means less time for fault tolerance, and vice versa. Therefore, there is a tradeoff between low energy consumption and high fault-tolerance. In accordance with that, we designed one heuristic-based DVFS algorithm and used it for RTSs analysis. The analysis refers to the power consumption and fault tolerance through time redundancy for different number available operating frequency levels of the processor used in the RTSs.

The rest of the paper is organized as follows. The first part of Section II describes real-time system, power, fault and feasibility models we used in the paper. The second part of Section II introduces our proposed heuristic-based DVFS algorithms. Section III gives the simulation results and finally, Section IV presents our conclusions.

## II. MODELS AND ALGORITHM DESCRIPTION

### A. Models description

For a system model, we assume one uniprocessor RTS with variable processor's operating frequency  $f_j$  ( $j = 1, \dots, m$ ) where  $f_j < f_{j+1}$ . Changing the operating frequency of the processor the voltage also changed and it could be switched between  $m$  values. This system can be used for one real-time task set execution. We assume a set of  $n$  periodic real-time tasks,  $\Gamma = \{\tau_1, \dots, \tau_n\}$  where each tasks are defined by a period  $T_i$ , worst case execution time (WCET)  $C_i$ , deadline  $D_i$  and priority  $p_i$ . We assume that  $D_i \leq T_i$  for  $i = 1, 2, \dots, n$ . The WCET of real-time tasks corresponds to executing the task at the maximum frequency  $f_m$ . For simplicity, we assume that the WCET of a task scales linearly with the processing speed. So, if we scale the operating frequency by a factor  $\alpha$ , then WCET must be scaling by factor  $1/\alpha$ , i.e.

$$C_i(f_j) = C_i(f_m) f_m / f_j.$$

Power consumption of an active processor can be modeled as

$$P_A(f) = P_d(f) + P_{ind},$$

where  $P_d(f)$  and  $P_{ind}$  are frequency dependent power and frequency independent power respectively [7]. Frequency dependent power is

$$P_d(f) = V^2(f) C_{ef} f$$

where  $V$  is supply voltage and it is a function of operating frequency,  $C_{ef}$  is the switch capacitance and  $f$  is the operating frequency. Beside power, for DVFS techniques energy is equally important and it is defined as the integral of power over time.

We assume that faults can occur during execution of any task. We consider transient faults and assume that the consequences of a fault can be eliminated by simple re-execution of the affected task at its original priority level and at its original operating frequency. The re-execution of the

corrupted task must not violate timing constraints of any task in  $\Gamma$ .

For checking the feasibility of fault tolerant real-time task set we use the response time analysis (RTA). In the RTA, the fault-tolerance capability of a RTS is represented by a single parameter,  $T_F$ , which corresponds to minimum time interval between two consecutive faults that the RTS can tolerate. More about RTA can be found in [8], [9]. The basic equation characterize for RTA is Eq. (1).

$$R_i^{n+1} = C_i + \sum_{j \in hp(i)} \left[ \frac{R_j^n}{T_j} \right] C_j + \left[ \frac{R_i^n}{T_F} \right] \max_{j \in hp(i) \cup i} (C_j) \quad (1)$$

With Eq. (1) the response time  $R_i$  of a task  $\tau_i$  could be calculated. This equation has three main addends. The first is WCET  $C_i$  for a task  $\tau_i$ . The second presents interference due to preemption by higher priority tasks. We use  $hp(i)$  to denote the set of tasks with higher priorities than  $i$ ,  $hp(i) = \{\tau_j \in \Gamma \mid p_j > p_i\}$ . The third addend refers to possible faults in the system. If we assume that inter-arrival time between faults is  $T_F$  then

there can be at most  $\left\lceil \frac{R_i}{T_F} \right\rceil$  faults during the response time  $R_i$

of task  $\tau_i$ . Since these faults could occur during the execution of task  $\tau_i$  or any higher priority task which has preempted  $\tau_i$ , each fault may add  $\max_{j \in hp(i) \cup i} (C_j)$  to the response time of

task  $\tau_i$ . So, the third addend in Eq. (1) presents an extra time needed tasks recovery due to faults.

Since  $R_i$  appears on both sides Eq. (1) is recurrence relations which starts with  $R_i^0 = C_i$ . The solution is found when  $R_i^{n+1} = R_i^n$ . If during the iteration process we get that  $R_i^{n+1} > D_i$  then task  $\tau_i$  is infeasible and iteration process must be terminated.

### B. Algorithm description

In order to solve the tradeoff problem between low energy consumption and high fault-tolerance, we propose one heuristic DVFS algorithm. The proposed algorithm has to find appropriate execution frequency for each task, from the real-time tasks set, such that energy consumption is minimal when faults are presence. Fig. 1 gives the algorithm in pseudo code form.

For this purpose we created a heuristic-based algorithm to The RTA is the basic of our proposed algorithm. This analysis is used to guarantee feasibility of real-time tasks set and fault tolerance. The input parameters for the algorithm are:

- frequency  $f_j$  ( $j=1, \dots, m$ ) where  $f_j < f_{j+1}$  and  $m$  is number of frequency levels;
- characteristics for all  $n$  real-time tasks from the set: period  $T_i$ , worst case execution time  $C_i$ , priority  $p_i$  and deadline  $D_i$ , for  $i=1, \dots, n$ ;
- minimum time interval between two consecutive faults  $T_F$ .

Input: operating frequency levels  $f_j$  ( $j=1..m$ ), characteristics for  $n$  real time tasks ( $C_i, D_i, T_i, p_i$ ), fault tolerant constraint ( $T_F$ )

- (1) for each *Task* in *TaskSet* set *Task's\_Freq* to  $f_m$  and set *Task's\_Key* to true;
- (2) repeat step (3) to (7) until there are true *Task's\_Key* in the *TaskSet*;
- (3) for each unlock *Task* in *TaskSet* do
- (4) temporarily set *Task's\_Freq* to *Lower\_Task's\_Freq*;
- (5) if new *TaskSet* is not feasible
- (6) then set *Task's\_Key* to false;
- (7) else calculate  $\Delta Power$  as  $Power(Task's\_Freq) - Power(Lower\_Task's\_Freq)$ ;
- (8) find *Task* with maximum  $\Delta Power$  and set *Task's\_Freq* to *Lower\_Task's\_Freq*;

Output: *TaskSet* with new frequency assigne to each *Task*

Fig. 1. Pseudo code of the proposed algorithm

The algorithm starts with assigning the maximum operating frequency,  $f_m$ , to each real-time task, step (1). Also, at the beginning, all tasks are allowed to change the frequency - we say that all tasks are unlocked. An iteration of the algorithm decreases the frequency of one task for one frequency level. The chosen task is one for which the frequency decrement yields maximum power reduction among all unlocked tasks provided that tasks set remains feasible. To find such task, the algorithm checks all currently unlocked task. For example, frequency index of one unlock task  $\tau_i$  is temporarily decreased for one frequency level, i.e. from  $f_j$  to  $f_{j-1}$ , step (4), and feasibility of task-set is tested using Eq. (1), step (5). If task-set is not feasible,  $\tau_i$  is locked, step (6). Otherwise, if task-set is feasible, the difference between power consumption of  $\tau_i$  at lower ( $f_{j-1}$ ) and higher ( $f_j$ ) frequency is calculated, step (7). Then,  $\tau_i$ 's frequency is changed back to  $f_j$ . After checking all tasks, one that remains unlocked and provides the maximal power reduction is selected, and its frequency index is decremented, step (8). Additionally, the selected task is locked if its new frequency equals 1, i.e. corresponds to the lowest execution frequency,  $f_1$ . After that, the algorithm enters the next iteration. The algorithm finishes when there are no more unlocked tasks. The frequency assignment to each task is algorithm's output. We previously proved the proposed heuristic algorithm and more about that can be found in [10].

### III. SIMULATION RESULTS

We realized simulator based on our proposed heuristic DVFS algorithm. The input parameters of the simulator are number of real-time task and their real-time characteristics: minimum inter-arrival time  $T_i$ , worst case execution time  $C_i$  on maximum operating frequency  $f_m$ , deadline  $D_i$  and priority  $p_i$ . Also, input parameters are processor's voltage and frequency levels and fault constraints  $T_F$ .

TABLE I  
TASKS SET FROM GENERIC AVIONICS PLATFORM

$\tau_i$	$p_i$	$T_i=D_i$ (ms)	$C_i$ (ms)
Nav_Status	1	1000	1
BET_E_Status_Update	2	1000	1
Display_Stat_Update	3	200	3
Display_Keyset	4	200	1
Display_Stores_Update	5	200	1
Nav_Steering_Cmds	6	200	3
Tracking_Target_Upd	7	100	5
Display_Hook_Update	8	80	2
Display_Graphic	9	80	9
Nav_Update	10	59	8

We performed simulations with a number of synthesized real-time task sets and few real-world applications. The characteristics of one real-world application are summarized in Table I. It is a task set taken from the Generic Avionics Platform (GAP) used in [11]. For the processor's frequency levels we used data for Transmeta Crusoe processor from [12]. The relevant parameters for the processor are listed in Table II.

TABLE II  
PROCESSOR FREQUENCIES, VOLTAGES AND POWER

CPU Frequency (MHz)	Voltage (V)	CPU Power (W)
300	1.2	1.3
400	1.225	1.9
533	1.35	3
600	1.5	4.2
667	1.6	5.3

First, we assumed that there were no faults in the system. With this assumption, we used our proposed algorithm to find the appropriate execution frequencies for each real-time task that lead to the maximum energy savings. Fig. 2 shows the simulation results for GAP task set and Transmeta Crusoe processor. We performed simulation for different number of available frequency (voltage) levels. That is represented on the x-axis where 2, 3, 4 and 5 frequency levels include set of frequencies (667MHz, 300MHz), (667MHz, 600MHz, 300MHz), (667MHz, 600MHz, 400MHz, 300MHz), (667MHz, 600MHz, 533MHz, 400MHz, 300MHz) respectively. The y-axis represents the power reduction calculated in percents. This reduction is presented as power saving with respect to the power consumption at maximum frequency.

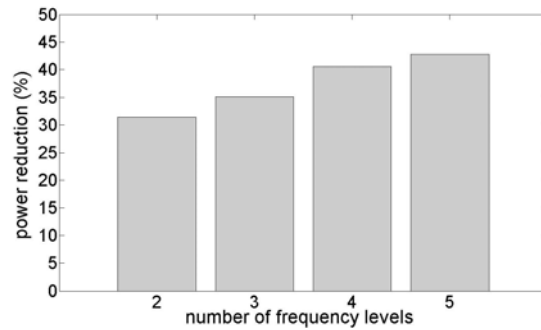


Fig. 2. Power consumption according to number of frequency levels in the absence of faults in the RTS

It can be concluded that power reduction is better when more voltage levels are included. The maximum energy savings is 42.8% for 5 levels and the minimum savings is 31.5% for only 2 frequency levels. The energy reduction is significant even for low number of frequency levels and this clearly shows the effectiveness of our proposed algorithm.

Our next step, in the simulation process, was to consider possible faults appearance in the RTS. This is represented by a single parameter,  $T_F$ , which corresponds to minimum time interval between two consecutive faults that the RTS can tolerate. For the input parameters of the simulator we used the same task set and the same processor. Fig. 3 shows the simulation results for different number of available frequency levels. We used the same sets of frequencies (667MHz, 300MHz), (667MHz, 600MHz, 300MHz), (667MHz, 600MHz, 400MHz, 300MHz), (667MHz, 600MHz, 533MHz, 400MHz, 300MHz). The x-axis of the Fig. 3 represents the ratio of  $T_{Fmax}$  to  $T_F$ .  $T_{Fmax}$  is minimum time interval between two consecutive faults that the task set can tolerate on maximal executing frequency and  $T_F$  is input simulation parameter. This axis represents the normalized  $T_F$  value which is proportional to fault tolerance of the task set. As fault tolerance proportional to time redundancy this axes also could represent free slack time in the systems. The y-axis represents the power saving with respect to the power consumption at maximum frequency calculated in percents.

According to number of available frequency levels the simulation was done for four possible scenarios. All four scenarios indicate the same fact that power reduction leads to less fault tolerance and vice versa. Now due to simulation results, we can better perceive the tradeoff between power consumptions and fault tolerance. For example, let's suppose that power reduction demands are between 40% and 45%. It can be seen, from the Fig. 3, that processor with 4 or 5 frequency levels could fulfill these demands. Also, fault tolerances vary for the given power reduction interval. The best is to choose one with maximal tolerances.

Also, it can be concluded that power reduction is better when more voltage levels are included. With larger number of frequency levels there are more possible task-frequency mapping, so the chance of finding solutions with lower energy becomes higher.

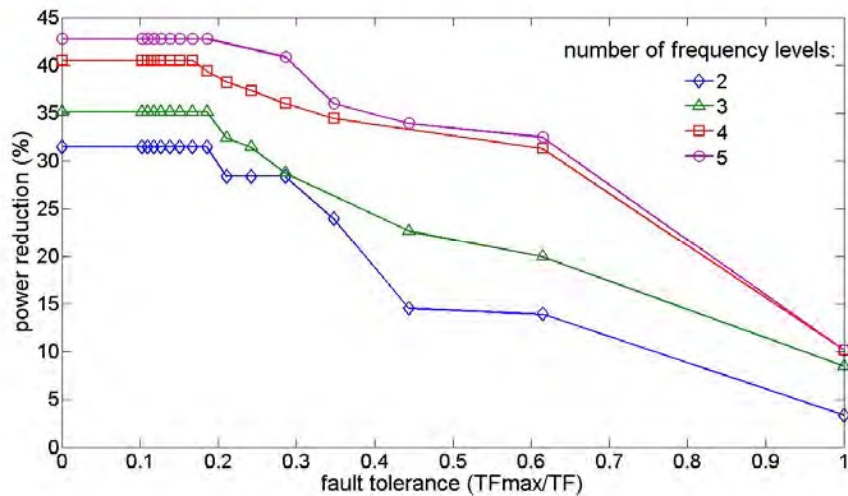


Fig. 3. Power consumption according to fault tolerance for different number of frequency levels

IV. CONCLUSION

In this paper the power consumption of the real-time systems according to fault tolerance through time redundancy analysis are given. The analysis is based on the heuristic DVFS algorithm which we realized. The proposed algorithm offers the possibility to study the trade-off between energy efficiency and fault tolerance for real-time task sets. Generally, this trade-off in discrete systems is NP-hard, so the heuristic-based approach is imposed as possible solution for RTSs analysis.

On the basis of proposed heuristic-based algorithm we realized simulator. Our simulations results show that power reduction is better when more operating frequency levels are included. This is valid for the case of the absence or the presence of the faults, in the RTS. Our opinion is that this simulator could be successfully used in the RTS design process.

ACKNOWLEDGEMENT

This paper is supported by Project Grant III44004 financed by Ministry of Education and Science, Republic of Serbia.

REFERENCES

[1] S. Došić, M. Jevtić, "Scheduling in RTS Using Time Redundancy for System Recovery After Faults", Proceedings of papers, Indel 2004, Banja Luka, pp. 146-149, November 2004.  
 [2] S. Došić, M. Jevtić, M. Damjanović, "Analysis of possibilities to overcome the transient faults in real-time systems with time redundancy", XLVI International scientific conference on information, communication and energy systems and technologies, ICEST 2011, Proceedings of Papers, volume 2, pp 417- 420, Serbia, Niš, June 29 - July 1, 2011.

[3] K. Woonseok, S. Dongkun, Y. Han-Saem, K. Jihong, M. Sang, "Performance Comparison of Dynamic Voltage Scaling Algorithms for Hard Real-Time Systems", Proceedings of the Eighth IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS'02), pp. 219 – 228, 2002.  
 [4] A. S. Ahmadian, M. Hosseingholi, A. Ejlali, "A Control-Theoretic Energy Management for Fault-Tolerant Hard Real-Time Systems", 2010 IEEE International Conference on Computer Design, pp. 173-178, 2010.  
 [5] R. M. Santos, J. Santos, J. D. Orozco, "Power saving and fault-tolerance in real-time critical embedded system", Journal of system Architecture 55, pp. 90-101, 2009.  
 [6] "Intel PXA270 Processor Electrical, Mechanical and Thermal Specification Data sheet", www.phytec.com/pdf/datasheets/PXA270\_DS.pdf, 2005.  
 [7] Z. Dakai, R. Melhem, D. Mosse, "The Effects of Energy Management on Reliability in Real-Time Embedded Systems", Proceedings of the 2004 IEEE/ACM International conference on Computer-aided design, pp. 35-40, 2004.  
 [8] S. Došić, M. Jevtić, "Analysis of Real-Time Systems Timing Constrains", SSSS2010, 3<sup>rd</sup> Small Systems Simulation Symposium 2010, February, 12-14, Faculty of Electronic Engineering, Niš, Serbia, pp 56-60, 2010.  
 [9] G. Lima, A. Burns, "An Optimal Fixed-Priority Assignment Algorithm for Supporting Fault-Tolerant Hard Real-Time Systems", IEEE Transaction on Computers, Vol. 52, No. 10, pp. 1332-1346, October 2003.  
 [10] S. Došić, M. Jevtić, "Dynamic voltage scaling for real-time systems under fault tolerance constraints", accepted for publication on 28<sup>th</sup> International Conference on Microelectronics, MIEL2012.  
 [11] Locke, C. D., Vogel, D. R., Mesler, T. J., "Building a Predictable Avionics Platform in Ada: A Case Study", Proceedings of IEEE Real-Time Systems Symposium, pp. 181–189, 1991.  
 [12] Ying Zhang, Krishnendu Chakrabarty, "Task Feasibility Analysis and Dynamic Voltage Scaling in Fault-Tolerant Real-Time Embedded Systems" Proceedings of Design, Automation and Test in Europe Conference and Exhibition, Vol.2, pp. 1170 – 1175, 2004.

# CHALLENGES OF PERSONALIZATION AND COLLABORATION LEARNING PROCESS BY USING BLOGS

Teodora Bakardjieva<sup>1</sup>, Boyka Gradinarova<sup>2</sup>

**Abstract** – Semantic blogging is a recently emerging technology that attempts to solve the problems of traditional blogging by integrating the features of the semantic web. However, the semantic capabilities currently implemented for semantic blogging are still limited. It is difficult to obtain blog entries relevant to a topic in an aggregated and organized form. A new framework for semantic blogging has been developed capable of organizing results relevant to user requirement. At the centre of this approach are the challenges of personalization and collaboration. Rather than integrating different tools into a centralized system, the idea is to provide the learner with tools and hand over control to him/her to select and use the tools the way the learner deems fit.

**Keywords** – blogging, semantic relations, personalization, collaboration.

## I. INTRODUCTION TO BLOGS

Blog posts or blogs are primarily textual and can vary widely in their content. They can be devoted to politics, news and sharing opinions or dedicated to technical developments. Blog entries are usually maintained in chronological order, but are usually displayed in reverse chronological order. Nardi et al. (2004) identified five reasons why blogs are used:

- to update others on activities and whereabouts;
- to express opinions to influence others;
- to seek others' opinions and feedback;
- to "think by writing";
- to release emotional tension.

The learning specialists Fernette and Brock Eide cited by Will Richardson (2006) identified the Potential benefits of using blogs in educational process as following:

- Can promote critical and analytical thinking.
- Can promote creative, intuitive and associational thinking (creative and associational thinking in relation to blogs being used as brainstorming tool and also as a resource for interlinking, commenting on interlinked ideas).
- Can promote analogical thinking.
- Potential for increased access and exposure to quality information.

<sup>1</sup>Teodora Bakardjieva is Director of Institute of Technology at Varna Free University, гр. Варна 9007, к.к. Чайка, Bulgaria, E-mail: bakardjieva@vfu.bg.

<sup>2</sup>Boyka Gradinarova is with the Department of Computer Sciences and Technology at Technical University of Varna, I Studentska Str. Varna 9000, Bulgaria. E-mail Bgradinarova@tu-varna.bg

- Combination of solitary and social interaction.

**Blogging** is increasingly finding a home in education (both in school and university), as not only does the software remove the technical barriers to writing and publishing online - but the 'journal' format encourages students to keep a record of their thinking over time. Blogs also of course facilitate critical feedback, by letting readers add comments - which could be from teachers, peers or a wider audience.

Students use of blogs are far ranging. A single authored blog can be used to provide a personal space online, to pose questions, publish work in progress, and link to and comment on other web sources. However a blog needn't be limited to a single author - it can mix different kinds of voices, including fellow students, teachers and mentors, or subject specialists. Edu-blogging pioneer Will Richardson (author of the main books devoted to Blogs, Wikis and Podcasts) in 2001 used the blog software Manila (<http://manila.userland.com>) to enable his english literature students to publish a readers guide (<http://weblogs.hcrhs.k12.nj.us/bees>) to the book *The Secret Life of Bees*. Richardson asked the book's author, Sue Monk Kidd, if she would participate by answering questions and commenting on what the students had written - to which she agreed. The result was a truly democratic learning space.

Richardson marked 10 years since his first blog post, a full decade of writing and sharing online. He defines the education reform: "We don't need better, we need different" (Richardson, 2011)

Today's students are immersed in the digital age, but can our educational system keep up? Best-selling author Will Richardson's comprehensive collection of posts from his acclaimed blog (<http://weblogg-ed.com>) outlines the educational reform we must achieve to stay ahead of the curve:

- Project-based learning
- Student-created media that develops critical thinking
- Extending learning beyond the classroom and school hours
- Cooperative and collaborative learning
- Student empowerment and career readiness

The necessary shift will not magically happen, but experts agree that it must happen now. This compilation will inspire educators and parents to engage in the technology their children already embrace, and to take an active role in transforming education to meet the challenges of the digital revolution..

II. OBSERVATIONS AND DISCUSSIONS

Herring et al. (2004) defined three types of blogs: personal journals, “filters” (because they select and provide commentary on information from other websites) and “knowledge logs”. The majority of blogs are the online diary type. Bloggers are interested in reading new information, sharing knowledge and being connected with other users. While blog writers are more extroverted, blog readers are more consumerist.

The use of blogs and semantic blogs has recently been associated with a decentralised form of knowledge management (Cayzer, 2004, Breslin & Decker 2007). Semantic blogging is a technology that builds upon blogging and enriches blog items with metadata. For publishing information such as research publications, there is need of some structure and semantic blogging provides this. Items may be classified using ontologies. Semantic links may exist between items (Cayzer, 2004b). Semantic blogging uses desirable features of both blogging and the semantic web to deal with the challenges of traditional blogging. The semantic web is well suited for incrementally publishing structured and semantically rich information. On the other hand, the easy publishing nature of blogging can boost the semantic web by publishing enough data and resources (Cayzer 2004a; Cayzer, 2004b).

Semantic blogging can help users discover items of interest in blogs. Navigation through the blogosphere can be more flexible and meaningful due to interconnections among various items and topics. Aggregation of useful materials across multiple blogs and the semantic web is possible. Semantic blogging can extend blogging from simple diary browsing to informal knowledge management (Cayzer, 2004b). Publication is easy in semantic blogs too because only some additional metadata data have to be added compared to traditional blogs. The users do not need to put any effort to enjoy the additional features provided. Hence, there is not much effort added in using a semantic blog instead of a conventional one. The rich metadata and semantic structure work behind to give the user the added value experience of semantic blogging. However, the semantic capabilities currently implemented for semantic blogging are still limited. It is difficult to obtain blog entries relevant to a topic in an aggregated and organized form.

There is newly developed framework for semantic blogging capable of organizing results relevant to user requirement (Shakya, 2006). Attempts for implementation of that framework are made at Varna Free University (VFU) to provide more effective navigation and search by exploring semantic relations in blogs.

The system is built upon a blogging infrastructure backed up by an RDF metadata store. The metadata schema enriches the blog entries input. The metadata schema also helps the query processor to search by metadata. Users input queries to the system according to their information requirement. The query processor searches for matching blog entries and

instances in the ontology of the domain of application. Integrated with the ontology is the inference engine, which can deduce implicit relations from the ontology. All the blog entries related to the relevant ontology instances are obtained from the blogontology mapping. The total relevant blog entries obtained are finally organized into an aggregated and navigable collection by the organizer. The system also produces output in RSS format which computers can understand and aggregate.

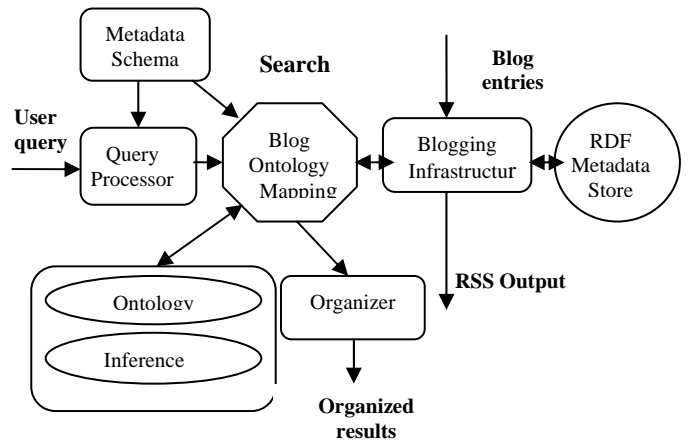


Fig. 1. System architecture of the semantic blogging framework

Some edu-blogs that are used at Varna Free University (Fig. 2, Fig. 3, Fig. 4):



Fig. 2. Edu-blog for the Ranking System for the Bulgarian Universities

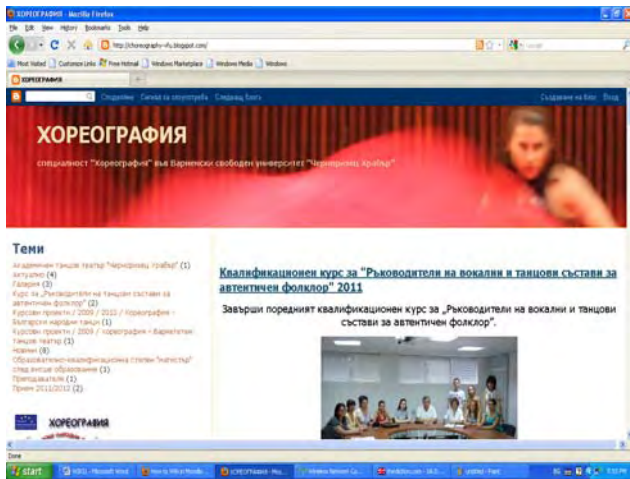


Fig. 3. Edu-blog for Choreography



Fig. 4. Edu-blog for Spatial Design

### III. CONCLUSION TO BLOGS

Teachers are using blogs to provide up-to-date information and commentary on their subject areas, as well as posting questions and assignments and linking to relevant news stories and websites.

They are preferred tools because they have advantage to provide reliable and safe information protected from spam.

Ontology has been introduced to utilize semantic relations, enhanced by inference. Blog entries are mapped to the ontology using language processing. Search results are organized by introducing semantic aggregation. Blog entries are enriched by metadata and an annotation mechanism has also been developed. The framework has been tested and evaluated by implementing a system for the Institute of Technology domain ontology at VFU. Experiments have shown quite good results. Single sample ontology is created for demonstration.

### REFERENCES

- [1] Bates, T. (2011), E-Learning Outlook for 2011 <http://www.tonybates.ca/2011/01/16/e-learning-outlook-for-2011/>
- [2] Bates, Tony (2010), Blackboard acquires Elluminate and Wimba: the end of LMSs? <http://www.tonybates.ca/2010/07/11/blackboard-acquires-illuminate-and-wimba-the-end-of-lms/>
- [3] Boulos, M. N. K., Maramba, I. & Wheeler, S. (2006) Wikis, blogs and podcasts: A new generation of web-based tools for virtual collaborative clinical practice and education, BMC Medical Education, 6(41). <http://www.biomedcentral.com/content/pdf/1472-6920-6-41.pdf> (Last visited January 2011)
- [4] Breslin, J.G, S. Decker (2007), "The Future of Social Networks on the Internet: The Need for Semantics", IEEE Internet Computing, vol. 11, no. 6, pp. 86-90pp.
- [5] Cayzer, S. (2004a). Semantic blogging and decentralized knowledge management. Communications of the ACM, 47(12): 47-52. <http://portal.acm.org/citation.cfm?id=1035164&coll=GUIDE&dl=ACM&CFID=49229987&CFTOKEN=13649580&ret=1#Fulltext>
- [6] Cayzer, S. (2004b). Semantic Blogging: Spreading the Semantic Web Meme. <http://citeseer.ist.psu.edu/698724.html> Last visited July 2011
- [7] Chao, J. (2007) Student project collaboration using Wikis. Proceedings of the 20th Conference on Software Engineering Education and Training (CSEE&T2007), Dublin, Ireland: July 3-5.
- [8] Chatti, M.A., Jarke, M. and Frosch-Wilke, D., (2007). The Future of e-Learning: a Shift to Knowledge Networking and Social Software, International Journal of Knowledge and Learning, IJKL, 3 (4), p. 404-420
- [9] Chatti, M.A., Srirama, S., Kensch, D. & Cao, Y. (2006). Mobile web services for collaborative learning, Proceedings of the 4th International Workshop on Wireless, Mobile and Ubiquitous Technologies in Education (WMUTE 2006), 16-17 November, Athens, Greece.
- [10] Godwin-Jones, R. (2003). Blogs and Wikis: Environments for Online Collaboration. Language Learning and Technology 7(2): 12-16. <http://llt.msu.edu/vol7num2/pdf/emerging.pdf> Last visited July 2011 Stephen Downes, E-Learning 2.0: Notes from Stephen Downes <http://www.downes.ca/http://www.speedofcreativity.org/2006/10/08/elearning-20-notes-from-stephen-downes/>
- [11] Gradinarova, B. (2009). Importance of web 2.0- mediated competence for the educational demands on learners. In Proceedings of the 10th International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing on International Conference on Computer Systems and Technologies (CompSysTech '09),
- [12] Kakizawa, Y. (2007), "In-house use of Web 2.0: Enterprise 2.0", NEC Technical Journal, Vol. 2 No. 2, pp. 46-9.
- [13] Kirkpatrick, M. (2006) The flu wiki: A serious application of new web tools. <http://marshallk.blogspot.com/2005/07/flu-wiki-serious-application-of-new.html> (Last visited March 2011)
- [14] Lamb, B. (2004) Wide open spaces: Wikis, ready or not. EDUCAUSE Review, 39(5) (September/October), 36-48. <http://www.educause.edu/pub/erferm04/erm0452.asp?bhcp=1>

- [15] McAfee, A.P. (2006), "Enterprise 2.0: the dawn of emergent collaboration", MIT Sloan Management Review, Vol. 47 No. 3, pp. 1-28.
- [16] McAfee, A.P. (2006), "Enterprise 2.0: the dawn of emergent collaboration", MIT Sloan Management Review, Vol. 47 No. 3, pp. 1-28.
- [17] McFedries, P.: The Web, Take Two Technically Speaking. IEEE Spectrum, June (2006)
- [18] Mentzas, G., Kafentzis, K. and Georgolios, P. (2007), "Knowledge services on the Semantic Web", Communications of the ACM, Vol. 50 No. 10, pp. 53-8.
- [19] Meyer, M. (2010) Creating an Open Web 2.0 Cloud E-Learning Experience, 2010 Educause Annual Conference <http://www.educause.edu/E2010/Program/SESS124> (Last visited March 2011)
- [20] Mott, J. (2010) [Envisioning the post-LMS era: the Open Learning Network](#) Educause Quarterly, Vol. 33, No. 1
- [21] Nabeth, T., Angehrn, A. and Roda, C. (2002), "Towards personalized, socially aware and active knowledge management systems", in Stanford-Smith, B., Chiozza, E. and Edin, M. (Eds), Proceedings of E-business and E-work – Challenges and Achievements in E-business and E-work, Vol. 2, IOS Press, Amsterdam, pp. 884-91.
- [22] Olson, C. (2006) New tools for learning. <http://faculty.eicc.edu/golson/tools.htm> (Last visited March 2011)
- [23] O'Reilly, T.: What Is Web 2.0. In Internet: <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>
- [24] Oren Eyal, Max Vöölkel, John G. Breslin & Stefan Decker (2006), Semantic Wikis for Personal Knowledge Management, S. Bressan, J. Këung, and R. Wagner (Eds.): DEXA 2006, LNCS 4080, pp. 509–518, 2006, Springer-Verlag Berlin Heidelberg [http://www.johnbreslin.org/files/publications/20060919\\_dexa2006.pdf](http://www.johnbreslin.org/files/publications/20060919_dexa2006.pdf)
- [25] Paul, C. and Schofield, A. (2010) *e-Portfolios for teacher candidates* Vancouver BC: UBC Faculty of Education (retrieved from <http://ctl.ubc.ca/2009/09/22/e-portfolios-for-teacher-candidates/>, December 6, 2010)
- [26] Pettenati M.C., M. Ranieri, Informal learning theories and tools to support knowledge management in distributed CoPs, E. Tomadaki and P. Scott (Eds.): Innovative Approaches for Learning and Knowledge Sharing, EC-TEL 2006 Workshops Proceedings, ISSN 1613-0073, p. 345-355, 2006.
- [27] Razmerita, L. (2005), "Exploiting semantics and user modeling for enhanced user support", electronic version, Proceedings of HCI International 2005
- [28] Rogers, E. (2011) How Web 2.0 Is Influencing E-Learning <http://ezinearticles.com/?HowWeb-2.0-Is-Influencing-E-Learning&id=5968742> (Last visited March 2011)
- [29] Shakya A.(2006), A Semantic Blogging Framework for better Utilization of Information, Thesis submitted in partial fulfillment of the requirements for the degree of Master of Engineering <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.115.7250&rep=rep1&type=pdf> (Last visited July 2011)
- [30] Surowiecki, J. (Ed.) (2005), The Wisdom of the Crowds, Anchor Books, New York, NY
- [31] Taylor, C. (2005). It's a Wiki, Wiki World. Time 165(23): 40-42.
- [32] Vanderwal, T., (2005). Explaining and Showing Broad and Narrow Folksonomies, <http://www.vanderwal.net/random/entrysel.php?blog=1635> (Last visited March 2011)
- [33] Wheeler, S, Kelly P., & Gale, K. (2005) The influence of online problem-based learning on teachers' professional practice and identity. ALT-J 2005, 3(2):125-137.
- [34] Wiki, Matrix - Compare Them All. (2007) Retrieved April 19, 2007 from <http://www.wikimatrix.org/>
- [35] Will, Richardson, Blogs, Wikis, Podcasts and Other Powerful Web Tools for Classrooms <http://weblogg-ed.com/>
- [36] Will, Richardson, Learning on the Blog <http://www.corwin.com/books/Book235915?siteId=corwin-press&subject=C00&q=richardson>



# Implementation of Web 2.0 in the Bitola Museum - Successful Marketing Tool

Pargovski Jove<sup>1</sup>, Irena Ruzin<sup>2</sup> and Aleksandra Lozanovska<sup>3</sup>

**Abstract** – This paper presents the Web 2.0 strategies implemented in the Bitola museum. The project started in 2010 and it has already proved itself as successful marketing tool, resulting with increased number of visits, higher level of communication and friendlier image of the museum in the public.

**Keywords** – Web 2.0, Museum, Social media, facebook

## I. INTRODUCTION

The permanent exhibition in the Bitola museum was renovated and reopened in the end of 2009. It is one of the largest and richest museum collections in Republic of Macedonia, chronologically describing the history of the Bitola region. As soon the exhibit was presented, an appropriate strategy was developed, toward the popularization of the museum and its activities.

The evolution of the internet technologies and introduction of Web 2.0, have opened a whole new realm of opportunities towards the increased operational effectiveness of the museums. Web 2.0 refers to the transition from static HTML web pages to a more dynamic, second generation of the World Wide Web, where people could collaborate and share information online. This introduces a new moment in the current practices of the museum, influencing many changes in the decision making processes, influenced by the constantly monitored public opinion.

Also another advantage of this kind of approach is more “friendlier” image of the museum, resulting with increased number of collaboration projects and cooperation with the public on many levels.

According to many experts, the Bitola museum is the leader in the area of innovative approach, aimed towards popularization of the cultural heritage in Republic of Macedonia.

## II. NI INSTITUTE AND MUSEUM BITOLA

The Institute for protection of cultural monuments and Museum Bitola is a national museum institution which main goal is the protection, systematization, scientific processing

<sup>1</sup>Pargovski Jove is with the Cultural Heritage Protection Office of Republic of Macedonia, Gjuro Gjakovic No 61, Skopje 1000, Macedonia, E-mail: j.pargovski@uzkn.gov.mk.

<sup>2</sup>Irena Ruzin is program director at NI Institute and Museum Bitola, Republic of Macedonia, Kliment Ohridski bb, Bitola 7000 Macedonia, E-mail: irenaruzin@yahoo.com.

<sup>3</sup>Aleksandra Lozanovska is director at Gauss Institute – Bitola, Republic of Macedonia, Pitu Guli 27, Bitola 7000 Macedonia, E-mail: director@gaussinstitute.org.

and presentation of the cultural heritage in the Municipality of Bitola and the nearest regions.

It is the largest museum in South West region of Republic of Macedonia and first museum institution, awarded as “The Best Museum” by Macedonian National Committee of ICOM in 2010. The Municipality of Bitola also awarded the Museum as “The Best Cultural Institution in 2010”.

The main reason for these awards was the new established permanent exhibition, which is among the richest museum collections in Republic of Macedonia.

In the framework of the Institute and Museum of Bitola the following edifices are included:

- The Museum building - Old Barrack,
- Archeological site Heraclea Lyncestis,
- Art gallery (Yeni Mosque),
- Hajdar Kady Mosque,
- Memorial House of Goce Delcev
- Memorial House of Stevan Naumov Stiv
- Memorial museum in village Smilevo
- The Magaza gallery

In 2010 the museum started several innovative projects, with main goal is to increase its operational effectiveness. The digitization program, the modernized website and the social media activities, made this museum the first candidate from Macedonia, nominated for the prestigious EMA Micheletti award in 2012.



Fig. 1. NI Institute and Museum Bitola

## III. MUSEUMS AND WEB 2.0

It is a fact that museums are constantly struggling with the decline in the number of visits and reduced interest in their activities. The evolution of the internet technologies has oversaturated the market with events and knowledge from all over the world, so an evident change can be recorded in the people’s everyday lives. The reduced interest for the museums exhibits and activities can also be contributed to the old

fashioned rigid approach of the cultural heritage presentation and the lack of appropriate marketing strategies.

Many museum activities are created not having any kind of market research or user survey, but based only on the curator’s idea. This also is the case with the museum exhibits, where the curators often have an abundance of objects which they want to present in the exhibit, but there is always the constraint of space and money. In some way this can be categorized as “telling the curator’s version of the story”, which can be good or bad but never complete.

How visitors are experiencing the exhibits and can they contribute in some way, are the factors that are many times neglected.

The modern technologies had in great deal influenced the peoples life’s and also have introduced many new ways of cultural heritage presentation and visitor collaboration.

According to ICOM, “Museum is non-profit, permanent institution in the service of society and its development, open to the public, which acquires, conserves, researches, communicates and exhibits the tangible and intangible heritage of humanity and its environment for the purposes of education, study and enjoyment”. [1]

Having this definition in mind, the museums must use all the available means and technologies in order to be successful in all previously mentioned areas. With the information age, many museums along with all other type of information have gradually moved from real into virtual space.

Many critics have argued that museums need to move from merely supplying information to providing usable knowledge and tools that enable visitors to explore their own ideas and reach their own conclusions [2]. This argument seems especially relevant today, when technology gives individuals access to communication, information gathering and analysis.

The evolution of the internet technologies and introduction of Web 2.0, is opening a whole new realm of opportunities towards the increased operational effectiveness of the museums.



Fig. 2. A tag cloud (a typical Web 2.0 phenomenon in itself) presenting Web 2.0 themes

Web 2.0 is a set of economic, social, and technology trends that collectively form the basis for the next generation of the

Internet—a more mature, distinctive medium characterized by user participation, openness, and network effects. [3]

Basically Web 2.0 refers to the transition from static HTML web pages to a more dynamic, second generation of the World Wide Web, where people could collaborate and share information online.

There have been many successful examples where the implementation of the new techniques, have increased the quality of work of the museums. The Bitola museum is one such example, this kind of strategy, resulted with many benefits for the museum.

#### IV. THE BITOLA MUSEUM WEBSITE

The old website of the Bitola museum existed since 1999 and it was static HTML website offering only one way presentation of the museum.

In 2006 for the first time in Republic of Macedonia, the “Virtual exhibits” were presented, offering new approach towards the presentation of the cultural heritage. The virtual exhibits have proved themselves as a successful tool, resulting with increased number of “online” and “real” visits to the museum.

With the introduction of the new museum exhibit and the necessity of its adequate presentation, new strategy was developed, based on the proven virtual technologies, supported by the use of Web 2.0. The new website of the Bitola museum was presented at the beginning 2010.

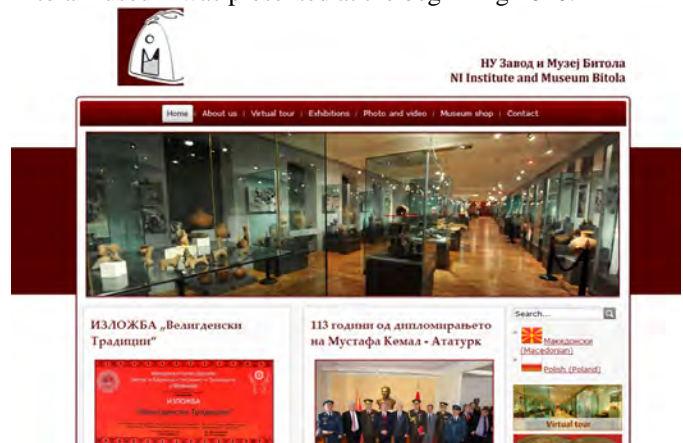


Fig. 3. The web site of the Bitola museum ([www.bitolamuseum.org](http://www.bitolamuseum.org))

According to the website statistics, as predicted most visited sections from the start were the virtual tours, photo galleries, the permanent, temporary exhibitions and the educational section.

It was also noticed that large quantities of website traffic come from the social media, facebook especially. By analysis of these visits, it was determined that users were also significantly contributing to the quality of the published data. People were viewing, sharing and commenting the various sections of the website, constantly contributing towards the quality of the published content.

V. THE BITOLA MUSEUM ON FACEBOOK

Being on the Internet is not just having a website. Almost every company today has a Facebook profile or a Twitter account in order to stay in touch with its clients. With estimated 1.2 billion users, the social media sites represent a significant tool for promotion of the companies, their products and activities.

However, Web 2.0 is not only about social connections or promotion. Constantly evolving, the Web 2.0 technologies provide new smart and interactive formats to show information.

Social media is not just about opening up another marketing channel. It enables customer participation on many levels. For museums, it offers various ways to support the museum experience and even to extend it beyond the actual visit. It is not about virtualizing the museum, but about developing new models of participation and feedback [4]

With more than 830 million users Facebook is the most used social media site, and it was the first choice for social profile for the Bitola museum.

In Republic of Macedonia the number of facebook users in March 2012 was 936 300, or 49,15 % of the total population and 88,55 % of the internet users [5]. According to Social Bakers, more than 50% of the facebook users are under the age of 34, which is target group for many museums. It is known fact that the younger population is spending less time visiting the museums, so new approaches must be developed in order to popularize the cultural heritage.

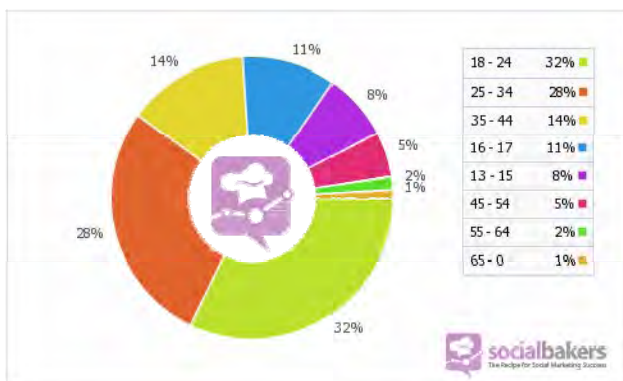


Fig. 4. User age distribution on Facebook in Macedonia

The facebook profile of the Bitola museum was created the same time when the website was launched. At first it was used as a tool for boosting the website traffic, but the preliminary results uncovered its true potential.

Small test was conducted, where one photo gallery from the website was shared as external link, compared to the same photos from the gallery uploaded and shared as album on facebook. The first approach resulted in increased number of visits on the website but with no comments. The facebook album was consecutively shared, but also individual images were tagged and commented. By analysis of the facebook album valuable information was gathered concerning the content of the photos, but also on the visitor's behavior. For example, some photos were shared and commented more than others, meaning they were more interesting to the public.

Soon after all the galleries and videos were transferred to the facebook account, which made this account very popular and generated large circle of friends in a short time. This is also another advantage of the use of social media, creation of circle of friends with similar interests. The online friends are not limited by the place and time barrier. For example city like Bitola has a very large scientific and nonscientific community outside its borders, holding large quantity of knowledge and assets in different areas.

Also through the analysis of the comments, it was determined that in some areas of interest a large quantity information was contained in the "normal" users also. Either in form of historical fact they have witnessed, some person they knew or a story they have heard.

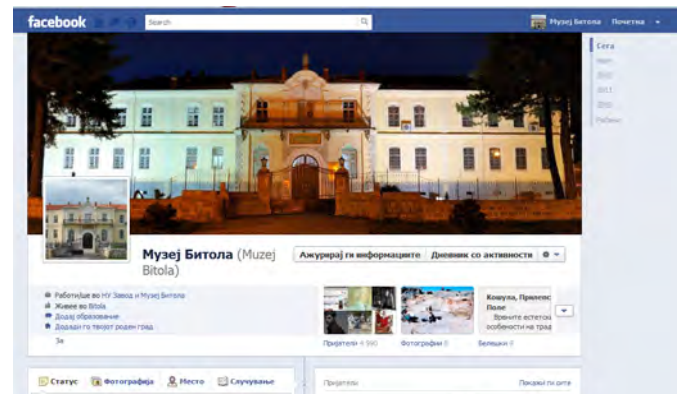


Fig. 5. Facebook profile of the Bitola museum

Using the sport as always interesting topic, some photos from the early days of the local soccer club "Pelister" were published. The museum had those photos but with only small legend describing their date and the clubs that played. Soon after, some of the players or their close relatives and friends were tagged, and in very short time the images were completely described, involving all kind of information. Even greater asset was the generated circle of friends tagged on the photos, which can be used in future researches.



Fig. 6. Donated photo from the online friends

Figure 6 is good example of the benefits of using Web 2.0 strategies, since it is photo donated to the museum by facebook friend. The photo also carries the facebook tags of many members, friends and relatives of the team and complete list of names of all team members.

The facebook approach proved as complete success and soon after new quizzes and educational games were

developed, which embraced the user interactivity with the museum. In some way this created more “friendly” image of the museum, resulting with more users donating images to the museum.

During the campaign “Bitola then and now”, more than thousand old photos were collected by the museum from the online friends. More than 70 % of these photos were new for the museum and as a result of this campaign temporary exhibition was organized in occasion of the day of liberation of Bitola – 4<sup>th</sup> November.

Trough the online communication the museum gathered a large quantity of new information and managed significantly to increase its inventory trough donations and acquisitions.

The online friends have accepted with great pleasure this new approach and many of them have contacted the museum concerning different matters. Based on this communication a large quantity of new ideas have evolved, which are constantly implemented by the museum. According to many experts, the Bitola museum is the leader in the area of innovative approach in Republic of Macedonia

The number of visitors is increased thanks to the different types of educational and interactive programs for all categories and excellence promotion in the media and International fairs.

## VI. EDUCATIONAL WORKSHOP : “MUSEUM DETECTIVE”

The museum of Bitola has significant experience in the organization of educational games and activities for the younger audience. In 2011 new project was conducted, aimed towards the education and familiarization of the museum by the young audience. Those educational workshops are part of long-term project “Learning about us”

The main goal of the “Museum detective“ was to provoke the participants to be interactive using different types of techniques and gain knowledge about the cultural heritage (during the “game”).

Before the task all participants attended lectures according to their age and the possibilities for further research were presented to them. They were divided in groups and every group had received a task to locate specific object from the museum exhibition. The museum collection of the Bitola museum is among the richest in Republic of Macedonia with abundance of objects. Having in mind the competition aspect of the game, it was determined later that the participants who did pre-game online research, were more successful than their competitors. A large number of participants were involved in these workshops and even greater interest is recorded for the second phase planned for May 2012.

This project resulted with some interesting conclusions, or expressed trough statistics:

- In the following 4 months 35 % more primary and secondary school visits were recorded in the museum, than the same period in 2010
- 78 % of the participants have visited the museum before the workshops to learn more
- 92 % have used the museum web site before the workshops

- 92 % wanted to continue with workshops and to be involved for the next year
- All of the participants had facebook profile and more than 70 % have shared their museum activity on their profile
- More than 70 % of schools asked for extra workshop on other subject (art, geography...)
- More than 80 % of the participants have never been in a museum before the workshops
- More than 40 % of the participants, claimed that their close relative or friend have visited the museum on their suggestion after the workshop.

After the workshop a meeting was held with the participants, where the mutual experience was shared. Many of the participants said they liked the Web 2.0 activities implemented by the museum and today some of them are also among the most active participants and supporters. According to them, “The Bitola museum represents an excellent place for cultural heritage education and sometimes fun by participating in the online quizzes”.

## VII. CONCLUSION

It is evident that the new internet technologies have changed the people’s everyday lives. If the museums want to be successful in their work, they are left with only two choices: to resist the changes and have only minor individual efforts in this area, or to embrace the new technologies and make the required organizational change. The benefits of the second approach are evident in many fields of the museum work and significantly increase the quality of work in the museums.

## REFERENCES

- [1] <http://icom.museum>, International Council of Museums.
- [2] Lynda Kelly, Ph.D., “How WEB 2.0 Is Changing the Nature of Museum Work”, [http://museumstudies.si.edu/webcast\\_052110.html](http://museumstudies.si.edu/webcast_052110.html).
- [3] John Musser, Tim O’Reilly & the O’Reilly Radar Team “Web 2.0 - Principles and Best Practices”
- [4] Vogelsang, A. and B. Minder. “Audience+: A Holistic Approach to Developing Social Media Guidelines for Swiss Museums. In J. Trant and D. Bearman (eds).” Museums and the Web 2011: Proceedings. Toronto: Archives & Museum Informatics. Published March 31, 2011.
- [5] <http://www.socialbakers.com>

# Attacking the cloud

Vlad Andrei Poenaru<sup>1</sup>, George Suciuc<sup>1</sup>, Cristian George Cernat<sup>1</sup>, Gyorgy Todoran<sup>1</sup> and Traian Lucian Militaru<sup>1</sup>

**Abstract** – Cloud is emerging as a new technology for serving web applications. There is little literature on how DDOS affects clouds and how other types of attacks fare versus a well designed cloud architecture. Is cloud resilient enough to be a solution or normal web architectures with good caching and reverse proxy is going to be better always? This is what paper is trying to answer by testing two architectures, SlapOS and an architecture with nginx in front and a few web servers behind it against different types of attacks DDOS, slowloris, RA flood attacks.

**Keywords** – cloud, DDOS, cloud architecture, flood attacks.

## I. INTRODUCTION

In this paper we present a study about distributed denial of service attacks (DDOS) in open source cloud platform SlapOS, the first open source operating system for Distributed Cloud Computing. This will include writing security testing scripts, collecting results and automating scripts for improving security of software deployment and configuration on cloud nodes.

We develop a test platform for cloud computing and use it as a case study for testing and monitoring different security threats. We use different types of attacks and monitor important information such as CPU load, number of processes and intrusion level from installed sensors. The sensors will transmit intrusion detection data from our cloud platform in real-time, display it in our web-based visualization application and get detailed recommendations when and where security threats did occur - resulting in optimized automating patching.

Also we will introduce in this article SlapOS, the first open source operating system for Distributed Cloud Computing. SlapOS is based on a grid computing daemon called slapgrid which is capable of installing any software on a PC and instantiate any number of processes of potentially infinite duration of any installed software. Slapgrid daemon receives requests from a central scheduler the SlapOS Master which collects back accounting information from each process. SlapOS Master follows an Enterprise Resource Planning (ERP) model to handle at the same time process allocation optimization and billing. SLAP stands for “Simple Language for Accounting and Provisioning”.

This structure has been implemented for cloud-based automation of ERP and CRM software for small businesses and aspects are under development under the framework of

<sup>1</sup>The authors are with the Faculty of Electronics, Telecommunications and Information Technology at Politehnica University of Bucharest, Bd. Iuliu Maniu, nr. 1-3, Bucharest 060042, Romania, E-mails: vlad.wing@gmail.com, george@beia.ro, cernatcristi@gmail.com, toloran.gyorgy@gmail.com, gelmosro@yahoo.com.

the European research project “Cloud Consulting” [1]. We will use our platform hosted on several servers running Ubuntu Linux – Apache – MySQL template with current software release. On our cloud testing environment we provide the platform for processing information from hundreds different sensors, enabling the analysis of security data through a large sample of logs.

We demonstrate that open source cloud platforms are well-developed and mature technologies offering a secure environment for deploying a growing number of applications.

## II. PROBLEM FORMULATION

SlapOS is an open source Cloud Operating system which was inspired by recent research in Grid Computing and in particular by BonjourGrid [2] a meta Desktop Grid middleware for the coordination of multiple instances of Desktop Grid middleware. It is based on the motto that “everything is a process”. SlapOS is now an OW2 project. Fig. 1 shows the current architecture.

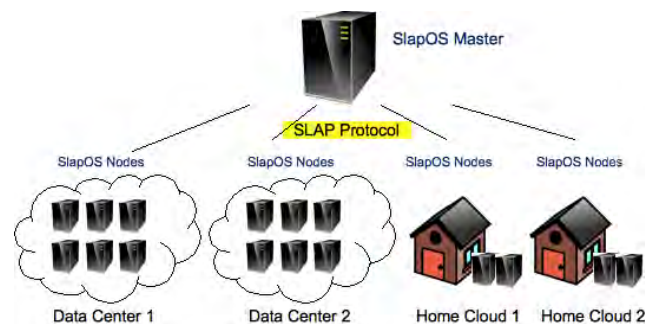


Fig. 1. The SlapOS Architecture

SlapOS defines two types of servers: SlapOS Nodes and SlapOS Master. SlapOS Nodes can be installed inside data centers or at home. Their role is to install software and run processes. SlapOS Master acts as a central directory of all SlapOS Nodes, knowing where each SlapOS Node is located and which software can be installed on each node. The role of SlapOS Master is to allocate processes to SlapOS Nodes.

SlapOS Nodes and SlapOS Master exchange are interconnected through the HTTP and XML based SLAP protocol. SlapOS Master sends to each SlapOS Node a description of which software should be installed and executed. Each SlapOS Node sends to SlapOS Master a description of how much resources were used during a given period of time for accounting and billing purpose.

From a user point of view, SlapOS Node looks like an online shop for Cloud Computing resources. The user connects to SlapOS Master through a simplified front end,

selects which software he or she needs. SlapOS Master then allocates the software onto a SlapOS Node and provides the connection information to the user. The allocated software can be of any type: virtual machine, database server, application server, web cache front end, etc.

From a developer point of view, as seen in Fig. 2, SlapOS is a simple and universal API to create instances of any software daemon through a programmatic interface.



Fig. 2. An example of SlapOS front-end

A simple code allows a developer to request a new instance of a memcache server by invoking the request method of SlapOS API. Memcache [3] is a widely adopted key-value store protocol which is used to cache values in large scale web infrastructure. It is usually installed and configured by system administrators using packaging systems such RPM or DEB. In this example, a single method call does in a few seconds what a human system administrator would have done in few minutes at best.

### III. PROBLEM SOLUTION

SlapOS is implemented as an extension of widely adopted open source software: GNU/Linux, Buildout [4] and Supervisor [5] and as depicted on Fig. 3. The only new software introduced by SlapOS is Slapgrid, a daemon in charge of implementing the SLAP protocol on each SlapOS Node.

Each time slapgrid receives a request from SlapOS master to install a software, it downloads a description of that software in the form of so-called buildout profile. It then runs the buildout bootstrap process to install the software. Buildout is a Python-based build system for creating, assembling and deploying applications from multiple parts, some of which may be non-Python-based. Buildout can be used to build C, C++, ruby, java, perl, etc. software on Linux, MacOS, Windows, etc.

Buildout can either build applications by downloading their source code from source repositories (subversion, git, mercurial, etc.) or by downloading binaries from package repositories (rpm, deb, eggs, gems, war, etc.). Buildout excels in particular at building applications in a way which is

operating system agnostic and to automate application configuration process in a reproducible way.

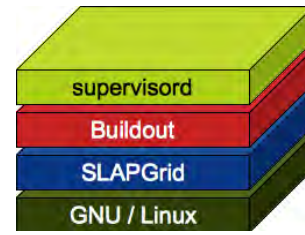


Fig. 3. The SlapOS kernel

Each time slapgrid receives a request from SlapOS master to run a software as a new process, it calls first buildout to create all configuration files for that process then delegates to supervisor the execution of the process. Supervisor is a client/server system that allows its users to monitor and control a number of processes on UNIX-like operating systems. It provides a higher abstraction and flexibility than traditional sysinit.

After some time, a typical SlapOS Node will include multiple software applications and, for each software application, multiple instances, each of which running in a different process. For example, both Mediawiki and OS Commerce could be installed onto the same SlapOS Node, with six instances of each being run as processes. By running software instances as processes, rather than by creating a virtual machine for each software instance as one would do with Amazon EC2 [6], SlapOS is able to use hardware resources and RAM in particular more efficiently.

SlapOS Master runs ERP5 Cloud Engine, a version of ERP5 open source ERP capable of allocating processes in relation with accounting and billing rules. Initial versions of SlapOS Master were installed and configured by human. Newer versions of SlapOS Master are implemented themselves as SlapOS Nodes, in a completely reflexive ways. A SlapOS Master can thus allocate a SlapOS Master which in turn can allocate another SlapOS Master, etc.:

After running security testing scripts and collecting results, as shown in Table 1, we conclude that our open source cloud platform delivers better performance in attacks against it.

TABLE I

Architecture	SlapOS cloud	Nginx Web
CPU load		
(1) DDOS	(1) 85%	(1) 95%
(2) slowloris	(2) 90%	(2) 99%
(3) RA flood attacks	(3) 60%	(3) 80%
Number of processes (slowloris)	200K	14K
Exploits detected	5238/5545	5211/ 5545

#### IV. CONCLUSION

SlapOS can be described as a cloud operating system in which “everything is a process” unlike Unix in which “everything is a file”. If one has to manage thousands of servers with thousands of processes, hundred different applications in multiple different releases or versions, SlapOS can help you a lot by making the whole security management process well specified, automated and under control.

Therefore cloud security is shared with the processes of the applications running on the nodes.

The second result with SlapOS is that the best way to create a reliable and secure cloud computing system is to follow the original principles of the Internet: distribution and simplicity.

Our system can also help keeping track of exploit development, optimize patching for zero-days threats and to produce log auditing to improve security risk management.

As future work we envision an early warning system of cloud attacks that applies intrusion prevention measures based on sensor information from different partitions on the distributed nodes.

#### ACKNOWLEDGEMENT

This paper is presented as part of the project “*Valorificarea capitalului uman din cercetare prin burse doctorale (ValueDoc)*” Project co-financed from the European Social Fund through POSDRU, financing contract POSDRU/107/1.5/S/76909 and part of the project “Cloud Consulting”.

#### REFERENCES

- [1] George Suciu, Octavian Fratu, Simona Halunga, Cristian George Cernat, Vlad Andrei Poenaru, Victor Suciu, “Cloud Consulting: ERP and Communication Application Integration in Open Source Cloud Systems”, 19th Telecommunications Forum - TELFOR 2011, IEEE Communications Society, pp. 578-581, 2011
- [2] Heithem Abbes, Christophe C´erin, and Mohamed Jemni. Bonjourgrid as a decentralised job scheduler. In APSCC 08. Proceedings of the 2008 IEEE Asia-Pacific Services Computing Conference, pages 89–94, Washington, DC, USA, 2008. IEEE Computer Society.
- [3] Memcached: a free and open source, high-performance, distributed memory object caching system. <http://memcached.org/>
- [4] Buildout - software build system reloaded <http://www.buildout.org/>
- [5] Supervisor: A Process Control System <http://supervisord.org/>
- [6] "The impact of virtualization on network performance of Amazon EC2 data center", Tze Ng, Guohui Wang, IEEE INFOCOM 2010 - 029th IEEE International Conference on Computer Communications, Vol. 29, no. 01, March 2010

# An Implementation of SMS Communication with Patients in a Medical Information System

Ivica Marković<sup>1</sup>, Aleksandar Milenković<sup>2</sup> and Dragan Janković<sup>3</sup>

**Abstract** – In order to achieve high quality in providing healthcare services an appropriate communication between staff and patients is required. For a large majority of patients SMS communication with a healthcare institution is a good solution because it uses simple and widely adopted technology. On the healthcare institution side, automating this communication can significantly reduce amount of work which needs to be done by the personnel. Therefore a carefully designed IT solution for this purpose can help improving certain processes in a healthcare institution.

**Keywords** – Medical information system (MIS), SMS communication, Patients scheduling.

## I. INTRODUCTION

SMS (abbreviated from Short Message Service) stands for a communication protocol which enables mobile phones users to send and receive short textual messages [1]. There are many different standards for exchange of textual messages and they are continuously being developed and improved by adding new features (e.g. sending and receiving formatted text, images, audio and video). Nevertheless, in the context of this paper, text messaging typically refers to SMS, with its 160 ASCII characters limitation.

In this paper we firstly discuss the advantages of using SMS in communication between a healthcare institution and the patients by using SMS protocol. Then we analyze architecture and implementation of our system for automated sending and receiving of textual messages. Finally we describe one typical use case of our system.

It is important to notice that our system for automated sending and receiving of textual messages is implemented as a part of medical information system Medis.NET. Medis.NET is being developed at Faculty of Electronic Engineering in Niš and more details about the system can be found in [2]. Anyway, system for automated sending and receiving of textual messages is a complete system itself and can be simply used in conjunction with other medical information systems as well as with other information systems which can be improved by adding SMS communication with their users.

<sup>1</sup>Ivica Marković is with the University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, P.O. Box 73 18000 Niš, Serbia, E-mail: ivica.markovic@elfak.ni.ac.rs.

<sup>2</sup>Aleksandar Milenković is with the University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, P.O. Box 73 18000 Niš, Serbia, E-mail: aleksandar.milenkovic@elfak.ni.ac.rs.

<sup>3</sup>Dragan Janković is with the University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, P.O. Box 73 18000 Niš, Serbia, E-mail: dragan.jankovic@elfak.ni.ac.rs.

## II. SMS COMMUNICATION – CONS AND PROS

We must agree that SMS is, in terms of information technology, relatively old standard (with the first officially transferred message in 1992.) [1]. It is also a fairly limited technology allowing users to transfer only 160 ASCII characters. Actually most of modern mobile devices do allow their users to use different character sets and multi-part messages, but it is a good practice when designing a system to keep in mind the worst case scenario. And there are many more characteristics of SMS which encourage us to use this kind of communication integrated with a healthcare information system. Here are listed some of them.

### A. Number of devices which support SMS

According to data from [3], there are less than 2 billions of Internet users and more than 5 billions of mobile phone users in the world registered in 2010. It is further estimated that 98% of those phones have text messaging capabilities.

### B. Geographic availability of mobile networks

The fact is that mobile phone networks cover significantly larger areas comparing to fixed phone networks and cable or wireless Internet. Although Internet coverage is constantly growing, the advantage on the SMS side comes from its minimal demands regarding the quality of the network. This is especially important in areas with low population density where Internet service providers do not have economic interests to build their networks. In such areas quality of mobile phone network is often below the level required for mobile phone conversation and SMS remains as the only one way of communication.

### C. Simplicity of usage

The most of mobile phone users agree that it is very simple to type and send/receive a short textual message. Also, for a number of these users it is too complicated to use Internet or send/receive an e-mail.

### D. Immediacy of action and response

Most of the mobile users keep their mobile devices always by themselves and turned on so the users can receive important information with the least possible delay and act accordingly. This is an advantage of SMS communication over e-mail and web where most of the users are not always



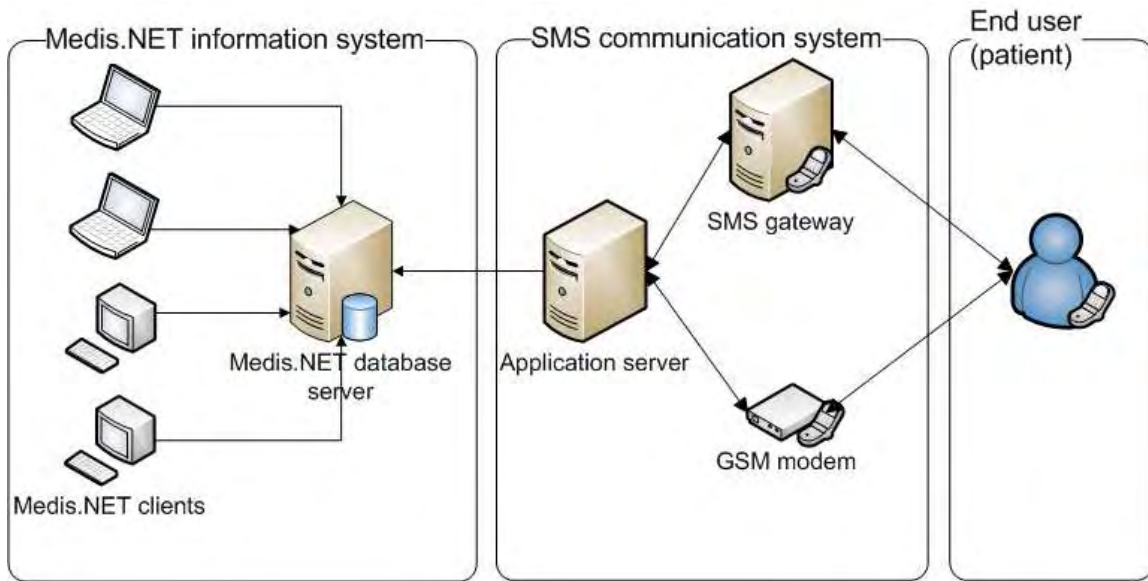


Fig. 1. System for automated sending and receiving of textual messages

online and healthcare institutions lack a possibility to quickly notify such users of some important information if needed.

*E. Broadcast possibilities*

It is possible to quickly and easily transfer information to a large number of users through SMS (for example to all patients registered in a healthcare institution database). In this case SMS outperforms traditional phone calls.

**III. IMPLEMENTATION**

Our system for automated sending and receiving of textual messages operates as a “middle layer” between existing Medis.NET information system and patients (Fig. 1.).

Left part of the diagram shows existing Medis.NET information system. Medis.NET manages patient’s Electronic Health Records (EHR), patient’s visits scheduling, work scheduling for employees, maintains financial data regarding provided services etc. All the data is kept in a central database and can be accessed by the employees of healthcare institution through Medis.NET client application.

Central part of the diagram shows our system for automated sending and receiving of textual messages. Core logic of our system is running as a service on application server node. This service listens to existing Medis.NET database for data modifications. If an important data for a patient who is subscribed to this service is modified then the service sends appropriate text message to the patient (e.g. some biochemical laboratory analysis for the patient are finished and service sends some most important results in the message also reminding him/her to collect full list of results at the healthcare institution).

Core messaging service also maintains the data about patients’ subscriptions to our system and stores them in a small database on the same server.

Third role of core messaging service is to choose appropriate interface for sending messages. As shown in Fig.1. there are two possible interfaces – SMS gateway and GSM modem.

SMS gateway is a service provided by a GSM service provider company. SMS gateway can be exposed as a web application or a web service and depending on that our core messaging service accesses it by using HTTP for web application or SOAP for web service.

GSM modem is a specialized type of modem which accepts a Subscriber Identity Module (SIM card), and operates over a subscription to a mobile operator, just like a mobile phone. From the mobile operator perspective, a GSM modem behaves just like a mobile phone. From our perspective GSM modem compared to standard mobile phone has some advantages (can be connected at the same time both to external power supply and to our server by a cable connection).

Provided with these two interfaces, our core messaging service can make a choice which one to use based on configuration parameters. Usual configuration settings are such that SMS gateway is used with a higher priority because of its lower price per message and higher throughput. GSM modem is used only as a back-up interface in case of failure of the SMS gateway.

**IV. A TYPICAL USE CASE SCENARIO**

In this chapter one interesting usage of our system for automated sending and receiving of textual messages will be described. We use it here to improve an existing system for patients’ visits appointments.

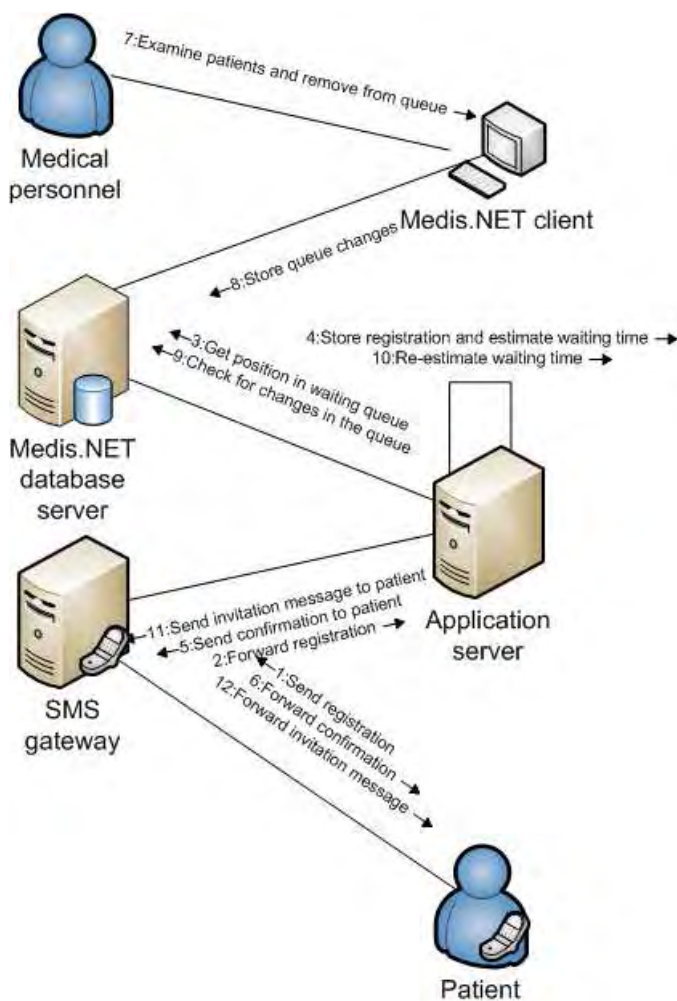


Fig. 2. Using system for patient notification

Medis.NET information system already has a sophisticated module for scheduling employees work time and appointing patient's visits [4]. This system supports automatic generating of work time schedule for employees and splits generated work time into time slots for patients. Duration of time slot is configurable, based on employee's specialization and work area (e.g. 15 minutes for general practice medicine, 20 minutes for dentistry etc.). For most specializations and work areas this approach works fine but there still some departments which make appointments for their patients on daily bases and not for precise hour and minute. Also for some work areas time needed per patient can significantly vary, for example from 15 to 30 minutes.

In such cases our system for automated sending and receiving of textual messages can enable patients to quickly check their appointment status, get estimation when they will be examined and/or get an invitation message 30 minutes before they should come to doctor's office.

Take for example a patient who already knows that he has an appointment today, but healthcare institution can't provide him exact time of his visit. Instead of waiting several hours at

the institution the patient can use our SMS service as shown on Fig. 2.

1. Patient registers for our service by sending an SMS containing service keyword and his Social Security Number.

2. SMS gateway forwards the message to core messaging service.

3. Core messaging service checks appointments and position in waiting queue for this patient.

4. Core messaging service estimates time when the patient will be examined and stores his registration in messaging database.

5. Core messaging service creates a confirmation message containing estimated time when patient's visit should start and information that he will receive one more message 30 minutes before that.

6. SMS gateway forwards confirmation message to the patient.

7. Doctor at healthcare institution examines patients and removes examined patients from the waiting queue using Medis.NET information system (Fig. 3.).

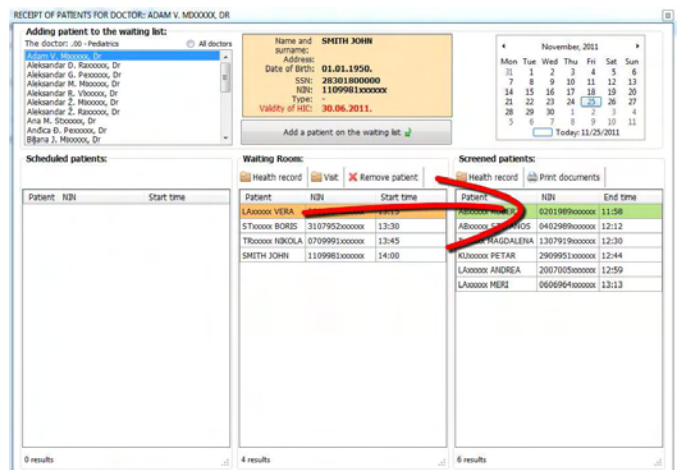


Fig. 3. Medis.NET application used by medical personnel shows lists of scheduled patients, patients in the waiting room and examined patients

8. Modifications from Medis.NET client are transferred to Medis.NET database.

9. Core messaging service detects modifications in Medis.NET database.

10. Based on new data in the waiting queue, core messaging service makes new estimation of the waiting time.

11. If new estimation of the waiting time is about 30 minutes core messaging service sends invitation message which contains new estimation for examination start time.

12. SMS gateway forwards invitation message to the patient (Fig. 4.).

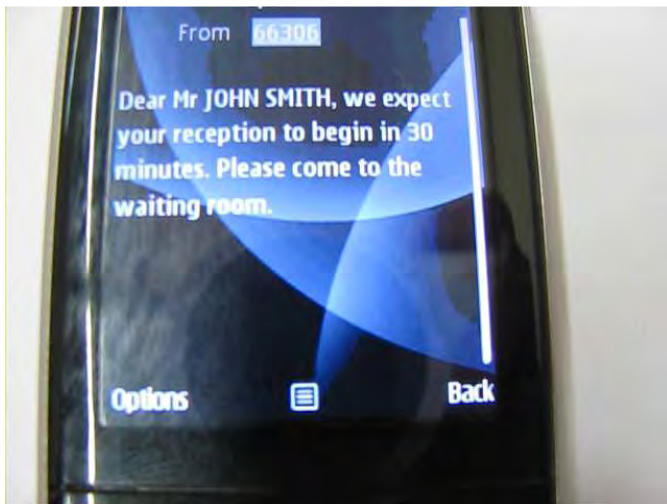


Fig. 4. Invitation message received on patient's phone

## V. CONCLUSION

In this paper we presented our implementation of a system for automated sending and receiving of textual messages. Motivated by existing systems in different areas such as e-learning, e-banking, marketing etc. ([5], [6], [7]) we designed our system as an improvement of a medical information system.

Use case presented here is planned to assist medical personnel in scheduling patients' visits and to reduce waiting time for the patients. Our plans for future work include some more services for patients based on the system for automated sending and receiving of textual messages. One possible extension is related to biochemical laboratory. When some analyses for the patient are finished the service would send most important results in the short message also reminding patient to collect full list of results at the healthcare institution. It is also possible to create a notification service which would warn a patient that a period of time has expired since his/her last visit to a doctor (for example 6 months after visiting dentist) and remind patient to consider scheduling a new visit. Another possibility is to create a service which would broadcast warning messages about changes in weather conditions to all registered patients with chronic diseases which could be affected.

Although we listed plans for extending our messaging service in healthcare area, it is important to notice that the service is not limited to this area and that it could easily be coupled with other kinds of information systems.

## REFERENCES

- [1] G. Le Bodic, *Mobile Messaging Technologies and Services: SMS, EMS & MMS*, Chichester, West Sussex, England, John Wiley & Sons Ltd, 2005.
- [2] 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.
- [3] P. Vrgović, I. Jošanov-Vrgović, B. Jošanov, "SMS Information Service: Innovative Thinking for the Successful Solution", Management Information Systems, vol. 6, no. 3, pp. 22-28, 2011.
- [4] I. Marković, S. Cvetković, D. Janković, "An Implementation of a Scheduling Tool in a Medical Information System", ICEST 2010, Conference Proceedings, pp.327-330, Ohrid, Macedonia, 2010.
- [5] A. Stone, "Mobile scaffolding: an experiment in using SMS text messaging to support first year university students", ICALT 2004, IEEE International Conference on Advanced Learning Technologies Proceedings, Joensuu, Finland, 2004.
- [6] A. Dickinger, P. Haghirian, J. Murphy, A. Scharl, "An investigation and conceptual model of SMS marketing", HICSS-37, Proceedings of the 37th Annual Hawaii International Conference on System Sciences, Hawaii, 2004.
- [7] L. Prabhakaran, W. Yan Chee, K. Chong Chua, J. Abisheganaden, W. Mun Wong, "The use of text messaging to improve asthma control: a pilot study using the mobile phone short messaging service (SMS)", Journal of Telemedicine and Telecare, vol. 16 no. 5, pp. 286-290, 2010.

# A Comparative Analysis of Mobile AR Software with the Application to the Archeological Site Medijana

Dušan Tatić<sup>1</sup>, Časlav Stefanović<sup>2</sup>, Dragan Stanković<sup>3</sup>

**Abstract** - As Augmented Reality is becoming more and more popular, it is also getting different ways of use. A particular area of application is in the field of digitalization of cultural heritage. Consequently, there are numerous tools for the application development in this area. This paper provides an overview and a comparative analysis of the software for mobile augmented reality. Based on this analysis, we selected a software tool to augment and improve the archaeological sites, by integrating 3D models of the artifacts with images of the real environment. As a case study, this paper presents the implementation details for the design of the fountain in the audience part of the imperial palace at the archaeological site of Medijana, Niš, Serbia.

**Keywords** – Augmented Reality (AR), AR browsers, Mobile AR SDK, Cultural heritage

## I. INTRODUCTION

Augmented Reality (AR) is defined as a real-time direct or indirect view of a physical real-world environment that has been enhanced by adding virtual computer-generated information to it [1]. Improvements in computer graphics technology have led to the world of virtual reality becoming more and more real, while the possibilities of combinations with real (actual) worldly environment opens a new dimension in the understanding of the concept of reality (the real world). In cases where the combination of real space and virtual elements is needed, the practical application indicates that there are no fixed limits.

According to Milgram [2], the possible combination states are shown in Fig. 1. Through the implementation of virtual objects into a real environment we create the augmented reality.



Fig. 1. Milgram's reality-virtuality continuum [2].

Augmented reality technology (AR) has achieved great progress over the last few years. The ever-increasing implementations of augmented reality are a result of the development of both hardware and software solutions. In this paper we show the implementations of augmented reality in the field of cultural heritage by using mobile devices.

<sup>1</sup>Dušan Tatić, Faculty of Electronic Engineering, University of Niš, Serbia, e-mail: dusan@dragongroup.org

<sup>2</sup>Časlav Stefanović, Faculty of Electronic Engineering, University of Niš, Serbia, e-mail: caslav.stefanovic@gmail.com

<sup>3</sup>Dragan Stanković, Faculty of Technical Science, Kosovska Mitrovica, Serbia, e-mail: sfgagi@gmail.com

## II. COMPARATIVE ANALYSIS OF AR SOFTWARE

The development of mobile devices over the last few years has contributed to the implementation of the augmented reality technology on these devices. Thus, the mobile augmented reality software has also undergone development, which has facilitated the process during which users', in an interactive manner (using images from the real world), search for information (without the use of a keyboard and mouse). Studying the field of augmented reality from the aspect of development and implementation of AR applications we can identify two groups: AR browsers and AR software development kits.

### A. AR browsers

This group includes AR browsers (ARBs), that are, applications of augmented reality aimed at helping the search of the real environment by using mobile devices cameras. By following the search of the images of the real environment, they are augmented with multimedia contents (html, image, audio, video or 3D models). AR browsers index the content through media streams (termed channels, layers, or worlds). ARBs, generally, accesses remote resources using web protocols and web services.

All ARBs work in a similar way and consist of the following components (Fig. 2.):

1. The ARB represents the client on the mobile device.
2. The publishing website - a website on which developers can register new channels (layers, worlds), and manage their channels and accounts.
3. The ARB server is the basis of this system and represents the interface towards other components (ARB, the publishing web site, and external servers).
4. The external server – a platform on which are stored all the multimedia contents (images, text, audio, and video) as well as the description of the location (POI - point of interest).

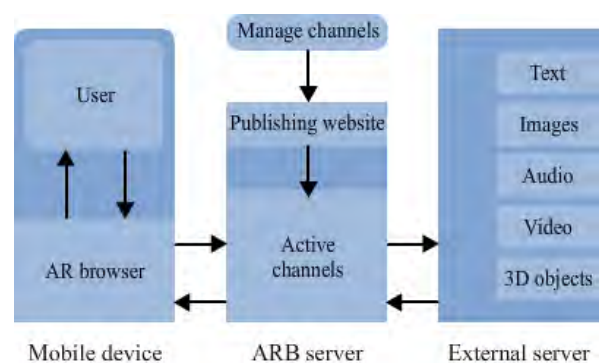


Fig. 2. AR browser system overview.

On the basis of previous research [4], shown in Fig. 3., we carried out an analysis, of the three most frequent AR browsers from the first group: Layar, Junaio, and Wikitude.

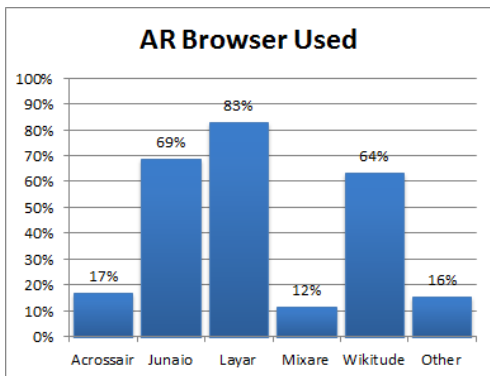


Fig. 3. Usage of AR browsers [4].

**A.1 Layar AR browser**

Layar [5] was one of the first ARB that appeared on the market, designed for the Android and iOS operating systems. It enables both location based and vision based tracking. This browser, the creation of channels also known as layers is enabled. POIs are shown, in the form either of an icon or a 3D object, on the display of the mobile device. By clicking on a POI it is possible to obtain additional information, to activate some of the multimedia contents or launch another application. The Layar browser gets data on the layer channel with the JSON structured data.

**A.2 Junaio AR browser**

Junaio [6] is a powerful AR browser which offers the support for location based tracking and both marker and markerless image recognition. For location based tracking, except for the outdoor version, an indoor version was also created, that is, one which supports AR even in those locations where GPS is not available and where the compass does not work. In order to determine the user's position in these conditions, we use the Latitude Longitude Altitude - Markers (LLA Markers).

Location based channels show points of interest. Virtual objects floating at the position of a POI. These virtual objects can take the form of: texts, thumbnails, or can be animated and static 3D objects. Each of these POIs can be linked to other images, sounds, videos, and websites.

GLUE (image based) channels are used for vision tracking and enable one to attach or "glue" virtual 3D models to any real object. The Junaio server transforms markers into tracking xml which then helps the Junaio browser to perform an optical analysis. To describe layers and POIs we use a specific XML code.

**A.3 Wikitude AR browser**

The Wikitude World browser [7] is used for location based tracking. POIs are grouped into worlds, just like layers are

formed in the case of the Layar browser. It supports 2D multimedia contents. This browser uses its own ARML language (Augmented Reality Markup Language) which is based on KML (Google) and enables the addition of more data onto the POI including addresses, telephone-numbers, links, different icons and images. Using the ARML database we can exchange data between the client (the mobile device), the Wikitude server, and external servers. This ARB can be used on iOS and Android platforms.

**B. AR software development kits**

The second group of software includes software development kits (SDKs) used to create mobile applications running on iOS and Android devices, providing mobile AR functionality. This software is by its structure and functionality significantly more complex than the software from the first group. Unlike the AR browser, the AR software development kit enables greater freedom in the development of applications.

From this group we will present two types of AR SDK software: metaio Mobile SDK and Vuforia SDK.

**B.1 metaio Mobile SDK**

Using this tool we can develop applications both for the Android and the iOS platforms. This software uses fast and modular libraries and algorithms since it was designed to be used on mobile devices. It offers location based, marker-based, and markerless 2D and 3D object tracking. In addition, the metaio Mobile SDK [8] enables both QR Code and Barcode scanning. With the help of this software, the AR browser Junaio was also developed. Fig. 4. shows the metaio Mobile SDK architecture.

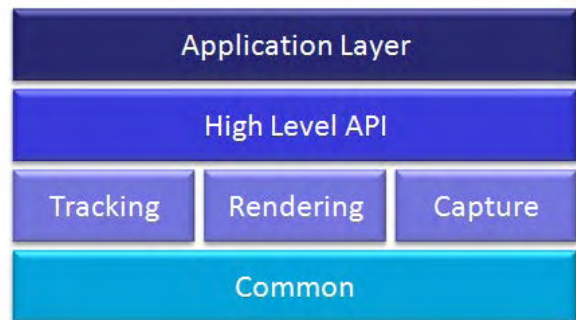


Fig. 4. Metaio Mobile SDK architecture [8].

The metaio Mobile SDK can be used in different ways depending on the type of application. For simpler applications we use the Application Layer. Applications can be configured based on an XML configuration file. The high level API development enables the realization of complex applications. At this level, knowledge of programming is required. This SDK platform also allows access to low level components. This approach is used in the case of special demands, when some other 3D graphics engine is required, as well as in connection with capturing and tracking modules.

## B.2 Vuforia SDK

Qualcomm [9] has developed a platform for augmented reality known as Vuforia. Vuforia SDK supports the Android and iOS platforms. It uses natural-feature tracking and frame markers. In addition to these characteristics, Vuforia also supports the option of virtual buttons, which registers the contact between the user and the marked parts of the marker. The tracking of objects which are rectangular in shape or something with a similar shape is also made possible. For the purpose of tracking the Vuforia SDK via one's Target management system, a dataset configuration XML file is created. The Vuforia uses OpenGL ES for rendering 3D objects. The Vuforia has an AR extension for Unity that enables vision detection and tracking functionality.

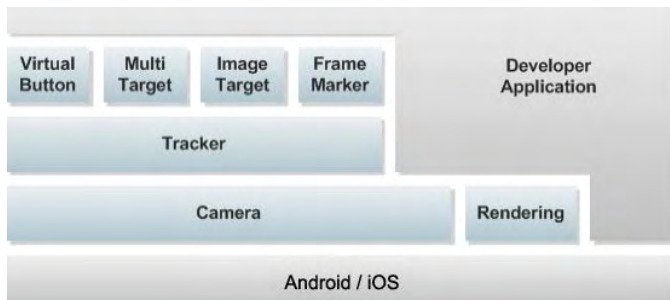


Fig. 5. System overview of Vuforia SDK [9].

## III. EVALUATION AND IMPLEMENTATION

### A. Evaluation

After experimenting in practice with the software briefly discussed above, a comparative analysis of AR browsers and AR software development kits can be summarized as in Table I. The most important elements of the software are shown: Platform – the system platform on which the described software is implemented. Tracking – The type of tracking which can be supported by the given software. Vision based – tracking which is based on the natural features (NFT) or classical (fiducial) markers [10]. Location based – this type of tracking is realized with the help of the GPS sensor and compass which are the constituent parts of a mobile device.

AR view content – the type of multi-media contents used to augment reality. 2D means classic multimedia contents (html, image, audio, video) while the 3D refers to the possibility of displaying 3D objects. 3Dani marks the possibility of displaying 3D animations.

We can conclude that the Layar and Junaio, judging by the given elements, are very similar. They can be implemented on the same platforms and support the same type of tracking except that Junaio supports location based tracking in the closed space where GPS is not available. From Table 1 we can note that Wikitude supports only location based tracking and only has the possibility of 2D contents.

In the case of AR software development kits Metaio and Vuforia, they support the same platforms. We can note a difference can only be found in the case of the location based tracking which, for now, is not supported by Vuforia AR SDK.

### B. Implementation

Taking into consideration the characteristics of the described software systems, we have opted for the Junaio AR browser. Using this AR browser we have designed an application for the archaeological site Medijana. The application was designed so that when the visitor, while touring the archaeological site, using the camera on the mobile device, detects a mosaic, the screen of the device shows a 3D model of the Roman fountain in the exact location where it was once located (Fig. 6.).

In order to realize the application, we used tracking: NFT (natural feature tracking), that is, vision based markerless tracking. In our case the tracking object is the image of the mosaic.

The AR view content represents a 3D model of the fountain. The model of the fountain is given in the \*.obj which is encrypted with the help of the online Junaio Model Encrypter. These AR content is stored on the external server.

The structure of the application follows the one given in Fig. 2. On the external server we generated a program code (php) which is used to connect with the Junaio server. This code enables the execution of other processes: the authentication process, the selection of multi-media contents as well as the transfer of the necessary parameters (XML).

The creation of channels is provided through the user interface on the Junaio server. We selected the Junaio GLUE channel. When forming the channel, we defined the other necessary data: the URL of the external server, the link for the homepage, and refresh time to search the request to the external server. It is possible to define the channel for it to be either of the type public or private. For additional searches of the channel the necessary tags are defined.



This application is experimentally verified on the tablet Asus transformer with the Android Honeycomb platform.



Fig. 6. Application example.

Augmented Reality is one of the upcoming technologies the usage of which becomes simpler thanks to the development of mobile devices. The described application realizes a new quality of the presentation of museum exhibitions. By using augmented reality technology is it possible to perform the restructuring of the archaeological sites and objects which do not exist and whose reconstruction is complex.

TABLE I  
A COMPARATIVE ANALYSIS OF MOBILE AUGMENTED REALITY SOFTWARE

AR browsers									
name	platform		tracking				AR view content		
			vision based		location based		2D	3D	3D ani
			marker based	marker less	outdoor	indoor			
Layar	✓	✓	✗	✓	✓	✗	✓	✓	✗
Junaio	✓	✓	✓	✓	✓	✓	✓	✓	✓
Wikitude	✓	✓	✗	✗	✓	✗	✓	✗	✗
AR software development kits									
Metaio Mobile SDK	✓	✓	✓	✓	✓	✓	✓	✓	✓
Vuforia AR SDK	✓	✓	✓	✓	✗	✗	✓	✓	✓

IV. CONCLUSION

REFERENCES

In this paper we carried out an analysis and a comparison of mobile augmented reality software. An application was designed for the virtual projection of the fountain in the auditorium of the emperor’s palace at the archaeological site of Medijana. We have shown one of the possible ways for displaying cultural and historical heritage at the archaeological sites.

By using these applications, the desire and interest for the acquisition of new knowledge regarding cultural heritage or history can be increased. Learning is made more complete and extensive at the very location, enhancing the learning process. In order to use this technology it is necessary to perform the digitalization of cultural heritage, which contributes to its preservation.

This type of systems offers the visitor the possibility of selecting his own choice of additional contents regarding the object being viewed, in accordance with the available time and his personal interests.

Further research on this topic should be focused on the use of more inclusive software from the AR software development kit group for the realization of multi-media guides based on AR technology which could replace standard audio guides.

ACKNOWLEDGMENT

The work presented here was supported by the Serbian Ministry of Education and Science (project III 044006).

[1] J. Carmigniani, B. Furht, M. Anisetti, P. Ceravolo, E. Damiani, and M. Ivkovic, Augmented reality technologies, systems and applications, in: *Multimedia Tools and Applications*, pp. 341-377, 2011.

[2] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino, Augmented reality: A class of displays on the reality-virtuality continuum, *Telemanipulator and Telepresence Technologie*, vol. 2351, pp. 282-292, 1994.

[3] R. Azuma, A Survey of Augmented Reality, In *Presence: Teleoperators and Virtual Environments*, vol. 6, pp. 355-385, 1997.

[4] Augmented Reality Browser Survey, Jens Grubert, Tobias Langlotz, Raphael Grasset, Institute for Computer Graphics and Vision Graz University of Technology, Austria , 2011.

[5] Layar, <http://www.layar.com/browser/>, Mar. 2012.

[6] Junaio, <http://www.junaio.com/develop>, Mar.2012.

[7] Wikitude, <http://www.wikitude.com/developer/wikitude-augmented-reality-for-developers>, Mar. 2012.

[8] Metaio Mobile SDK, <http://www.metaio.com/software/>, Mar. 2012.

[9] Vuforia AR SDK, <https://ar.qualcomm.at/qdevnet/>, Mar. 2012.

[10] Michael Gervautz, and Dieter Schmalstieg, Anywhere Interfaces Using Handheld Augmented Reality, *Computer*, Feb. 2012.

# A comparative analysis of dynamic programming languages for application in multi-agent systems

Ana Stankovic<sup>1</sup>, Dragan Stankovic<sup>2</sup>, and Dusan Tatic<sup>3</sup>

**Abstract** – In recent years the interest for dynamic programming languages has risen together with increasing popularity of Web 2.0 applications. Many web frameworks based on popular dynamic programming languages such as Rails (developed in Ruby), Grails (Groovy) or Django (Python), were created in order to improve the efficiency of web applications development by promoting Agile methodology of work and simple, maintainable code. In this paper we have analysed if the same efficiency gains can be achieved in development of multi-agent systems (MAS). In our analysis we have observed the quality and size of the code written in dynamic programming languages: Groovy, Python, and Ruby, by comparing it with the code written in statically typed Java programming language. The analysis is based on independent implementations of the asynchronous dynamic programming algorithm in all four programming languages. Obtained results can be generalized to other MAS algorithms.

**Keywords** – multi-agent systems, dynamic programming languages.

## I. INTRODUCTION

Dynamically typed programming languages have recently turned out to be really suitable for specific scenarios such as Web development, application frameworks, game scripting, interactive programming, rapid prototyping, dynamic aspect-oriented programming and any kind of runtime adaptable or adaptive software. The main benefit of these languages is the simplicity they offer to model the dynamicity that is sometimes required to build high context-dependent software. Common features of dynamic languages are meta-programming, reflection, mobility and dynamic reconfiguration and distribution [1]. Out of special interest to us when considering dynamically typed programming languages in the context of multi agent systems is rapid prototyping with additional benefit of adaptability which is not the subject of this paper.

Russel and Norvig in [2] define artificial intelligence as a scientific study of agents that are able to perceive the environment and perform actions. Since the research of multi-agents is still in its infancy, there is no universal consensus on an unequivocal definition of the concept of agent.

<sup>1</sup>Ana Stankovic is with the Faculty of Information Technologies at Metropolitan University, Tadeusa Kosusca 63, 11000 Belgrade, Serbia, E-mail: ana.stankovic@metropolitan.ac.rs

<sup>2</sup>Dragan Stankovic is with the Faculty of Technical Sciences, University of Pristina, Kneza Milosa No. 7, Kosovska Mitrovica, Serbia, E-mail: sfsagagi@gmail.com

<sup>3</sup>Dusan Tatic is with the Faculty of Electronic Engineering, University of Nis, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: dusan@dragongroup.org

Nevertheless, definition provided by Wooldridge and Jennings [3], [4] is becoming widely accepted by a growing number of researchers, which is why it can be regarded as being one of the most complete definitions. According to this definition, an agent is:

“a computer system that is situated in some environment, and that is capable of autonomous action in this environment in order to meet its design objectives”.

Interest in studying multi-agent systems usually stems from the interest in artificial (software or hardware) agents, such as the agents living on the Internet, for example. Examples of those agents are trading agents, game-playing agents that assist or replace human players in multi-player games, autonomous robots in multi-robot environments and the like.

Software agents can be regarded as a natural extension of the concept of software objects. Object-oriented programming has introduced abstraction entities – objects to the structural programming paradigm. Similarly, agent-based programming introduces new entities – agents, which, in contrast to objects, have an independent execution thread. Therefore, in comparison to objects, agents have the ability of acting in a goal-directed manner, for example, by interacting with other agents, reading sensors or sending commands to effectors, while objects only passively respond to procedure calls. In short, it can be stated that agents represent intelligent, adaptable software applications, designed with the purpose of meeting different, user-defined requirements.

In most of the cases, even separate action of agents can be useful. Nevertheless, agents achieve their fullest potential by interacting with other agents, thus making multi-agent systems. Most of these systems are heterogeneous because they consist of different types of agents that have different functions within the observed system. Agents act either in synergy with the purpose of achieving the common goal or competitively with the purpose of achieving contradictory goals.

What follows are detailed definitions of important concepts associated with agents and multi-agent systems. After that, comparison of one simple MAS-used search algorithm implemented in dynamic programming languages Ruby, Python and Groovy with statically typed Java programming language will be performed.

## II. CONCEPT OF AN INTELLIGENT AGENT

It has already been pointed out that there is no universally accepted definition of the term agent. Debates concerning this issue are still under way. In fact, while there is a general consensus that autonomy is something that is always associated with agents, not all the details have been cleared up. Perhaps the main reason for which one universally



accepted definition is difficult to find lies in the fact that agents are used for finding solutions to problems within various domains. Therefore, for some applications, the ability of agents to learn on the basis of their previous experience is very important. However, for some other applications learning is not only unnecessary but also undesirable at times (an example supporting these facts is found in the air-traffic control system; the passengers would probably not like the situation in which the system modified flights schedule at run time on the basis of previously learned facts).

Agents can be regarded as an approach to structuring and development of software that offers certain advantages and that is suitable for certain types of applications (some papers see agents as evolutionary in relation to objects) [5]. Agents' characteristics to reduce the interdependence of application components can represent their most advantageous characteristics. Agents are autonomous, which can be regarded as some kind of encapsulation [5]. While objects have their own methods that are controlled by external entities, agents do not allow external entities to control them. When an agent gets a message, being autonomous, it decides what is to be done with that message by itself. Interdependence of application components is reduced not only by the agents' autonomy, but also by their robustness, reactivity and proactiveness. For example, when an agent enters goal-directed phase, agent itself is responsible for the process of realization of that goal. It is not necessary to perform constant supervision and checking. Analogously, object can be regarded as a reliable employee that has no initiative or sense of responsibility; supervision of that employee requires increased level of communication. On the other hand, agent can be regarded as an employee that takes initiative and has the sense of responsibility. Therefore, supervision of that employee does not require increased level of communication, which is why it can be stated that there exists lower level of interdependence.

Reduced interdependence leads towards software systems that are more modular, more decentralized and more easily changeable. This resulted in the fact that agents started being used in wide specter of applications, especially in applications that are regarded as open systems, that is, applications which have been designed and written by different authors without their mutual communication. Of course, this entails the introduction of certain standards. Examples of these systems include semantic web and grid computing.

The fact that some agents are proactive and reactive makes their mode of problem solving similar to human. That feature resulted in a great number of applications in which agents are used as substitutes for humans within some limited domains. One such example is an application in which software agents are used to replace human pilots in military simulations [6]. Another example are computer games. The game Black & White uses agents that are based on BDI (Belief-Desire-Intention) model. Another field within which agents have been practically applied is the film industry. Producers of the film Lord of the Rings used the software package Massive to generate armies of orcs, elves and humans. Each individual character was modeled as an agent. Other types of applications where agents show their advantages include

intelligent assistants, e-trade, production and modeling of business processes [7], [8].

### III. DYNAMIC PROGRAMMING LANGUAGES

In the last few years, the development of Web 2.0 applications has brought about the increase of interest in dynamically typed programming languages. Great number of frameworks that enable efficient development of web applications and promote Agile application development methodology have been developed. Terms such as DRY (don't repeat yourself), KISS (keep it simple but not simpler) and convention over configuration have been adopted by the programmers and these stand for the major directions the programmers follow in the process of application development. Growing interest in programming languages that increase productivity of web developers has resulted in further expansion of their use in the desktop applications domain making languages such as Ruby, Python and Groovy extremely popular nowadays. What follows is a short description of these programming languages and their important features that improve efficiency of rapid prototyping of multi-agent systems algorithms.

#### A. Ruby

Ruby is a dynamic programming language which is characterized by a complex but very expressive grammar and a good core class library with a rich and powerful API. Ruby is based on elements of Lisp, Smalltalk and Perl, but its grammar is such that C and Java Programmers find it easier to learn. Ruby is a programming language that is completely object-oriented, but it is also suitable for procedural and functional programming styles. Ruby includes powerful metaprogramming mechanisms and can be used for the creation of new languages which are suitable for certain domains or for the creation of DSLs (Domain Specific Languages). [9]

#### B. Python

Python is a dynamic, object-oriented programming language which can be used for various forms of software development. It offers strong support for integration with other languages and tools, huge range of standard libraries and can be learned in a few days. Many programmers who had had experience with Python programming reported substantial productivity gains and easier and more maintainable code development. [10]

#### C. Groovy

Groovy is a developing dynamically typed programming language for the Java Virtual Machine. It builds upon the strength of Java but it also possesses additional features that are inspired by languages such as Python, Ruby and Smalltalk. It supports DSLs and test driven development. Its

main advantage lies in the fact that it smoothly integrates with Java objects and libraries. In fact, “Groovy is Java and Java is Groovy”. Groovy is the second referent language for Java platform (Java programming language being the first), which further explains its relation with Java [11].

D. Support for DSL creation

Common feature of all languages that have been described so far is that they can create DSLs, that is, new programming languages that enable more efficient development of applications for specific domains. This feature is important because it enables, for example, the creation of a specific DSL for the domain of multi-agent systems, which, in turn, enables more efficient modeling and development of multi-agent applications without losing the interoperability of the code written in that new DSL with standard libraries.

IV. DISTRIBUTED PATH FINDING PROBLEM

Majority of problems that occur within multi-agent systems are focused on how to meet some global constraints in a distributed way, that is, how the agents can optimize some objective function in a distributed manner. In most cases, it is achieved with the help of four families of techniques and specific problems. Those techniques are:

- Distributed dynamic programming (applied here to the path planning problem)
- Distributed solutions for Markov Decision Problems (MDP – Markov Decision Problems)
- Algorithms of optimization algorithms of economic functions (matching and scheduling problems)
- Coordination on the basis of social laws and conventions (example of traffic regulations)

With the purpose of illustration, distributed dynamic programming will be applied to the path planning problem. Path planning problem consists of a weighted directed graph with a set of  $n$  nodes  $N$ , directed links  $L$ , a weight function  $w: L \rightarrow R^+$  and two nodes  $s, t \in N$ . The goal is to find a directed path from  $s$  to  $t$  that will have minimal possible total weight. Generally speaking, a set of goal nodes  $T \subset N$  can be considered, and the shortest path from  $s$  to any of the goal nodes  $t \in T$  can be looked for.

This kind of abstract framework can be applied in various domains. It can certainly be applied in cases of some specific networks (for example, transportation or telecommunication network). Nevertheless, it can be applied to other problems as well. For example, in a planning problem the nodes can be states of the world and the arcs can be the actions that the agent performs. In that case, the weights stand for the cost of each action (for example, the time needed for the action) (37) (38).

V. ASYNCHRONOUS DYNAMIC PROGRAMMING

The problem of finding the best path is the problem that has been thoroughly studied in computer science. Distributed

solution will be considered here, in which each node performs a local computation with insight only into the state of neighboring nodes. The principle of optimality underlies the solutions that will be illustrated: “if node  $x$  belongs to the shortest path from  $s$  to  $t$ , then the part of the path from  $s$  to  $x$  (or from  $x$  to  $t$ ) must also be the shortest path between  $s$  and  $x$  (that is,  $x$  and  $t$ ). This principle enables an incremental divide-and-conquer procedure, also known as dynamic programming.

Let  $h^*(i)$  represent the shortest distance from any node  $i$  to the goal node  $t$ . In that case, the shortest distance from  $i$  to  $t$  via node  $j$  neighboring  $i$  is shown as:  $f^*(i, j) = w(i, j) + h^*(j)$ , and  $h^*(i) = \min_j f^*(i, j)$ . Having these facts in focus, ASYNCHDP algorithm has each node perform the procedure shown in Fig. 1. Within this procedure, each node  $i$  maintains a variable  $h(i)$  that stands for an estimate of  $h^*(i)$ .

It can be proved that ASYNCHDP procedure always converges to the true values, that is, that  $h$  will converge to  $h^*$ . In this case, convergence will require additional step for each node in the shortest path, which means that in the worst case convergence will require  $n$  iterations. However, this is not so good for realistic problems. Not only will convergence be slow, but this procedure also assumes the existence of agent for each node. In typical search spaces it is not possible to enumerate all nodes in an efficient way and allocate each of them a separate process. (For example, chess has approximately  $10^{120}$  positions). For that reason, programmers

```

procedure ASYNCHDP (node  $i$ )
if  $i$  is a goal node then
     $h(i) \leftarrow 0$ 
else
    initialize  $h(i)$  arbitrarily (e.g. to  $\infty$  or 0)
repeat {
    forall neighbors  $j$  do
         $f(j) \leftarrow w(i, j) + h(j)$ 
     $h(i) \leftarrow \min_j f(j)$ 
    }
    
```

Fig. 1. Asynchronous dynamic programming algorithm

often turn to heuristic versions of the procedure that require smaller number of agents.

VI. RESULTS

We have implemented the above mentioned algorithm in Java, Groovy, Python, and Ruby and used the number of lines of code as a measure for evaluating their rapid prototyping abilities. It can be further discussed whether the number of lines of code is a measure that can be suitable for the estimate of efficiency of some programming language in a specific domain (multi-agent systems). The lines of code will depend on developer’s experience with certain programming language and the applied code style rules. The code that we used here

and the way it was generated are sufficient for the process of drawing general conclusions. One such conclusion is that there is a big difference between statically typed Java and dynamically typed languages that were considered here. The graph shown in Fig. 2 illustrates this difference in the best possible way.

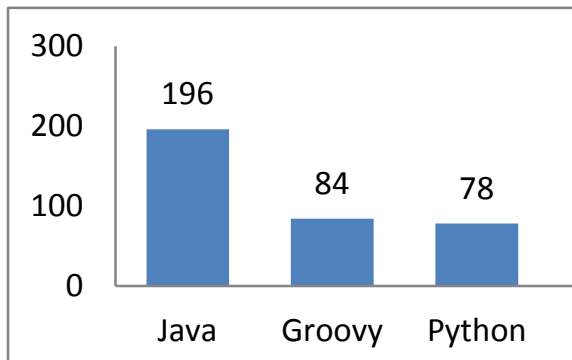


Fig. 2. Lines of code (LoC) for the complete test application that relies on AsyncDp algorithm for different programming languages

## VII. CONCLUSION

In the domain of rapid prototyping of MAS algorithms dynamically typed programming languages clearly have advantages over classic statically typed languages, such as Java. These advantages, as we have shown, are noticeable even in the simplest examples. Apart from the advantages that are reflected in reduced number of code lines and higher productivity, big advantage is also seen in the increased code readability and subsequent easier influx of broad community of developers in what was previously done, as well as in more efficient code maintenance and iterative improvement.

The conclusion that can be drawn is that dynamically typed programming languages should be given preference over statically typed languages whenever possible.

The future will probably bring increased interest in functional programming languages in the domain of multi-agent systems as well. Environments and tools with the most efficient support and broadest community of users will become dominant while the remaining projects will disappear in time.

## ACKNOWLEDGEMENT

The work presented here was supported by the Serbian Ministry of Education and Science (projects III44006 and III42006).

## REFERENCES

- [1] F. Ortin, "Type Inference to Optimize a Hybrid Statically and Dynamically Typed Language", *The Computer Journal*, vol. 54, No. 11, 2011
- [2] S. Russell, P. Norvig, "Artificial Intelligence - A Modern Approach", Prentice Hall, 2009.
- [3] M. Wooldridge, "An Introduction to MultiAgent Systems", Chichester : John Wiley & Sons, 2002.
- [4] M. Wooldridge, N. Jennings, "Intelligent Agents: Theory and Practice", *Knowledge Engineering Review*, 10(2): pp. 115–152, 1995.
- [5] M. Wooldridge, "Intelligent Agents. In: G. Weiss (Ed.), *Multiagent Systems. A Modern Approach to Distributed Artificial Intelligence*", The MIT Press, Cambridge, Massachusetts, pp. 27-78, 1999
- [6] G. Tidhar, C. Heinze, M. Selvestrel, "Flying together: modelling air mission teams", *Applied Intelligence*. 1998, Vol. 8, 3, pp. 195–218.
- [7] W. Shen, D. Norrie, "Agent-based systems for intelligent manufacturing: a state-of-the-art survey", *Knowledge and Information Systems, An International Journal*. 1999, Vol. 1, 2, pp. 129–156.
- [8] N. Jennings et al., "Autonomous agents for business process management", *International Journal of Applied Artificial Intelligence*. 2000, Vol. 14, 2, pp. 145–189.
- [9] D. Flanagan, Y. Matsumoto, "The Ruby Programming Language", O'Reilly Media, 2008.
- [10] [Online] <http://www.python.org/>
- [11] [Online] <http://groovy.codehaus.org/>

# GPU Accelerated Construction of Characters of Finite Abelian Groups

Dušan B. Gajić<sup>1</sup> and Radomir S. Stanković<sup>1</sup>

**Abstract** – In group theory and Fourier analysis on finite Abelian groups, the group characters are an essential concept. In many applications, as for instance, spectral processing of logic functions (binary or  $p$ -valued), it is often required to construct the table of group characters for the specified group. This can be a computationally demanding task, both in terms of space and time, when dealing with large groups, since the group characters are viewed in matrix notation as rows of  $(p^m \times p^m)$  matrices, where  $p$  is the cardinality of the set where the given logic function and its variables take values, and  $m$  is the number of variables. The graphics processing unit (GPU), as a highly parallel computational platform, may facilitate this complex task.

This paper discusses the application of the GPU processing to the construction of tables of group characters for finite Abelian groups represented as a direct product of cyclic subgroups of order  $p$ . We exploit the Kronecker product structure of these tables permitting redistribution of the related computing tasks over GPU resources. Experimental results confirm that the presented solution offers a considerable speed-up over the C/C++ implementation of the same character construction method processed on the central processing unit (CPU).

**Keywords** – Abstract harmonic analysis, finite Abelian groups, group characters, Kronecker product, GPU computing, OpenCL.

## I. INTRODUCTION

*Abstract harmonic analysis* is a mathematical discipline that evolved from the classical *Fourier analysis* by the replacement of the real group  $R$  with an arbitrary locally compact Abelian or compact non-Abelian group [3, 6, 9, 14, 15]. This implies the transition from the exponential functions, used in classical Fourier analysis and viewed as the group characters of  $R$ , to the *group characters*, in the case of Abelian groups, and the *group representations*, in the case of non-Abelian groups [6, 9]. Abstract harmonic analysis provides foundations for the formulation of many methods with significant applications in electrical engineering and computer science [9, 16, 17, 18, 19]. In these methods, it is often required to construct the group characters of various Abelian groups and use them in further computations. With that motivation, this paper presents a method for an efficient construction of group characters of finite Abelian groups

<sup>1</sup>Dušan B. Gajić and Radomir S. Stanković are with the University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mails: dule.gajic@gmail.com, radomir.stankovic@gmail.com.

using the graphics processing unit (GPU). This choice of hardware is made due to the fact that contemporary GPUs are highly parallel computing engines which can simultaneously serve as programmable graphics processors and scalable parallel computational platforms [1, 8, 13]. For a given group  $G$ , the construction of group characters can be expressed in terms of the Kronecker product of characters of its subgroups of smaller orders. In this formulation, the algorithm for the construction of group characters expresses a substantial inherent parallelism and, therefore, the GPU is a natural choice of hardware for the implementation of this algorithm. The experimental comparisons of the proposed implementation on the GPU and the C/C++ implementation of the same algorithm processed on the central processing unit (CPU) confirm this assumption.

The rest of the paper is organized as follows. The background theory is introduced in Section 2. In Section 3, we propose a mapping of the algorithm for the construction of group characters to the GPU and discuss the details of the respective programming implementation. The experiments are discussed in Section 4. We close the paper with Section 5, by presenting some conclusions and possible directions for further research.

## II. BACKGROUND THEORY

In this section, we give a brief introduction to the theoretical background of the paper. For more detailed discussion of these topics, we recommend classical works such as [3, 15, 17], or more recent references [6, 9, 14].

We consider finite Abelian groups of the form  $G = C_p^m = (\{0, 1, \dots, p-1\}^m, \oplus_p)$ , where  $C_p$  is the cyclic group of order  $p$ , and  $\oplus_p$  is the componentwise addition modulo  $p$ .

The group characters  $\chi_\omega^{(p)}(z)$ ,  $z = 0, 1, \dots, p^m-1$ , of the group  $G$  are defined as [9, 16, 17]:

$$\chi_\omega^{(p)}(z) = \exp\left(\frac{2\pi}{p} i \sum_{s=0}^{m-1} \omega_{m-1-s} z_s\right), \quad (1)$$

where  $i = \sqrt{-1}$ ,  $\omega_s, z_s \in \{0, 1, \dots, p-1\}$ , and

$$\omega = \sum_{s=0}^{m-1} \omega_s p^{m-1-s}, \quad z = \sum_{s=0}^{m-1} z_s p^{m-1-s}. \quad (2)$$

**Example 1** The group character tables, for the cyclic groups  $C_p$  of orders  $p = 2, 3$ , and  $4$ , are given in Table I, where  $i = \sqrt{-1}$ ,  $e_1 = -0.5 \cdot (1 - i\sqrt{3}) = \exp(2\pi i / 3)$ , and  $e_2 = e_1^* = -0.5 \cdot (1 + i\sqrt{3}) = \exp(4\pi i / 3)$ .

TABLE I CHARACTER TABLES OF CYCLIC GROUPS

Cyclic group	$C_2$	$C_3$	$C_4$
Character table	$\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 & 1 \\ 1 & e_1 & e_2 \\ 1 & e_2 & e_1 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{bmatrix}$

The group  $G=C_p^m$  is the direct product of  $m$  elementary cyclic subgroups  $C_p$ . It follows, see for instance [3, 6, 15], that the character table of the group  $G$  is the Kronecker product of  $m$  character tables of its cyclic subgroup  $C_p$ .

**Example 2** For the group  $C_3^2$ , the character table can be computed as the Kronecker product of two character tables of its cyclic subgroup  $C_3$ . In this way, only the character table of  $C_3$  is computed through (1) and the character table for  $C_3^2$  is generated as:

$$[C_3^2] = [C_3] \otimes [C_3] = \begin{bmatrix} 1 & 1 & 1 \\ 1 & e_1 & e_2 \\ 1 & e_2 & e_1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 1 & 1 \\ 1 & e_1 & e_2 \\ 1 & e_2 & e_1 \end{bmatrix} = \begin{bmatrix} 1 \cdot \begin{bmatrix} 1 & 1 & 1 \\ 1 & e_1 & e_2 \\ 1 & e_2 & e_1 \end{bmatrix} & 1 \cdot \begin{bmatrix} 1 & 1 & 1 \\ 1 & e_1 & e_2 \\ 1 & e_2 & e_1 \end{bmatrix} & 1 \cdot \begin{bmatrix} 1 & 1 & 1 \\ 1 & e_1 & e_2 \\ 1 & e_2 & e_1 \end{bmatrix} \\ 1 \cdot \begin{bmatrix} 1 & 1 & 1 \\ 1 & e_1 & e_2 \\ 1 & e_2 & e_1 \end{bmatrix} & e_1 \cdot \begin{bmatrix} 1 & 1 & 1 \\ 1 & e_1 & e_2 \\ 1 & e_2 & e_1 \end{bmatrix} & e_2 \cdot \begin{bmatrix} 1 & 1 & 1 \\ 1 & e_1 & e_2 \\ 1 & e_2 & e_1 \end{bmatrix} \\ 1 \cdot \begin{bmatrix} 1 & 1 & 1 \\ 1 & e_1 & e_2 \\ 1 & e_2 & e_1 \end{bmatrix} & e_2 \cdot \begin{bmatrix} 1 & 1 & 1 \\ 1 & e_1 & e_2 \\ 1 & e_2 & e_1 \end{bmatrix} & e_1 \cdot \begin{bmatrix} 1 & 1 & 1 \\ 1 & e_1 & e_2 \\ 1 & e_2 & e_1 \end{bmatrix} \end{bmatrix}. \quad (3)$$

This property of the character table will be exploited in the mapping of the computation of the character table to the GPU.

### III. GPU CONSTRUCTION METHOD

#### A. GPU Computing

The technique of performing general-purpose algorithms on graphics processors, known as GPGPU (*general-purpose computing on GPUs*) or *GPU computing*, has recently become a subject of a fast growing research interest and practical application [1, 13].

This interest is mainly the result of two factors. First is the evolution of the GPU hardware towards a scalable, programmable, and highly parallel computing platform [1, 13], and the second is the development of the *Nvidia CUDA* [13] and *OpenCL (Open Computing Language)* [10] programming frameworks, based on the C/C++ language, which made the immense GPU computational resources more accessible. For the implementation purposes, we use OpenCL, since it allows the development of the code that is both accelerated and portable across heterogeneous processing platforms (GPUs, FPGAs, DSPs) [8, 10].

#### B. Algorithm Mapping

The key task in porting algorithms to the GPU is their efficient mapping to the *SPMD (single program, multiple data) processing model* and the *multi-level memory hierarchy* of GPUs [1, 2, 8, 12, 13]. In the GPU SPMD model, a single data parallel function called a *kernel* is executed over a stream of data by many threads in parallel. A *thread* is the smallest execution entity and represents a single instance of the kernel. The execution of the kernel is controlled by the *host program* processed by the CPU.

The mapping of the algorithm for the construction of group character tables to the GPU is explained using Example 2.

The matrix  $[C_3^2]$  in (3) has the following block structure:

$$[C_3^2] = \begin{bmatrix} B_{00} & B_{01} & B_{02} \\ B_{10} & B_{11} & B_{12} \\ B_{20} & B_{21} & B_{22} \end{bmatrix}. \quad (4)$$

Blocks  $B_{x,y}$  ( $x, y = 0, 1, 2$ ) are the character tables for  $C_3$  multiplied by the elements of the matrix  $[C_3]$ . Therefore, each block can be represented as:

$$B_{x,y} = c_{x,y} \cdot [C_3] = c_{x,y} \cdot \begin{bmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \\ a_{20} & a_{21} & a_{22} \end{bmatrix} = c_{x,y} \cdot [a_{i,j}], \quad (5)$$

where  $c_{x,y}, a_{i,j} \in \{1, e_1, e_2\}$ ,  $x, y, i, j = 0, 1, 2$ .

To each block we assign a thread  $t = (x, y, a_{i,j})$ ,  $x, y, i, j = 0, 1, 2$ . Each thread performs a multiplication of  $[C_3]$  by a scalar, as in (5). Threads are organized into a two-dimensional  $(x, y)$  array corresponding to the matrices to be computed. Fig. 1 represents the mapping of the character table computations to the GPU threads. Each thread processes a single block, which is indicated by a different color in Fig. 1.

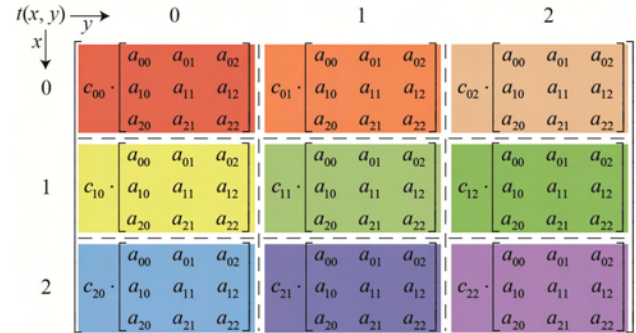


Figure 1. Mapping of the computations to the GPU threads for Example 2.

For the group  $C_3^2$  in Example 2, we have nine threads in the first and only step of the algorithm (since this example involves only one Kronecker product), each performing the operation from (5) in parallel. In this case, indices of memory locations, where a thread  $t(x, y, a_{i,j})$  stores the first element ( $c_{x,y} \cdot a_{0,0}$ ) of the block, are computed as:

$$startElement \leftarrow x \cdot 3^3 + y \cdot 3. \quad (6)$$

Indices of memory locations for the rest of the elements ( $c_{x,y} \cdot a_{i,j}$ ,  $x, y, i, j = 0, 1, 2$ , except for the case  $i = j = 0$ ) in a computed block are determined as:

$$nextElement \leftarrow startElement + i \cdot 3^2 + j. \quad (7)$$

The results of the computations are stored in the GPU global memory which has a linear layout. Formulas for the computation of the memory location indices ((6) and (7)) lead to the GPU global memory access pattern which is, for Example 2, depicted in Fig. 2. Coloring of the blocks and the memory locations in this figure corresponds to the thread coloring in Fig. 1.

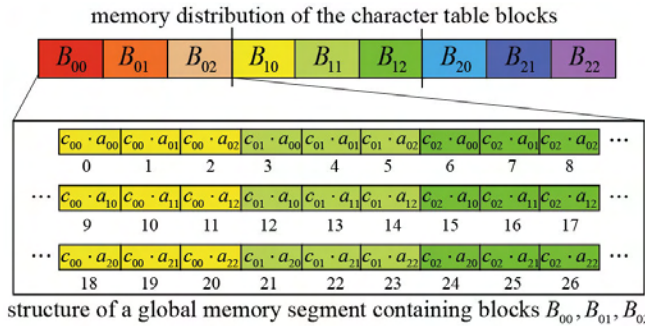


Figure 2. GPU global memory access pattern for Example 2.

In the general case, in the  $k^{\text{th}}$  step of the algorithm, we perform the Kronecker product of a  $(p^k \times p^k)$  matrix by the  $(p \times p)$  matrix, and the result is a  $(p^{k+1} \times p^{k+1})$  matrix. Therefore, there are  $p^2$  active threads in the first step of the algorithm, while in the  $k^{\text{th}}$  step, there are  $p^{2k}$  active threads. The index of the GPU memory location for the first entry ( $c_{x,y} \cdot a_{0,0}$ ) of the block is determined as:

$$startElement \leftarrow x \cdot p^{k+2} + y \cdot p, \quad (8)$$

The indices of the memory locations for the other elements ( $c_{x,y} \cdot a_{i,j}$ ,  $i, j = 0, 1, \dots, p-1$ , except for the case  $i = j = 0$ ) in a block are:

$$nextElement \leftarrow startElement + i \cdot p^{k+1} + j \cdot p. \quad (9)$$

### C. Features of the Mapping

The proposed method for computing the character tables has the following features:

1. The character table is stored as a vector of length  $p^{2m}$  obtained by the concatenation of rows of  $[C_p^m]$ . This allows reading the values of characters directly without any reordering.
2. Elements of  $[C_p^m]$  computed by threads with the same first index and the successive second index are stored in neighboring memory locations. This automatically allows memory coalescing, due to which multiple data accesses to the GPU global memory are performed as a single memory transaction [2, 12].

### D. Algorithm Implementation

A GPGPU program consists of two parts:

1. *Host program*, which executes on the CPU and creates and controls the context for the execution of kernels as well as allocates and transfers data to the GPU memory.
2. *Device program*, which is processed on the GPU and implements the SPMD kernels.

In the presented OpenCL implementation, the host program determines the character table for the cyclic subgroup  $C_p$  through (1). Notice that not all of the characters of  $C_p$  need to be computed by using (1), since, e.g.,  $e_{p-i} = e_i^*$ , for  $i = 1, 2, \dots, \lceil p/2 \rceil - 1$ . Thus, we compute half of the rows of the character table for  $C_p$ , while other rows are determined by using this property.

The host allocates GPU global memory space for two  $(p^m \times p^m)$  matrices that are used as buffers to store the results of the application of the Kronecker product. This minimizes the communication between the host and the device, which is a bottleneck in the GPU computing [8, 12, 13]. Note that we have to reserve the space for  $(p^m \times p^m)$  matrices at the beginning of the computation, since the size of the GPU buffers cannot be changed after their creation, otherwise, we would have to create buffers and transfer data between the host and the device for each step of the algorithm, as the resulting intermediate matrices increase in size. To minimize the memory bandwidth occupation on the GPU itself, we use the technique of buffer swapping [7]. For odd-numbered steps, the first matrix is used as the input to the kernel and the second matrix as the output. For even-numbered steps, the order is reversed.

The character table for  $C_p$  is stored in a  $(p \times p)$  matrix and it is used as the second operand in the Kronecker product operation in each step. Since it is of a small size, we keep it in the constant GPU memory, which is cached. This allows much faster access and leads to improved program performance [12].

The Algorithm 1 presents a pseudo-code for the device program. Code in lines 2 and 6 implements (8) and (9), respectively. Since the characters of finite Abelian groups are complex numbers, elements of  $[C_p^k]$ ,  $[C_p]$ , and  $[C_p^{k+1}]$  are stored in the GPU buffers using the *float2* OpenCL vector data type [10]. The first component in the vector variable stores the real part and the second component the imaginary part of the complex number.

---

#### Algorithm 1 Pseudo-code for the device program

---

- 1:  $x, y \leftarrow$  acquire thread indices in the two-dimensional grid
  - 2:  $startElement \leftarrow x \cdot p^{k+2} + y \cdot p$
  - 3:  $adr1 \leftarrow x \cdot p^k + y$
  - 4: **for**  $i = 0$  to  $p-1$  **do**
  - 5:   **for**  $j = 0$  to  $p-1$  **do**
  - 6:      $nextElement \leftarrow startElement + i \cdot p^{k+1} + j$
  - 7:      $adr2 \leftarrow i \cdot p + j$
  - 8:      $[C_p^{k+1}](nextElement).re \leftarrow [C_p^k](adr1).re \cdot [C_p](adr2).re - [C_p^k](adr1).im \cdot [C_p](adr2).im$
  - 9:      $[C_p^{k+1}](nextElement).im \leftarrow [C_p^k](adr1).re \cdot [C_p](adr2).im + [C_p^k](adr1).im \cdot [C_p](adr2).re$
-

TABLE II SPECIFICATION OF TEST PLATFORMS

Platform	A	B
CPU	AMD Phenom II N830 triple-core (2.1GHz)	Intel Core i7-920 quad-core (2.66GHz)
RAM	4GB DDR3 1066MHz	12GB DDR3-2000
OS	Windows 7 Ultimate (64-bit)	
IDE	MS Visual Studio 2010 Ultimate	
SDK	AMD APP 2.6	Nvidia GPU Computing 4.0
GPU	ATI Radeon 5650	Nvidia GTX 650 Ti
engine speed	650 MHz	900 MHz
memory	1 GB GDDR3 800 MHz	1 GB GDDR5 4.2 GHz
processors	80	384

I. EXPERIMENTAL RESULTS

The experiments reported in this section are performed using two hardware platforms, labeled **A** and **B**, respectively, and specified in Table II. The GPU kernel performance analysis is done through the application of AMD APP Profiler 2.4 (for **A**) and Nvidia Parallel Nsight 2.1 (for **B**), in accordance with instructions provided in [2, 12].

The referent C/C++ implementation uses the *complex* data type from the Standard Template Library (STL) for the representation of the values of group characters. This data structure best corresponds to the *float2* OpenCL vector data type [10] used for the same purpose in the GPU implementation. The referent C/C++ implementation is compiled for the *x64* platform using the MS C++ compiler set to the maximum level of performance-oriented optimizations.

The results of the experiments performed on both test platforms for the construction of the character table for the groups  $C_3^m$ ,  $m = 1, 2, \dots, 8$ , are presented in Fig. 3. Notice that for  $p = 3$  and  $m = 8$ , the size of the character table is  $p^m \times p^m = 3^8 \times 3^8 = 6561 \times 6561$ , and, therefore, to complete the task, we have to compute and store 43 046 721 complex numbers. The OpenCL implementation processed on the GPUs outperforms the referent CPU C/C++ implementation on both platforms and for all values of  $m$  used in the experiments. The speed-up is almost constant throughout the range for  $m$ , and it goes up to a factor of  $7.8 \times$ , on the test platform **A**, and up to a factor of  $8.2 \times$  on the platform **B**.

II. CONCLUSIONS

In this paper, we propose a method for the construction of characters of finite Abelian groups of the form  $G = C_p^m = (\{0, 1, \dots, p-1\}^m, \oplus_p)$ , using the graphics processing unit (GPU) as the computational platform. We identify the sources of the parallelism available in the algorithm for construction of the character table for  $G$  formulated in terms of the Kronecker product. Based on this analysis, we devise a mapping of the computations to the SPMD processing model of the GPU and develop an OpenCL implementation of the algorithm. The experimental results obtained through the comparison of the proposed solution and the referent C/C++ implementation of the same algorithm show speed-ups of up to  $7.8 \times$  and  $8.2 \times$ , depending on the platform, when using the GPU and, thus, confirm the validity of the proposed approach.

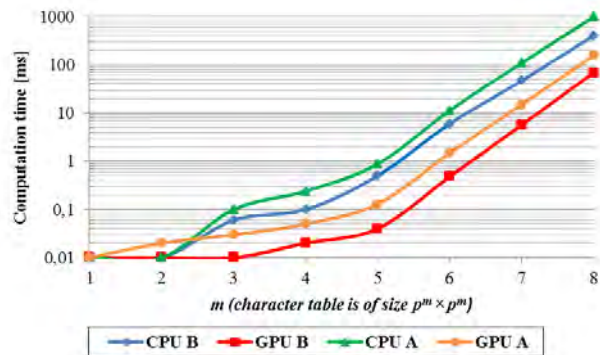


Figure 3. Computation times for the groups  $C_3^m$ ,  $m = 1, 2, \dots, 8$ , on CPUs and GPUs for the test platforms specified in Table II.

REFERENCES

- [1] T. M. Aamodt, "Architecting graphics processors for non-graphics compute acceleration", in *Proc. 2009 IEEE PacRim Conf. Comm., Comp. & Sig. Proc.*, Victoria, BC, Canada, 2009.
- [2] AMD, "AMD Accelerated Parallel Processing OpenCL Programming Guide", available from: <http://developer.amd.com/sdks/AMDAPPSDK>, [accessed 1 April 2012].
- [3] T. Apostol, *Introduction to Analytic Number Theory*, Springer-Verlag, New York, USA, 1976.
- [4] M. Clausen, "Fast generalized Fourier transforms", *Theoretical Computer Science*, No. 67, 1989, pp. 55-63.
- [5] J. W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex Fourier series", *Mathematics of Computation*, No. 90, 1965, pp. 297-301.
- [6] D. S. Dummit and R. M. Foote, *Abstract Algebra*, John Wiley & Sons, 2003.
- [7] D. B. Gajić, R. S. Stanković, "GPU accelerated computation of fast spectral transforms", *Facta Universitatis - Series: Electronics and Energetics*, Vol. 24, No. 3, University of Niš, Niš, Serbia, 2011, pp. 483-499.
- [8] B. R. Gaster, L. Howes, D. Kaeli, P. Mistry, and D. Schaa, *Heterogeneous Computing with OpenCL*, Elsevier, 2011.
- [9] M. G. Karpovsky, R. S. Stanković, and J. T. Astola, *Spectral Logic and Its Applications for the Design of Digital Devices*, Wiley-Interscience, 2008.
- [10] Khronos, "OpenCL Specification 1.2", Khronos OpenCL Working Group, 2011.
- [11] D. K. Maslen and D. N. Rockmore, "Generalized FFTs – A survey of some recent results", in *DIMACS Workshop in Groups and Computation*, 1998, pp. 183-238.
- [12] Nvidia, "OpenCL Best Practices Guide", available from: <http://developer.nvidia.com/nvidia-gpu-computing-documentation>, [accessed 1 April 2012].
- [13] J. Owens, M. Houston, D. Luebke, S. Green, J. Stone, and J. Phillips, "GPU computing", *Proc. of the IEEE*, Vol. 96, No. 5, 2008, pp. 279-299.
- [14] C. C. Pinter, *A Book of Abstract Algebra*, Dover, 2010.
- [15] W. Rudin, *Fourier Analysis on Groups*, Wiley, 1990.
- [16] R. S. Stanković, and J. T. Astola, *Spectral Interpretation of Decision Diagrams*, Springer, New York City, USA, 2003.
- [17] M. R. Stojić, M. S. Stanković, and R. S. Stanković, *Diskretne transformacije u primeni*, Nauka, Beograd, 1993, (in Serbian).
- [18] M. A. Thornton, "Spectral transforms of mixed-radix MVL functions", in *Proc. IEEE Int. Symp. on Multiple-Valued Logic (ISMVL)*, Tokyo, Japan, May, 2003, pp. 329-333.
- [19] M. A. Thornton, R. Drechsler and D. M. Miller, *Spectral Techniques in VLSI CAD*, Kluwer Academic Publishers, 2001.

# Modern Processor Architectures Overview

Danijela Jakimovska<sup>1</sup>, Aristotel Tentov<sup>1</sup>, Goran Jakimovski<sup>1</sup>, Sashka Gjorgjievska<sup>1</sup> and Maja Malenko<sup>1</sup>

*Abstract* – Complexity of the modern processor architectures and the constant race of the various computer technologies resulted in a wide range of computer architectures, each with its advantages and disadvantages, but with the ultimate goal to increase overall computer systems performances. This paper provides an overview of these architectures, including RISC, CISC, Superscalar, VLIW, EPIC and Data Flow processor architectures. All of them are presented together with their drawbacks. The concept of parallelizing instruction's execution is also described, in order to emphasize the goal for executing more than one instruction in a single clock cycle, and in that way to increase overall system performances. Moreover, examples are given of commercial implementations of the modern computer architectures.

*Keywords* – Processor architectures, RISC, CISC, ILP

## I. INTRODUCTION

Modern microprocessors are one of the most complicated systems that have been ever created by the human beings. Microprocessors have the main role in each system, since they handle the instruction and data flow, control the communication with memory and external devices and thus coordinate the whole system operation. Computer architects face with the challenge to maximize computer performance, while retaining the cost, power and functional requirements. Regarding this, they should consider three aspects of computer architecture design, including: instruction set architecture, organization (memory system, memory interconnect, internal processor), and hardware logic design. Therefore, in order to optimize the architecture design, an architect must be familiar with many technologies, such as compilers, operating systems, logic design and packaging.

One of the first computer architectures, such as the Intel IA-32, belongs to the Complex Instruction Set Computer (CISC) design which takes advantage of the microcode and is consisted of a wider range of variable-length instructions. In order to reduce the complexity of the instructions, Reduced Instructions Set Computing (RISC) was introduced, speeding up the process of decoding the instruction.

Superscalar processor and Very Long Instruction Word (VLIW) architectures were introduced in order to achieve Instruction Level Parallelism (ILP) exploiting the pipeline mechanism. The main issue in ILP is detecting and overcoming data dependencies, which Superscalar processors

provide with the hardware, while VLIW processors with software support.

Architectures like Explicit Parallel Instruction Computing (EPIC) and Data Flow were designed to solve some of the problems detected in the previous parallel architectures. EPIC processors overcomes the hardware dependencies and furthermore, Data Flow machines provide concurrency in program execution.

## II. RISC, CISC AND SUPERSCALAR

### A. CISC Architecture

CISC is designed to use complex instruction set which makes the assembly languages closer to the operations and data structures of the High Level Languages, [1]. CISC instruction set is proposed to take advantage of microcode, and is consisted of many variable-length instructions, which can specify a sequence of operations. The CISC instructions are characterized with complexity, in terms of instruction formats and addressing modes, and therefore require serial (slow) decoding algorithms, [2]. The memory references are usually combined with other operations, such as add memory data to register. CISC processors generally have few registers, and some of them may be special-purpose, which restricts the ways how they can be used, [3]. Accordingly, the CISC architecture complicates the instruction's decoding and scheduling, and therefore is not very suitable for pipelining [4].

The family line of x86 processors is the leader in CISC computing, used both as general purpose processors and for embedded systems. Using the x86 processors in embedded design is relatively novel approach and is mainly purposed for the 32- and 64-bit designs such as Intel's Atom, VIA's Nano, Athlon's Neo, and VIA's C7.

### B. RISC Architecture

The researches intended to improve the existing (mainly CISC) architectures led to reducing the set of instructions and their complexity, and thus RISC was designed. RISC instruction set consists of simple fix-sized operations, which are easy (quick) for decoding, and therefore suitable for pipelining. Contrary to CISC, RISC instructions are less complex, support simple addressing modes, and do not require a microcode for their implementation, [2], [4]. RISC processors have a relatively large number of general-purpose registers. RISC instructions reference to the main memory only via simple load and store operations. This is the main reason why the RISC processors are usually referred as load-store architecture designs, [5].

<sup>1</sup>D. Jakimovska (danijela@feit.ukim.edu.mk), A. Tentov (toto@feit.ukim.edu.mk), G. Jakimovski (goranj@feit.ukim.edu.mk), S. Gjorgjievska (sashka@feit.ukim.edu.mk) and M. Malenko (majam@feit.ukim.edu.mk) are with the Faculty of Electrical Engineering and Information Technologies, Dept. of Computer Science and Engineering, Karposh 2, Skopje, Republic of Macedonia



Widely known success stories behind the RISC architecture are AMD 29K, ARM, SPARC, Power/PC, and MIPS. ARM takes the lead in embedded systems, such as smart phones and tablet computers, due to its low cost and low power consumption.

### C. Instruction Level Parallelism

Further research led to an idea, which explains that splitting the work of a single processor to multiple processors will increase the productivity and will speed up the execution of the instructions. The main setback here is that programs were written and were meant to be executed in sequential manner, that is, one instruction at a cycle. This restriction of the sequential execution came from the data dependencies between the variables in a program.

Pipelining is a mechanism which enables instruction level parallelism, since parallel instructions are executed in parallel over multiple cycles. The theoretical increase of the instruction execution speed is proportional to the pipeline length, [2]. However, there are three potential hazard problems that can occur, during the pipeline execution of parallel instructions. Data hazards appear when an instruction result depends on the previous instruction; structural hazards happen when there is not enough hardware space for the parallel instructions execution and control hazards are result of the unexpected program counter change, [1]. There are several mechanisms targeting these problems, and the simplest solution is to stall the pipeline.

### D. Superscalar architecture

RISC architecture is very suitable for exploiting parallelism through pipelining, since all the RISC instructions are simple and take roughly the same time to finish. The instruction execute stage is usually the most time consuming pipeline operation, so if the processor architecture employs multiple execution units, it would always have a busy one, while the others would be idle, [2]. As a solution to this problem, computer architects proposed a superscalar architecture, characterized with parallel instructions execution on multiple executing units. Although this architecture utilizes more execution units simultaneously, the number of hazards is increased and the processor has to retire the instructions in program order if they are re-ordered (dynamic scheduling). Besides that, handling branch operations becomes very problematic, since a typical program executes a branch after each six or seven instructions. A possible solution would be to utilize a special hardware branch predictor, which would predict the target with some probability, on behalf of the previous branch results. However, this approach has still some disadvantages, such as speculatively wrong executed instructions. The processor architects overcome this difficulty, by extending the processor architecture with reorder buffer. This buffer is intended to store the results of speculatively executed instructions, and to update the real state after the correct instructions has completed, [1].

## III. VLIW AND EPIC ARCHITECTURES

### A. Very Long Instruction Word

Unlike the Superscalar architecture, which relies on hardware to detect and overcome data dependencies, Very Long Instruction Word (VLIW) processors, [3] use software solutions (compilers) to mark independent instructions. Usually compilers translate the code into intermediate language, optionally optimize it, and as well generate machine code for the specific architecture, [6]. VLIW processors use compilers that inspect the source code and thus concentrate on scheduling and optimizing the raw source code before translating it.

Scheduling is done using two structures: control flow and data flow graphs, [6]. Control flow graphs divide the source code into basic blocks that must be sequentially executed (usually the delimiter is a branch or a label), [6]. The data flow graph shows the dependencies between registers within a basic block and those which are independent can be executed in parallel, [6]. Here, the disadvantage is that a single basic block usually contains four to six operations and therefore, limits the amount of parallelism that can be achieved. To maximize parallelism, a technique called global scheduling is used, where instructions are moved from one block to another, [6].

Transmeta's Crusoe® and Texas Instruments 320C6x line of processors are the commercial breakthrough behind the VLIW architecture. Transmeta's Crusoe processor uses hybrid hardware-software implementation of the VLIW architecture. It is a 128 bit architecture that uses Code Morphing Software that detects and resolves data dependencies. The Code Morphing software also implements routines for power management and thermal dissipation, which makes this processor ideal for mobile devices. The Texas Instruments 320C6x line of processors is a general purpose DSP processor using the VLIW architecture, which is mainly used as a research processor. It has wide range of debugging tools and compilers available for research.

### B. Explicitly Parallel Instruction Computing

The explicitly parallel architecture was designed to overcome some essential limitations of the VLIW architecture, such as hardware dependence. The EPIC architecture solves the hardware dependence problem, by defining several mini-instructions which can be combined in groups, depending on the template type field, [8]. Therefore, the processors which are characterized with greater parallelism capabilities will simply exploit more bundles in parallel. EPIC processors utilize dispersal technique to issue two bundles at a time, and split-issue mechanism if the mini-instruction cannot be executed. Furthermore, the instructions can be predicted, reducing the cost of a branch operation. However, the cost for wrong prediction is very high, because branches appear very often. EPIC processors provide hardware support for the control speculation of loads and they allow parallel issuing of multiple prioritized branch

operations. Speculative load failures are resolved with poison bits utilization, [1], [2].

The EPIC architecture doesn't solve all the problems of the VLIW architecture. The difficulty with the increased code size (because of the empty slots in the bundles) still remains unsolved. This has a negative impact on cache performance and bus bandwidth utilization. Other very important feature of each EPIC processor is the good compiler support, [1].

This architecture has only one implementation, as part of the IA-64 processor architecture of Itanium family processors, [1]. Intel Itanium architecture processors have been designed from the ground up to meet the increasing demands for high availability, scalability and performance needed for high-end enterprise and technical computing applications. In its core, Itanium was designed to address a number of performance bottlenecks in computers, such as memory latency, control flow dependencies and memory address disambiguation. It enables the hardware to take advantage of the available Instruction Level Parallelism and to provide the necessary resources, while focusing on dynamic runtime optimizations.

Precision Architecture – Reduced Instruction Set Computer (PA-RISC) [7] was originally designed as a 32-bit architecture, intended to be easily scalable across a broad performance range, while providing for straightforward migration of applications from existing systems. It was rather conservative RISC design, but still competitive in terms of speed, especially for simultaneous multiprocessing and floating-point operations.

#### IV. DATAFLOW ARCHITECTURES

Concurrency is a major step in increasing computational performance, especially with today's technological limitations. Dataflow architecture offers an attractive alternative to the conventional control flow architecture in providing concurrency in execution of programs. Execution of each dataflow instruction depends only on the availability of its operands, which implies implicit synchronization of parallel activities. There are no constraints in sequencing of dataflow instructions, except for the conventional data dependencies in a program.

Data flow architecture differs from control-flow architecture, by two basic principles: asynchronous operations and functionality. Dataflow instructions are executed only when all input operands are available (assuming hardware resources are also available), in contrast to control-flow model which uses program counter for sequential ordering of instruction execution. The functionality rule implies that any two enabled instructions can be executed in either order or concurrently, only if they don't interfere with each other (don't have data dependences), which implies parallel processing.

A dataflow program is represented as a directed graph, where named nodes represent instructions and links represent data dependencies among instructions, [9] [10]. Dataflow graphs can be described as machine language for dataflow computers. Data is conveyed from one node to another in data packets called tokens. This flow of tokens enables nodes (instructions) which depend on them and fires them.

TABLE I  
COMPARISON OF PRESENTED ARCHITECTURES

Architecture/ Characteristics	ILP	Instructions per cycle	Instruction format
<b>RISC</b>	Yes, pipeline	Depends on the pipeline depth (usually 4 or 5)	Fixed length, Usually 16,32,64 etc
<b>CISC</b>	Yes, similar to superscalar	Depends on the operation's complexity	Variable length, Complex operations
<b>Superscalar</b>	Yes, Multiple pipelines	Depends on the pipeline depth (usually 8 or 10)	Fixed length, Usually 16,32,64 etc
<b>VLIW</b>	Yes, Fixed number of instructions in the word, Implicit parallelism	Depends on the number of execution units (usually 4 or 8)	Fixed length, Multiple instructions in one word
<b>EPIC</b>	Yes, variable, but limited number of instructions in the word, Explicit parallelism	Depends on the number of execution units (6 to 8)	Variable length, Multiple microinstructions in the instruction word
<b>Dataflow</b>	Yes, variable, but limited number of instructions (due to operands availability and hardware resources)	Depends on the number of execution/functional units	Packet format of instructions (PISC)

Dataflow architecture can generally be divided into pure dataflow architecture and hybrid dataflow architecture. Pure dataflow firing rule says that an instruction can be executed as soon as all input operands are available to it. It gives the dataflow model asynchronous behavior and self-scheduling of instructions.

Pure dataflow architecture is subsequently divided into static, dynamic and explicit token store architectures, while the hybrid architecture utilizes some known control flow mechanisms, [11]. Pure dataflow architecture executes a program by receiving tokens, each containing data and tag, processing instructions and sending out newly formed tokens, [9]. When a set of matched tokens (tokens with same tag) is available at the execution unit, processing starts by fetching the appropriate instruction (with the same tag) from the instruction store. The instruction is executed and the result is generated, containing data and tag of the subsequent instructions which depend on it.

Pure dataflow architectures have some serious drawbacks, the major being is bad single thread performance. Other problems are: the overhead produced by token matching, as well as implementing efficient unit for matching tokens (resolved to some extent in explicit token store architecture, [12]).

Unfortunately drawbacks made it difficult to achieve direct implementation of computers based on a pure dataflow model. For this reason, possibilities of converging dataflow and control-flow models were investigated and broad spectrum of hybrids (techniques and machines based on them) were developed: threaded dataflow, coarse-grain dataflow, RISC dataflow, dataflow with complex machine operations, [13], [14].

## V. CONCLUSION

This stream line of processor architectures is highly unlikely to die off. Other architectures are about to emerge, some similar to the previous architectures and other completely different. The advance of technology requires new and better processor architectures and exploitation of the concept of parallel computing. Overcoming the data dependencies issue has to be done in order to achieve better

results in executing a program in parallel and better overall performance. Also, we need to further investigate the possibilities to narrow the gap between memory response time and processors working frequencies. However, in order to be able to go forward and toward developing completely new different processor architecture one must make thorough investigation of existing processor architectures. That was the main reason for leading this investigation, and presenting the results of it within this paper.

## REFERENCES

- [1] John L. Hennessy, David A. Patterson, "Computer Architecture: A Quantitative Approach", 2007
- [2] Nicholas FitzRoy-Dale, "The VLIW and EPIC processor architectures", 2005
- [3] William Stallings, "Computer organization and architecture designing for performance", 8<sup>th</sup> edition, Prentice Hall, 2009
- [4] Sivarama P. Dandamudi, "Guide to RISC processors: for programmers and engineers", Springer, 2005
- [5] Prof. Vojin G. Oklobdzija, "Reduced Instruction Set Computers", 1999
- [6] Zomaya, A.Y.H. , Parallel and Distributed Computing Handbook, McGraw-Hill, Ch. 21, Thomas M. Conte, "Superscalar and VLIW Processors", 1996
- [7] Hewlett – Packard, Precision Architecture :The Processor, HP journal, August 1986
- [8] Mark Smotherman, "Historical background for EPIC", February 2011
- [9] J. Silc, B. Robic, T. Ungerer, "Processor architecture: From Dataflow to Superscalar and Beyond", Springer, 1999.
- [10] A. L. Davis, R. M. Keller, "Data flow program graphs", IEEE Trans. On Computers, February. 1982.
- [11] R. A. Iannucci, "Toward a dataflow/von Neumann hybrid architecture", Proc. 15th ISCA, May 1988.
- [12] G. M. Papadopoulos, "Implementation of a general-purpose dataflow multiprocessor", Tech. Report TR-432, MIT Laboratory of Computer Science, Cambridge, Aug. 1988.
- [13] J. Silc, B. Robic and T. Ungerer, "Asynchrony in parallel computing: From dataflow to multithreading, Parallel and Distributed Computing Practices", 1998.
- [14] R. Buehrer, K. Ekanadham, "Incorporating dataflow ideas into von Neumann processors for parallel execution", IEEE Trans. On Computers, Dec. 198.

# The mechanism for flexible symbology in mobile GIS

Miloš Roganović<sup>1</sup>, Bratislav Predić<sup>2</sup>, Dragan Stojanović<sup>3</sup>, Marko Kovačević<sup>4</sup>

**Abstract** – This paper presents an innovative approach for implementing visual portrayal of geographic features with point, polyline (complex polylines) and polygon geometry for application in mobile GIS applications. This approach relies on developed XML styling language for defining custom styles and mapping layer subclasses to the specific style. Benefits of this approach are extending basic styles and applying custom, user defined style to distinguish object that belong to the same layer. Visually separated objects can emphasize a specific attribute value that is crucial for fast identification of objects in time critical situations. Typical emergency situations require synergic response from various response team all accessing different aspects of the same geospatial data. Flexible and customizable style definition language is applied to general vector data allowing separate field services to have custom visualization of shared community spatial data.

**Keywords** – styling, GIS, XML, visualization.

## I. INTRODUCTION

A geographic information system (GIS) is special type of computer-based information system tailored to store, process, and manipulate geospatial data [1]. The ability of GIS to handle and process both location and attribute data distinguishes GIS from other information systems. It also establishes GIS as a technology important for a wide variety of applications [2].

The fundamental information unit that GIS deals with is called a geographic feature. Geographic feature is an abstraction of a real world phenomenon associated with a location relative to the Earth [3]. Every feature may have a number of properties. One or more of the feature's properties may be geometric. Geometry provides the means for quantitative description of the spatial characteristics of features, including dimension, position, size, shape, and orientation. A geometric object is a combination

of a coordinate geometry and a coordinate reference system. In general, a geometric object is a set of geometric points, represented by their coordinates. Basic geometric objects are points, polylines, and polygons [4].

The importance of the visual portrayal of geographic data in GIS cannot be neglected. The skill that goes into portraying data is what transforms raw information into an explanatory or decision-support tool. Fine-grained control of the graphical representation of geographic features is a fundamental requirement for any professional mapping community. Allowing user to define styling rules for visual portrayal of geographic features requires the existence of a styling language that the user and GIS application can both understand [5].

The paper is organized as follows: Section II presents previous work. Section III describes creating custom layer styles that can be applied to any vector layer. Section IV shows flexible symbology mechanism in mobile GIS application. Section V summarizes the achieved results.

## II. RELATED WORK

Mobile GIS application, as integral part of complete GIS solution in the domain of defense and security, is developed on basis of GiniMobile platform [6]. It is capable of visualizing geospatial data in the form of raster and vector layers [7]. All data required for specific field operation is prepared in advance using desktop GIS part of the complete solution and deployed to mobile devices using GIS server and standardized WMS and WFS interfaces. Geospatial data being deployed to mobile device is separated into two groups:

- Data which is not of primary interest for specific operation. Data in this group is not expected to change during field operation and it will not be changed by specific field team using the mobile device or any other field deployed team. This data is referred to as background data and is visualized as static raster background map.
- Dynamic data which is primary focus of specific field team. Data from this group is organized as vector layers and individual features of these layers can be modified by field team using the mobile device or any other field team involved in the operation.

<sup>1</sup>Miloš Roganović is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: milos.roganovic@elfak.ni.ac.rs

<sup>2</sup>Bratislav Predić is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: bratislav.predic@elfak.ni.ac.rs

<sup>3</sup>Dragan Stojanović is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: dragan.stojanovic@elfak.ni.ac.rs

<sup>4</sup>Marko Kovačević is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: markko.marce@gmail.com

Data from the first group mainly contributes to spatial orientation and landscape/terrain recognition. In the domain of defense, security and intelligence operations data from the second group is organized into situational map and this shared situation map becomes central hub for collaborative field operations. One example of combined visualization of background raster map with situational map overlaid is shown in Fig. 1.



Fig. 1. Combined maps in mobile GIS application

All general purpose GIS operations are implemented and supported. These include numeric coordinate transformations between various (well known and manually defined) coordinate systems, measurements (length, circumference, area etc.), zooming and panning map manually. Measurements and cartometric functions are shown in Fig. 2.

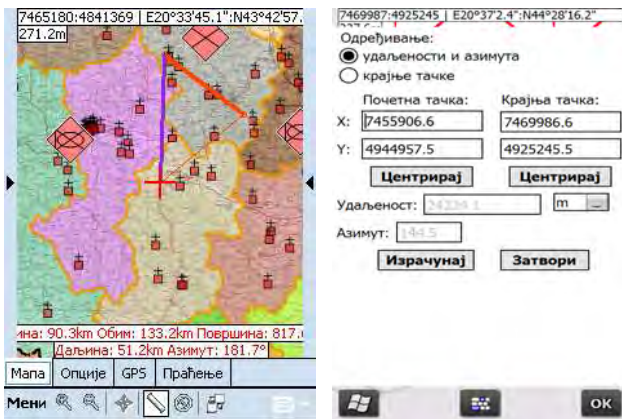


Fig. 2. Basic GIS functions in mobile GIS application

All operations are based on map view central point. User drags map and centers crosshairs on situational map elements. Crosshairs shape depends on focused type of operational map element and is shown in Fig. 3.



Fig. 3. Selecting operational map elements

### III. CREATE CUSTOM LAYER STYLE

While desktop applications offer more detail because of the greater amount of available resources (memory, processing power, and no restrictions on the battery), mobile application stand out in a more efficient use in field conditions. With the price of sensors recently declined significantly, sensors such as GPS, accelerometers, and gyroscopes are incorporated as standard in modern mobile devices. Application of mobile applications that will be discussed in this paper refers to the Geo-Information Systems. The paper will describe a mechanism for flexible symbology of geospatial objects. Mobile GIS application relies on the use of OGC defined standards.

Mobile GIS applications are most frequently used in navigation, but it is a narrow set of mobile GIS applications. In this paper we are focused on the use and modification of geospatial data using mobile applications. Data can be in .shp, .dbf format or any specific format, or may be obtained from the service.

Using this method it is possible to load data from OGC standardized data source memorized in the local file in format (.shp, .dbf). Also, data can be loaded from the service (WFS). It is important to emphasize that the objects which are loaded are vector objects whose location and other attributes can be changed in real time.

The loaded data contains one or more attributes related to the geometry of the object itself, and additional attributes that describes detailed information of the object. Each layer of geospatial data consists of data whose geometry is the same type (eg. point, line, polygonal). After you load the data layer is displayed in the default style.

For data that is loaded into the mobile application, later can be applied created style. Style depends on the geometry type of layer loaded. In this way, the mobile application styles are grouped into three classes: styles for point objects, line objects and polygon objects. For point sources, we can define the symbol color and shape and size of the spot object. An example

is given in Fig 4 a). In line objects can define the color and line type and line endings. An example is given in Fig 4 b). For polygonal objects, we can define the color and type of linje which marks the edge of the object, and then we can define the color and type of fill. An example is given in Fig 4 c).

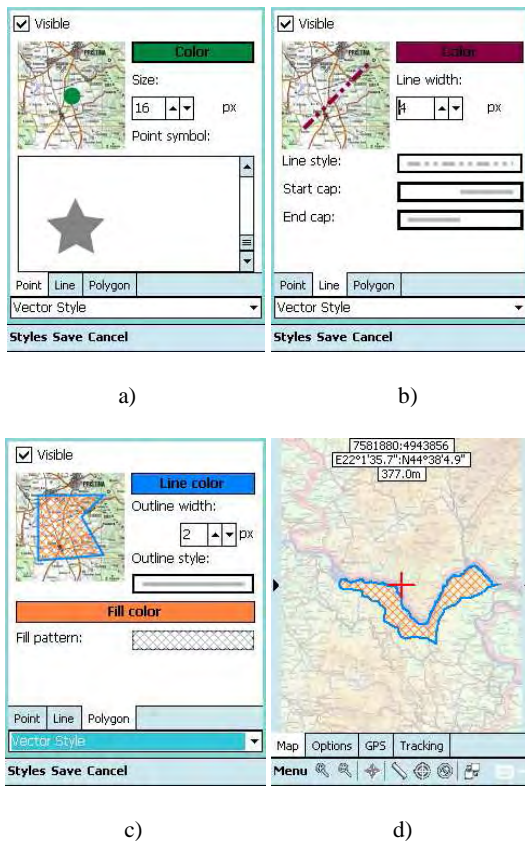


Fig. 4. Setup style for whole layer based on layer geometry

On Fig. 4.d) is shown one object that belongs to layer “National parks” of Serbia on which polygon style is applied.

Previously mentioned manner can indicate a particular layer or vector data to distinguish between two layers of the same geometry type, but has conceptually different meaning. All this is enough for use in mobile applications in navigation. If you need to show or indicate one type of points of interest (POI Eng.), but what if you want the visually differences of data that are belonging to the same layer. With that purpose a mechanism has been developed for flexible symbology in a mobile GIS.

#### IV. FLEXIBLE SYMBOLOGY MECHANISM IN MOBILE GIS

The mechanism is based on XML-defined mapping of attributes of mobile on visual representation of the object. It is necessary to define a hierarchy or a subset of data for a given layer. Once we define subsets of data for a given layer, it is necessary to prepare the appropriate raster or texture for the visual representation of instances of classes. In this way, all objects are classified into a special classes of one layer, if you look at climate change visualization of natural disasters can be more detailed graphically represented with visualization of measured values. Disaster management can greatly facilitate services on the ground to determine the appropriate plan of action.

As a data source can be any standard defined data source, you need to configure mapping of one attribute from source to a particular that can be used as symbol identifier. Each class in the hierarchy is defined by selected attribute. This can be done through a defined XML file that will be used for mapping or in the application. For specialized display, source defined attribute that serves to identify the symbols subclasses. Fig. 5 Shows mapping process, on Fig. 5. a) all source attributes are displayed unmapped. When user select an attribute, list of possible match from hierarchy is displayed and user can perform mapping. If source doesn't have all attributes from hierarchy, then those attributes are not considered during display process. Unmapped attributes from source are added as additional set of attributes beside hierarchy attributes, although they are not involved in display process.

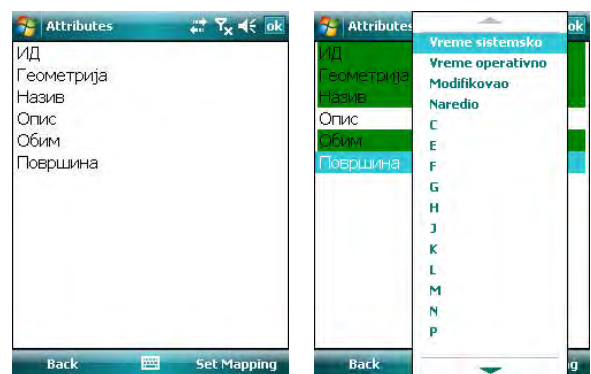


Fig. 5. Mapping an attributes from data source to specific domain hierarchy



Fig. 6. Applying different style on one layer objects

## V. CONCLUSION

To improve the use of mobile applications, a mechanism for flexible symbology of objects in mobile GIS applications has created. This mechanism extends basic styles and improves usability of mobile application. Big improvement has made in creation custom styles that apply on each object based on specified attribute value. Mechanism contains tree simple steps to configure custom mapping. First step is to classify subsets of objects for specified layer based on object attribute value. Second step is creation of raster images or textures. Final step is creating unique mapping. From user perspective, different kind of services can apply their own symbology in order to visually represent object from specific domain. This improvement allows extending visual representation of same geo-spatial data according to domain in which data are used.

Fig. 6. Shows four objects that belongs to same layer but they are visualized differently, based on previous mapping. Each object has geometry that indicates location of object, other attributes describe objects in details. We configure mapping between object attribute and symbol identifier. We create list of textures and include them in mapping.

User can create object using mobile application and see preview of the object based on previous mapping, shown on Fig. 7.

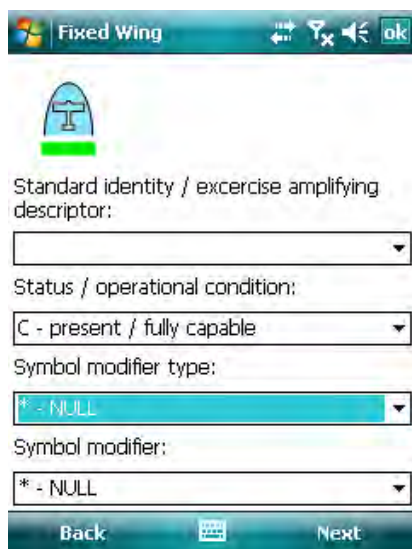


Fig. 7. Creating new raster image for visual representation of object with preview

## VI. ACKNOWLEDGMENTS

Research presented in this paper is funded by Ministry of Education and Science, Republic of Serbia as part of the projects 'Environmental Protection and Climate Change Monitoring and Adaptation', Nr. III-43007 and 'The infrastructure for electronically supported learning in Serbia', Nr. III-47003.

## REFERENCES

- [1] Worboys, M., and Duckham, M., *GIS: A Computing Perspective, Second Edition*, CRC Press, Boca Raton, FL, 2004.
- [2] Chang, K., *Introduction to Geographic Information Systems, Third Edition*, McGraw-Hill, New York, NY, 2005.
- [3] *The OpenGIS Abstract Specification, Topic 5: Features* (Version 5.0), document 08-126, Open Geospatial Consortium Inc., January 2009, <http://www.opengeospatial.org/standards/as>
- [4] *OGC Reference Model* (Version 2.0), document 08-062r4, Open Geospatial Consortium Inc., November 2008, <http://www.opengeospatial.org/standards/orm>
- [5] *Styled Layer Descriptor profile of the Web Map Service Implementation Specification* (Version 1.1.0), document 05-078r4, Open Geospatial Consortium Inc., June 2007, <http://www.opengeospatial.org/standards/sld>
- [6] Predic, B., Stojanovic, D., Djordjevic-Kajan, S., „Developing Context Aware Support in Mobile GIS Framework“, 9th AGILE Conference on Geographic Information Science, Visegrád, Hungary, 2006
- [7] Rančić, D., Predić, B., Dimitrijević, A. "Optimizations of raster map visualization in mobile GIS", Proceedings of the 10th WSEAS International Conference on COMPUTERS, Vouliagmeni, Athens, Greece, July 13-15, 2006 (pp1191-1195) ISSN:1790-5117, ISBN: 960-8457-47-5

# GinisED tools for spatial analysis of electric power supply network

Aleksandar Stanimirović<sup>1</sup>, Leonid Stoimenov<sup>2</sup> and Danilo Vulović<sup>3</sup>

**Abstract** – **GinisED is the geographic information system specially designed for electric power utility companies. GinisED was developed by CG&GIS Lab, Faculty of Electronic Engineering, University of Niš, and is deployed in ED Jugoistok Niš. A very important part of GinisED are tools for spatial analysis of electric power supply network, potential or actual events in electric power supply network and risk factors for a certain geographic area. In order to provide support for efficient spatial analysis we have developed a specific data model for representing electric power supply network in GinisED. Our data model is a combination of topological model (nodes and edges for representing geographic information) and typical graph data structure (for representing electric network attributes.**

**Keywords** –GIS, power supply network, spatial analysis, data model

## I. INTRODUCTION

Making decisions based on geography is basic to human thinking because geographic location is an important attribute of all human activities policies, strategies and plans. By understanding geography and people's relationship to location, we can make informed decisions about the way we live on our planet. Geographic information systems (GIS) are a special class of information systems that keep track not only of events, activities and things, but also of where these events, activities, and things happens or exist [1].

GIS enables capturing, storing, analyzing, and displaying geographically referenced information. It allows us to view, understand, query, interpret, and visualize data in a way that is quickly understood and easily shared. GIS technology can be used for scientific research, resource management, and development planning. Because geography location is very important for every day human activities, GIS solutions provide support for making decisions and solving many different problems: environmental monitoring, transportation management, public safety, facility security, disaster management, etc.

Management and visualization of underground utilities have been always of a great concern in many countries.

<sup>1</sup>Aleksandar Stanimirović is with the Faculty of Electronic Engineering at University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: aleksandar.stanimirovic@elfak.ni.ac.rs.

<sup>2</sup>Leonid Stoimenov is with the Faculty of Electronic Engineering at University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: leonid.stoimenov@elfak.ni.ac.rs.

<sup>3</sup>Danilo Vulović is with the Faculty of Electronic Engineering at University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: danilo.vulovic@elfak.ni.ac.rs.

Pipelines such as water supply, sewage, power supply, heat supply, industrial pipelines and communication lines, are essential infrastructures in cities. This tendency has been increased in the last decade while rebuilding and substituting existing principal pipelines to fit in with increased demands of citizens and industrial development. Moreover, newly designed underground pipelines for different purposes have been built along with various engineering projects. Insufficient, inaccurate and unclear information about the location and depth of cables and pipelines may cause various problems and may even result in tragic accidents [2].

Efficient functioning of utility companies engaged in the transmission and distribution of electricity cannot be achieved without proper record keeping and monitoring of the transmission and distribution network system [3]. Therefore, almost any electric power supply company has a need for the existence of specialized GIS solution that should provide mechanisms for collecting, storing and manipulating spatial data concerning transmission and distribution network system. This specialized GIS solution provides support for recording company assets, their locations their condition, how they are performing, and how much it costs to provide the service [4].

GIS tools for utility companies provide a very effective tool for generating maps and statistical reports from a database. However, GIS functionality far exceeds the purposes of mapping and generating reports. In addition to the basic functions related to automated cartography and data base management system, spatial analysis is the vital part of every GIS solution for utility companies. GIS is used to collect spatial data concerning distribution network, spatial analysis should be able to answer complex questions regarding space.

## II. RELATED WORK

The ability to take the geographic location of objects into account during search, retrieval, manipulation and analysis lies at the core of a GIS. How well these tasks can be accomplished is determined by the spatial data model. The starting point for modeling of geographic information is the geographic feature. A feature is an abstraction of a real world phenomenon. A geographic feature is a feature associated with a location relative to the Earth. A digital representation of the real world can be thought of as a set of features [5].

Among many of the commonly used GIS data models, simple features model has the most simple data structure. Every geographic feature has a unique persistent ID, set of thematic attributes (represented as alphanumeric data) and a number of geometric attributes. In this model, feature geometric attributes are limited to simple geometries. Simple geometries are defined as 2-dimensional geometries and linear



interpolation is used for curve representation. Basic geometry types are points, lines and polygons [5].

Another model commonly used in GIS is the topological model [1]. Topological features are simple features structured using certain topological rules. Introduction of topological relations between features simplifies data validation, modeling behavior of connected objects and optimization of different spatial analysis. The network model is just a special case of the topological model. The network data structures are used for modeling flows. Tree structure is used to model the radial network, while the graph structure is used to model cyclic networks. As new elements, topological model introduces nodes and edges. A node is a distinguished point that connects one or several arcs. An edge is a line composed by a start and an end node. The advantage of this model is the topological information it is containing: every object includes information about the elements it is related to.

Database management systems are vital part of every modern operational GIS [1]. DBMS solutions provide GIS with standardized approaches for storing and, more importantly, accessing and manipulating geographic data using some standard query language. GIS provides the necessary tools to load, edit, query, analyze, and display geographic data. Innovative work in the GIS field has extended standard relational DBMS solutions with specialized support for storing and managing geographic data.

There have been several attempts to define a standard for representing and processing geographic data in relational databases. The GIS community, working under supervision of ISO and OGC boards, has defined the core geographic types and functions to be used in a relational DBMS and accessed using the SQL language [6].

### III. GINISED – GIS FOR ELECTRIC UTILITY

GinisED is the geographic information system specially designed for electric power utility companies. It uses the most modern GIS technologies and methodologies for collecting, editing, visualization and analysis of spatial electric power supply network data. GinisED was developed by CG&GIS Lab, Faculty of Electronic Engineering, University of Niš, and is deployed in ED Jugoistok Niš (Serbian public electric power supply company). GinisED helps ED Jugoistok Niš in everyday operation and maintenance as it provides the accurate, reliable spatial and non-spatial information to the utility operational staff, and in turn help them better meet customer needs [7][8].

GinisED (Figure 1) tools can be divided into three groups [4][5]:

- Tools for collecting (digitization, map scanning and vectorization using GPS and other specialized equipment) and editing the spatial electric power supply network data.
- Tools for visualization of spatial electric power supply network data for a certain geographic area.
- Tools for spatial analysis of electric power supply network, potential or actual events in electric power supply network and risk factors for a certain geographic area.

Basic components of the GinisED system are (Fig. 1):

- *Centralized geospatial database* for storing data about electric power supply network. Elements of electric power supply networks are stored as a collection of geobjects. Each geobject is presented by its spatial component (coordinates or geometry) and thematic component (e.g. parameters for the conductors and protective devices such as voltage level, conductor type, length, construction and device type).
- *GinisED Editor* is a specialized tool for geographic editing of distribution network. It is a desktop application developed in accordance with carefully studied needs and requirements of customers. It is used for creation and editing of geographic schemes of the network, editing parameters of network elements and their connectivity.
- *GinisED Web* is a WebGIS application that allows quick and easy positioning on a specific geographic area, search and selection of parts of electric power supply networks. This application implements information integration functionalities and uses data from centralized geospatial database.
- *WMS [9], WFS [9] and custom Web Services* are components that provide GinisED Web and GinisED Mobile Server with raster maps and information considering geo-objects. WMS and WFS are components built according to Open Geospatial Consortium specifications. Custom Web Services are used for data integration and enable GinisED Web with searching and reporting capabilities.
- *GinisED Mobile Editor* is tool for GPS survey of electrical consumers and relevant electrical assets. *GinisED Mobile Server* supports mobile GIS applications and provides map segments and synchronisation between mobile database and centralized geospatial database.

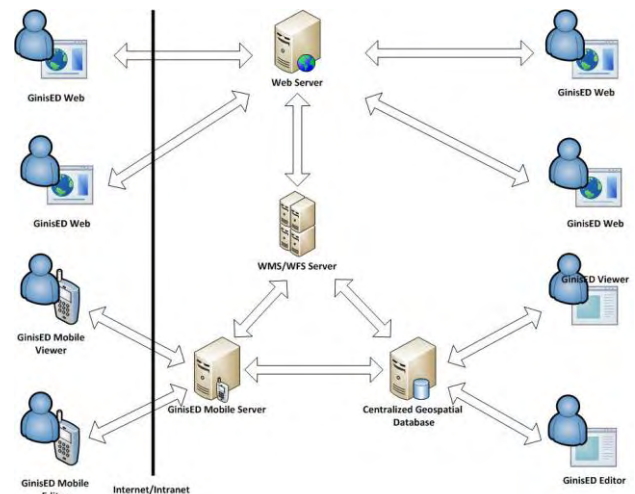


Fig. 1. GinisED architecture

### IV. MODEL FOR REPRESENTING POWER DISTRIBUTION NETWORK

In order to provide support for efficient spatial analysis we have developed a specific data model for representing electric power supply network in GinisED. Power distribution network usually contains features with line geometry (power

cables of different types) or point geometry (substations, poles, street lights, etc.). The GinisED system provides 2D visualization of all the elements of the power distribution network (Fig. 1). The major trace of power cables is mostly the same, i.e. under the streets, which results in overlapping lines. To avoid this overlap, in visualization we offset the multiple power cables to increase the readability of the map (schematic view).

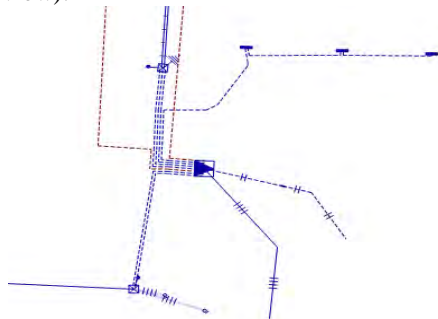


Fig. 2. 2D visualization of power distribution network (schematic view)

In order to represent specifics of power distribution network we have designed GinisED network model (Fig. 3). GinisED data model is a combination of topological network model (nodes and edges for representing geographic information) and typical graph data structure (for representing electric network attributes). The core part of GinisED network model is based on our GinisFrame object-oriented model for representing geographic features.

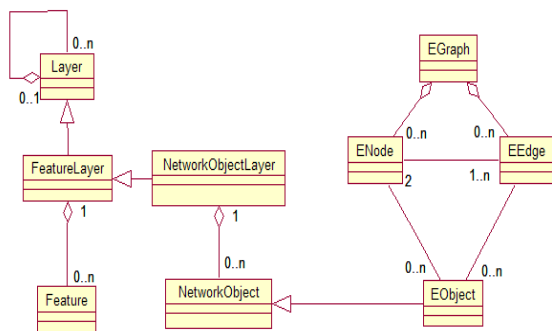


Fig. 3. The GinisED network model

Layer, FeatureLayer and Feature classes are derived from the original GinisFrame object model. These three classes define the basic organization of geographic information in GinisED system. Layer class is a basic organizational structure of our network model. Different layers can be organized in a hierarchical structure, thus providing a hierarchy and aggregation of geographic features. The FeatureLayer class defines layers of simple features (Feature class). This model is extended with NetworkObject class for network features and NetworkObjectLayer class for appropriate layers. Network objects, as opposed to simple features do not contain explicitly defined geometry. The geometry of the network features is indirectly contained in topological graph (Fig. 3).

EGraph class implements topological graph that defines geometry of power network elements. EObject class

implements power network elements and contains all thematic attributes. Geometry of power distribution network is defined through nodes (point geometry) and edges (line geometry). Additionally, every node element contains connection matrix that defines electrical connections between elements of power distribution network (Fig. 4).

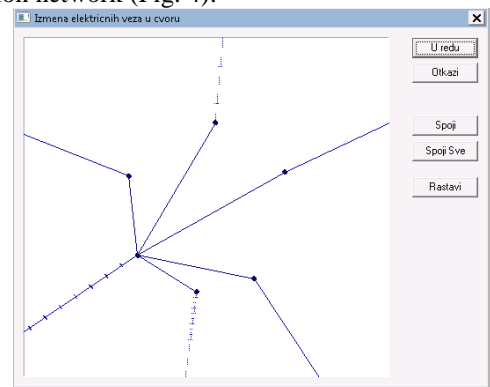


Fig. 4. Electrical connections in node

GinisED system uses centralized geospatial database for storing data about electric power supply network. This database is implanted using Oracle DBMS (Fig. 5). Each network feature on a map is stored as one logical record in EOBJECT table. A topological part of our model is stored in relational database using Oracle Spatial geometry model. Oracle Spatial geometry model has several simple geometry types and collections of them. One of the benefits of using Oracle spatial data types is that a quite extensive number of spatial queries can be performed at database level. In most scenarios spatial relationships of geometries such as “nearest to”, or “within a specified region” are needed. Using this mechanism, queries which can be formulated as “is a cable at a particular distance from a cadastral parcel” or “find the nearest buildings and their distance to a specific cable” or “find all underground networks in a given region” can be performed.

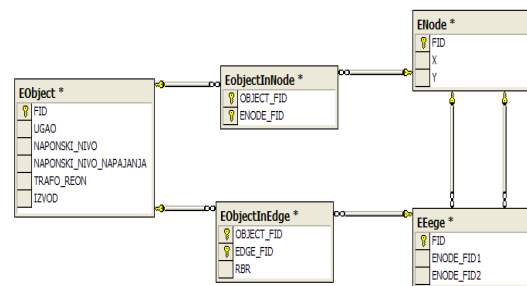


Fig. 5. GinisED geospatial database

## V. GINISED TOOLS FOR SPATIAL ANALYSIS

Tools for spatial analysis are one of the most valuable parts of GinisED system. Creating making maps alone and generating standard statistics reports does not justify the high cost of building a GIS. The greatest strength of GinisED system is possibility to utilize various types of information in

the spatial context and to generate new information and conclusions on the basis of this analysis.

A very important part of GinisED are tools for spatial analysis of electric power supply network, potential or actual events in electric power supply network and risk factors for a certain geographic area. Spatial analysis in GinisED involves two types of operations: attribute query (also known as non-spatial query) and spatial query. Attribute query requires the processing of attribute data without referencing spatial information. Spatial query involves selecting features based on location or spatial relationships, which requires processing of spatial information.

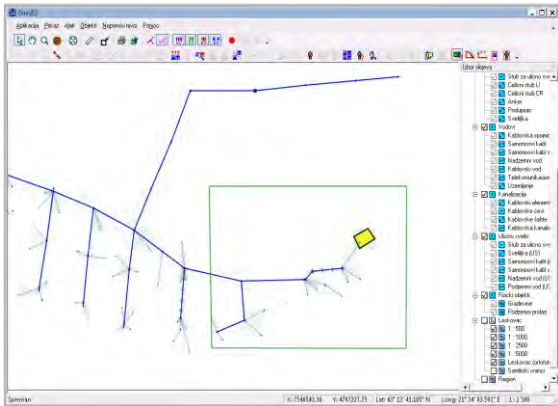


Fig. 6. Spatial query: Find all network elements in specified region

GinisED provides simple point-and-click query capabilities and sophisticated queries, analysis and display functions using visual, user-friendly techniques (Fig. 6). GinisED provides timely information and decision making support to GIS users. It helps users to solve problems, analyze geographic situations, extract necessary data and generate new information.

While basic spatial analysis involves some attribute queries and spatial queries, complicated analysis typically require a series of GIS operations including multiple attribute and spatial queries, alteration of original data, and generation of new data sets. In order to support effective spatial analysis GinisED provide powerful query builder that can combine different types of attribute queries and spatial queries.

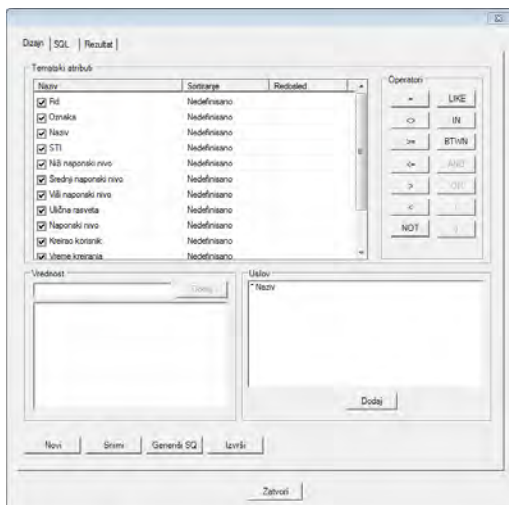


Fig. 7. GinisED Query builder

VI. CONCLUSION

Development of GinisED system for ED Jugoistok Niš started in 2004. Before GIS was implemented, data maintenance was erratic and time-consuming. Keeping track of changes was difficult because hard-copy maps or digital CAD drawings were shared among the utility’s different departments.

For the past six years implemented GIS solution has improved efficiency in overall company operations. GinisED quickly became an essential part of the ED Jugoistok Niš day-to-day business. Today, GinisED is considered very important data resource for the distribution analyses as it contains most facility information including major network topological structures. The circuit map display provided by the GinisED system is the most natural graphic user interface for engineers. In some applications, such as the trouble call analysis, geographic maps provide more information to engineers than the traditional one-line diagrams.

REFERENCES

- [1] P. A. Longley, M. F. Goodchild, D. J. Maquire, D. W. Rhind, *Geographic Information Systems and Science*, ISBN 978-0470721445, Wiley, 3rd ed., 2010.
- [2] S. Zlatanova, F. Döner, P. van Oosterom, “Management and visualization of utility networks for local authorities: a 3D approach” *Electronic Government and Electronic Participation, Joint Proceedings of Ongoing Research and Projects of IFIP EGOV and ePart, Schriftenreihe Informatik 37*, Trauner Verlag, pp. 459-474, 2011
- [3] D. Pickering, J. M. Park, D. H. Bannister, 1993, “Utility Mapping and Record Keeping for Infrastructure”, *Urban Management and Infrastructure - Urban Management Programme*, Washington, D.C., Vol. 10, pp. ix-11.
- [4] J. I. Igbokwe, E. J. Emengini, “GIS in Management of Electricity Distribution Network: A case study of Onitsha-North L.G.A., Anambra state, Nigeria”, [http://www.gisdevelopment.net/application/utility/power/utility\\_p0022pf.htm](http://www.gisdevelopment.net/application/utility/power/utility_p0022pf.htm) , accessed 12.01.2011
- [5] OpenGIS Reference Model (Version 0.1.3), document 03-040, Open Geospatial Consortium, Wayland, Mass., [http://portal.opengeospatial.org/files/?artifact\\_id=3836](http://portal.opengeospatial.org/files/?artifact_id=3836), March 2003.
- [6] OpenGIS Simple Feature Specification for SQL (Revision 1.1), document 99-049, Open Geospatial Consortium, Wayland, Mass., [http://portal.opengeospatial.org/files/?artifact\\_id=829](http://portal.opengeospatial.org/files/?artifact_id=829), March 2003.
- [7] 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”, *Proceedings of IX Triennial International Conference on Systems, Automatic Control and Measurements SAUM*, ISBN-86-85195-49-7, Niš, 2007.
- [8] L. Stoimenov, A. Stanimirović, N. Davidović, M. Bogdanović, A. Krstić, D. Nikolić, “GinisED Enterprise GIS - Framework for the Utility of the Future”, *CIRE2011*, Frankfurt, Germany, 6-9. June, ISSN:2032-9644, [www.cired2011.org](http://www.cired2011.org)
- [9] Open Geospatial Consortium, WMS and WFS Specifications, 2002, <http://www.opengeospatial.org/>

# Computer Methods and New Values for Cut Set Catalan Numbers

Iuliana Dochkova-Todorova<sup>1</sup>

**Abstract** – Improved methods to determine the values of the cut set Catalan numbers are presented. These methods require computing the number of special solutions of a system of linear inequalities. Moreover, the table of values for cut set Catalan numbers is completed up to  $n=29$ .

**Keywords** – cut set Catalan numbers, dissection of a convex polygon, triangulation.

## I. INTRODUCTION

Cut set Catalan numbers were introduced in [1] as a variation of the well-known classical Catalan numbers  $C_n$ . The Catalan number counts the number of dissections of a regular  $(n+2)$ -gon with labelled vertices by  $n-1$  non-intersecting diagonals into  $n$  triangles. The cut set Catalan numbers  $S_n$  are defined as follows: Let  $P_{n+2}$  be a regular convex  $(n+2)$ -gon with labelled vertices. Consider dissections of  $P_{n+2}$  by  $n-1$  non-intersecting diagonals into  $n$  triangles. Two such triangle sets  $T'=\{T'_1, \dots, T'_n\}$  and  $T''=\{T''_1, \dots, T''_n\}$  are said to be isomorphic if there exists a bijection  $\varphi: T' \rightarrow T''$  such that  $T'_i$  and  $T''_i$  are congruent for  $i=1, \dots, n$ . Then the cut set Catalan numbers  $S_n$  is defined to be the number of isomorphism classes of all such dissections of  $P_{n+2}$ .

In contrast to the classical Catalan numbers  $C_n$ , for  $S_n$  no explicit formula is known up to now. Some methods to determine  $S_n$  are presented in [1] and [2]. The new method in [4] is partly based on this results. We will now systemize the methods and analyze the ways to computing  $S_n$ . Moreover, the values of  $S_n$  are presented in [4] for  $n \leq 29$ .

The cut set Catalan numbers are registered in the On-Line Encyclopedia of Integer Sequences (OEIS, see [5]) with the sequence number A033961.

In [1] is proved that  $S_n$  equals the number of special solutions of a system of linear equalities. Then the method is improved in [2] and in [4]. Now we will present here these computer methods.

## II. SPECIAL LINEAR SYSTEM AND $S_N$

A dissection  $\Delta$  of a convex  $(n+2)$ -gon  $P$  means here a dissection by  $n-1$  non-intersecting diagonals into  $n$  triangles. The label  $l_T(s)$  of a side  $s$  belonging to a triangle  $T$  in  $\Delta$  is defined as the number of vertices of  $P$  between the

<sup>1</sup>Iuliana Dochkova-Todorova is with the Faculty of Mathematics and Informatics at St. Cyril and St. Methodius University of Veliko Tarnovo, 83 G. Kozarev str., Veliko Tarnovo 5000, Bulgaria, E-mail: doshkova@uni-vt.bg.

endvertices  $u$  and  $v$  of  $s$  on that part of the boundary of  $P$  not containing the third vertex  $w$  of  $T$ . Note that the label sum of the three sides belonging to  $T$  equals  $n-1$ . If  $a, b$  and  $c$  are the labels of the sides of a triangle  $T$  in  $\Delta$ , where  $a \leq b \leq c$ , then  $T$  is said to be of type  $(a,b)$ .

Let  $T_n$  be the set of the possible types of triangles:

$$T_n = \{(a,b) \in N_0 \times N_0; 0 \leq a \leq b \leq (n-1-a)/2\}. \quad (1)$$

We use  $\lambda_{a,b}$  for the number of triangles of type  $(a,b)$  in  $\Delta$ . It is easy to see that the upper bound  $\lambda_{a,b} \leq \lfloor (n+2)/(a+b+2) \rfloor$  holds for all  $(a,b) \in T_n$ .

A triangle of type  $(a,b)$  is called a boundary triangle if  $a=0$ , and an inner triangle if  $a \geq 1$ . The set of types for the inner triangles is

$$I_n = \{(a,b) \in N_0 \times N_0 | 1 \leq a \leq b \leq (n-1-a)/2\}. \quad (2)$$

A central triangle in  $\Delta$  is a triangle of type  $(a,b)$  where  $a+b \geq n/2-1$ . A central triangle can be only of certain types and the number of the central triangles in one dissection can be only one or two.

Since every diagonal  $d$  in  $\Delta$  is a side of two triangles, the sum of the two labels of  $d$  equals  $n$ . Thus, the number of occurrences of the label  $i$  in  $\Delta$  is equal to the number of occurrences of the label  $n-i$  in  $\Delta$  for  $i=1, \dots, n-1$ . These equalities and the requirement of the central triangle imply the following system of linear inequalities.

For  $(a,b) \in I_n, i=0, \dots, \lfloor (n-3)/2 \rfloor$  let

$$\zeta_{a,b}(i) = \begin{cases} 0, & \text{if } a \leq i \text{ and } b > i \text{ or if } a+b \leq i-1, \\ 1, & \text{if } a \geq i+1, \\ -1, & \text{if } a \leq i, b \leq i, \text{ and } i \leq a+b < n-1-i, \\ -2, & \text{if } a \leq i, b \leq i, \text{ and } a+b \geq n-1-i. \end{cases} \quad (3)$$

Then the cut set Catalan number  $S_n$  equals the number of solutions in nonnegative integers of the following system of  $\lfloor (n+1)/2 \rfloor$  linear inequalities with the  $\lfloor ((n-1)^2+3)/12 \rfloor$  variables  $\lambda_{a,b} \leq \lfloor (n+2)/(a+b+2) \rfloor, (a,b) \in I_n$ :

$$\begin{aligned} 0 &\leq 2 + \sum_{(a,b) \in I_n} \zeta_{a,b}(i) \lambda_{a,b} \quad \text{for } i=0, \dots, \lfloor (n-3)/2 \rfloor, \\ 0 &\leq 2 - \sum_{\substack{(a,b) \in I_n \\ a+b=n/2-1}} \lambda_{a,b} - 2 \sum_{\substack{(a,b) \in I_n \\ a+b \geq n/2}} \lambda_{a,b} \quad \text{if } n \text{ is even,} \\ 0 &\leq 1 - \sum_{\substack{(a,b) \in I_n \\ a+b \geq (n-1)/2}} \lambda_{a,b} \quad \text{if } n \text{ is odd.} \end{aligned} \quad (4)$$

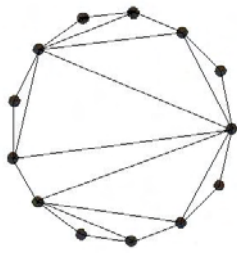


Fig. 1. A dissection of 13-gon

Thus, the values of  $S_n$  have been determined for  $n \leq 29$  without the need for examination of dissections of a polygon. The values for  $S_n$  known up to now are given in Table I.

TABLE I  
THE VALUES OF  $S_N$

n	2	3	4	5	6	7	8	9	10	11	12	13
$S_n$	1	1	2	2	4	6	11	17	35	57	115	203

N	14	15	16	17	18	19	20
$S_n$	412	745	1546	2838	5901	11154	23255

N	21	22	23	24	25	26
$S_n$	44263	93169	179214	377441	733151	1547068

n	27	28	29
$S_n$	3020878	6399874	12543862

In the system of linear inequalities two inequalities, for  $i=0$  and  $i=1$ , can be deleted because they follow from the other inequalities. For example the system of linear inequalities for  $n=11$  gives the value  $S_{11}=57$ :

$$\begin{aligned}
 0 &\leq 2 - \lambda_{1,1} - \lambda_{1,2} - \lambda_{2,2} + \lambda_{3,3} \\
 0 &\leq 2 - \lambda_{1,2} - \lambda_{1,3} - \lambda_{2,2} - \lambda_{2,3} - \lambda_{3,3} \\
 0 &\leq 2 - \lambda_{1,3} - \lambda_{1,4} - \lambda_{2,2} - \lambda_{2,3} - 2\lambda_{2,4} - 2\lambda_{3,3} \\
 0 &\leq 1 - \lambda_{1,4} - \lambda_{2,3} - \lambda_{2,4} - \lambda_{3,3}
 \end{aligned} \tag{5}$$

### III. THE METHODS FOR COMPUTING $S_N$

The determination of values of  $S_n$  is doing with the aid of computer programs in C. The first program is already presented in [1] and it gives  $S_n$  for  $n \leq 19$ . It counts the number of solutions of a linear system of equalities. In [2] the method is improved and we succeeded in evaluating  $S_n$  for  $20 \leq n \leq 25$ . This program counts the number of solutions of a linear system of inequalities from Eq. (4) in nonnegative integers. A equivalent set of the system is constructed and it has a smaller number of inequalities. At this time the evaluation for  $S_{25}$  took about 100 hours. By the new values  $S_n$  for  $26 \leq n \leq 29$  the algorithm is once again improved [4]. The count the number of solutions we proceeded here in two steps.

**Step 1.** We first considered the subsystem of a system in Eq. (4) consisting of the inequalities concerning  $i = \lfloor (n-1)/3 \rfloor, \dots, \lfloor (n-3)/2 \rfloor$  and the last inequality. From

$i \geq \lfloor (n-1)/3 \rfloor$  follow  $a \leq \lfloor (n-1)/3 \rfloor$ ,  $a < i$  and  $\zeta_{a,b}(i) \leq 0$ . We used a computer algebra system to determine all solutions of the subsystem in nonnegative integers. This step could be done in a few seconds.

**Step 2.** Using a simple algorithm, for each solution of the subsystem we counted the number of possibilities to choose the remaining variables  $\lambda_{a,b}$  such that the inequalities for  $i = 0, \dots, \lfloor (n-1)/3 \rfloor - 1$  and  $0 \leq \lambda_{a,b} \leq \lfloor (n+2)/(a+b+2) \rfloor$  are satisfied. For  $n=28$  and  $n=29$  the calculation were parallelized and distributed among more computers.

The program determines not only  $S_n$  but also all the relevant dissections of a polygon into triangles. The solutions of the linear system Eq. (4) give the numbers of the inner triangles and for every dissection the numbers of boundary triangles are determined from a system of equalities, see [2]. How can be composed the dissection corresponding to this set of triangles? We start with the central triangle or two central triangles (possible for  $n$  even). For every diagonal we draw another triangle (every diagonal is side in two triangles and has two labels). This procedure ends successfully because the relations between the labels in the system of inequalities.

This is one of the solutions of a system from Eq. (5):  $\lambda_{0,0}=5$ ,  $\lambda_{1,2}=\lambda_{0,1}=2$ ,  $\lambda_{1,4}=\lambda_{0,4}=1$ ,  $\lambda_{1,1}=\lambda_{1,3}=\lambda_{2,2}=\lambda_{2,3}=\lambda_{2,4}=\lambda_{3,3}=\lambda_{0,2}=\lambda_{0,3}=0$ . The corresponding dissection is given in Fig. 1.

### IV. CONCLUSION

At the time we work on the systems of inequalities for  $S_{30}$  and  $S_{37}$ . Based on a special classification of the dissections and the recursion formula from [2] we hope to get new values.

The cut set Catalan numbers are related to the other research areas: the maximum and minimum number of incongruent triangles in a dissection of regular convex polygon, see [3], the number of dissections by given number of triangle types, the numbers  $S_n(i)$  of cut set Catalan numbers with inner triangles etc.

### ACKNOWLEDGEMENT

This paper was supported by the project "Modern tendencies in the development of software technologies and algorithms, data processing and adequate training of specialists in these areas".

### REFERENCES

- [1] J. Dochkova, H. Harborth, I. Mengersen, "Cut set Catalan numbers", Congr. Numer. 130, pp. 133-139, 1998.
- [2] J. Dochkova, I. Mengersen, "Some new results on cut set Catalan numbers", Utilitas Math. 64, pp.159-166, 2003.
- [3] J. Dochkova, I. Mengersen, "Triangle dissections of convex polygons", Utilitas Math. 68, pp.255-269, 2005.
- [4] J. Dochkova, M. Krone, I. Mengersen, "New values for cut set Catalan numbers", Utilitas Math. (submitted).
- [5] The On-Line Encyclopedia of Integer Sequences, <http://oeis.org>

# Accelerating Strategies in Evolutionary Algorithms

Vassil G. Guliashki<sup>1</sup>, Leoneed Kirilov<sup>2</sup>

**Abstract** – We present three strategies designed to accelerate the convergence during the search process of evolutionary algorithms for convex integer optimization problems. The strategies realize a systematic diversification of the search. They are compared with performance of scatter search and particle swarm optimization.

**Keywords** – evolutionary algorithms, convex integer optimization problems, accelerating strategies.

## I. INTRODUCTION

We consider the convex integer programming problem in the form:

$$\begin{aligned} \text{Min } & F(x) & (1) \\ \text{subject to: } & g_i(x) \leq 0; \quad i = 1, \dots, m; & (2) \\ & l_j \leq x_j \leq u_j; \quad j = 1, \dots, n; & (3) \\ & x \in \mathbb{Z}^n, & (4) \end{aligned}$$

where  $x$  is an  $n$ -dimensional vector of integer variables  $x_j$ ,  $j = 1, \dots, n$ . By  $l_j$  and  $u_j$  are denoted the bounds (lower and upper) of  $x_j$ , and  $F(x)$  is the multimodal objective function.  $F(x)$  may not possess derivatives in an explicit analytical form. The functions  $g_i(x)$ ,  $i = 1, \dots, m$ ; are convex nonlinear functions and  $m$  is the number of nonlinear constraints (2).

The convex integer problems (see [6, 18]) belong to the class of NP-hard optimization problems. 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 (see [8, 11, 14, 17, 19, 21]).

To solve problem (1-4) many algorithms which mimic the natural evolution process of species have been designed in order to obtain a global optimum. They could be classified as “evolutionary” or “population based” algorithms (see [14]). The most familiar and powerful among them are Genetic Algorithms (GA) (see [11, 15]), Scatter Search (SS) (see [7, 9]), Tabu Search (TS) (see [8, 9, 10]), Ant Systems (AS) (see [1, 2, 3]) and Particle Swarm Optimization (PSO) (see [5, 16, 17]). The evolutionary algorithms use a population of feasible solutions (or characteristics of solutions), called *individuals*, *trial (dispersed) points*, *ants*, *particles* etc. In this paper is used the term *individuals*.

The evolutionary algorithms usually use an improvement sub-procedure to intensify the search process in some regions

of the search space. One such sub-procedure is the local search. It is supposed that each intensification period finishes with a found local optimum. To escape from the local optimality a diversification of the search process is necessary after the intensification period.

To be efficient an evolutionary algorithm for search a global optimal solution it should quickly perform the diversification of the search. Different ways for diversification of the search have been developed. For example, during the diversification phase the individuals could be modified independently – like the mutation in GA. But the results are unexpected in the sense that the modification does not lead necessarily to an improvement. A famous example for diversification is the Tabu list strategy used in the Tabu search algorithms. Some characteristics of solutions or movements (steps in given directions) are stored as forbidden (tabu) for certain number of iterations. In this manner the cycling and the trap of local optimality are avoided. Successful diversification of the search process is the use of non-convex combinations of parent solution vectors. In this way individuals that lie in new regions are systematically generated (see [7, 9]).

To achieve good convergence speed the successful global search methods combine usually two or more metaheuristics in hybrid methods. For example GA are combined with Tabu Search methods, or with a faster local search procedure, AS – with local search techniques (see [21]), GA – with clustering procedure (see [4]), SS – with TS or SS – with GA (see [9]). Tabu search can also be coupled with directional search approach. 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. This is an elitism – based approach for generating new offspring individuals (see for example [13]). A possible strategy is to combine the qualities of a directional type method with the good features of evolutionary algorithms. 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. Accelerating the systematic diversification is an open area for further development of search strategies.

In this paper three strategies for fast systematic diversification of the search process are proposed. The proposed accelerating strategies are described in Section II. An illustrative example is given in Section III. Some conclusions are drawn in Section IV.

## II. THE HEURISTIC ACCELERATING STRATEGIES

Considering the search process for global optimum there is

<sup>1</sup>Vassil G. Guliashki is with the Institute of Information and Communication Technologies – BAS, “Acad. G. Bonchev” Str. Bl. 2, 1113 Sofia, Bulgaria, E-mail: vggul@yahoo.com

<sup>2</sup>Leoneed Kirilov is with the Institute of Information and Communication Technologies – BAS, “Acad. G. Bonchev” Str. Bl. 2, 1113 Sofia, Bulgaria, E-mail: leomk@abv.bg

no reason the search to be directed to the region of the best found so far (local optimal or near optimal) solution, because in the most cases it will not coincide with the global optimal solution. The same is valid for all known local optimal solutions, as well for all explored regions of the feasible domain. The exploration of the whole feasible domain means that there is a guaranteed systematic diversification of the search process. A hybrid method performing systematically diversified search (SDS-method) by means of separating the feasible domain in sub-regions (cones having a common vertex) is proposed in [12]. The systematic diversification of the search consists in exploring the cones obtained one by one.

Let the feasible domain be denoted by  $X$  and let the Tchebicheff center (the point located at the maximal Euclidean distance from the constraint surfaces) be  $x_{tch} \in X$ . We assume that  $x_{tch}$  is obtained by means of a method for solving convex problems with continuous variables. Then  $x_{tch}$  is rounded off to the nearest integer point  $i_{tch}$ .

A. Wave-spreading strategy

**Step 1.** Generate a regular simplex with  $n+1$  vertices, using  $i_{tch}$  as one vertex. The other simplex vertices are generated in the following manner:

$$v^{(j)} = \begin{cases} i_{tch_j} + \varphi_1 & \text{if } j \neq i \\ i_{tch_j} + \varphi_2 & \text{if } j = i \end{cases} \quad i = 1, \dots, n; j = 1, \dots, n; \quad (5)$$

$$\varphi_1 = \alpha \cdot \left[ \frac{\sqrt{(n+1)} + n - 1}{n\sqrt{2}} \right] \quad (6)$$

$$\varphi_2 = \alpha \cdot \left[ \frac{\sqrt{(n+1)} - 1}{n\sqrt{2}} \right] \quad (7)$$

Let  $i_{tch}$  be denoted as  $v^{(0)}$ . Round off each  $v^{(j)}$ ,  $j = 1, \dots, n$ ; to its nearest integer point. There are  $(n+1)$  combinations of  $n$  vertices, correspondingly for each facet of the simplex.

**Step 2.** Calculate the components of the simplex weight center as follows:

$$cs_i = \frac{\sum_{j=0}^n v_i^{(j)}}{n+1}, \quad i=1, \dots, n \quad (8)$$

Round off each component  $cs_i$  to its nearest integer value.

**Step 3.** Create an initial population  $P_0$  around the weight center  $cs$ , containing  $p$  uniform distributed solution vectors, generated by using deviation of  $\pm\delta$ , where  $\delta$  is a constant % of corresponding component (for example  $\delta_{max} = \pm 1\%$ ).

**Step 4. Intensification phase**

Here is used a reflection like the idea in the simplex method by Nelder and Mead (see [20]).

**a)** Order the individuals (solutions) in the current population in increasing order of their  $F$ -values.

**b)** Calculate the weight center  $P_c$  of first  $k$  individuals:

$$P_c = \frac{1}{k} \sum_{j=1}^k x^{(j)} \quad (9)$$

Here  $k$  is chosen to cover about 10% to 50% of the individuals in the population.

**c)** Let  $x_{w1}, \dots, x_{wk}$  be the individuals in the current population having the worst (i.e. the greatest)  $F$ -values. Calculate the steps:

$$y_i = P_c - x_{wi}, \quad i = 1, \dots, k; \quad (10)$$

**d)** Reflect the  $k$  worst individuals towards  $P_c$  to generate  $k$  new individuals (solutions):

$$x_{newi} = P_c + y_i, \quad i = 1, \dots, k. \quad (11)$$

Round off each  $x_{newi}$  is rounded off to its near integer point. In case someone new solution is infeasible, i.e. the constraints (2)-(3) are violated, restrict the step length:

$$y = \theta y, \quad (12)$$

where  $\theta \in (0,1)$ .

**e)** In case someone of the so generated individuals is better than one of the current population, the better individual replaces the worse. If there aren't generated better individuals continue by **Step 5**, otherwise go to **a**).

**Step 5. Diversification phase**

**a)** Make step  $\beta(v^{(j)} - cs)$  along each ray starting at  $cs$  and passing through the simplex vertices  $v^{(j)}$ ,  $j = 0, \dots, n$ ; in outside direction, so that the new central solutions  $cs^{(j)}$ ,  $j = 0, \dots, n$ ; are generated.

**b)** Around each point  $cs^{(j)}$  are generated  $p$  uniform distributed solutions' vectors like in **Step 3** and build  $(n+1)$  new populations  $P_j$ .

**c)** Perform the *Intensification phase* for each new population  $P_j$ ,  $j = 0, \dots, n$ .

**d)** Make step  $\beta(cs - v^{(j)})$  along each ray starting at  $cs$  in the opposite of sub-step **a**) direction, so that the new points  $cs^{(j)}$ ,  $j = 0, \dots, n$ ; are generated. Perform the sub-steps **b**) and **c**).

**Step 6.** Alternate the *Diversification* and the *Intensification* phase in the same way until reaching the boundaries of the feasible region.

**Step 7.** Perform simple local search around each found locally optimal solution to precisely locate all found optima.

**REMARKS:**

The initial simplex gets larger and larger in the search space like a wave raised by a stone in a lake.

The parameter  $\beta$  depends on the size of feasible region. For relative small domains the greatest component of  $\beta(v^{(j)} - cs)$  is 10% of the greatest among the values  $Q_j = u_j - l_j$ , for  $j = 1, \dots, n$ ; For larger domains  $\beta$  should be chosen smaller.

For large feasible domains also the rays passing through  $cs$  and through each of the  $(n+1)$  weight centers of simplex vertices determining each simplex facet should be explored in the way described above.

B. Slicing strategy

In this strategy the feasible domain will be separated (sliced) in  $t$  sub-regions as follows:

**Step 1.** Compare the values  $Q_j = u_j - l_j$ , for  $j = 1, \dots, n$ ; and find out the maximal value  $Q_j^{(max)}$  for fixed  $j = j_{max}$ . Let  $q$  be the integer part of  $Q_j^{(max)}/t$ :

$$q = \left\lfloor \frac{Q_j^{(max)}}{t} \right\rfloor \quad (13)$$

and let  $l_j^1 = l_j$ ,  $u_j^1 = l_j^1 + q - 1$ ,  $l_j^i = u_j^{i-1} + 1$ ;  $u_j^i = l_j^{i-1} + q - 1$ ; for  $i=2, \dots, t-1$ , and  $l_j^t = u_j^{t-1} + 1$ ;  $u_j^t = u_j$ , where  $j = j_{\max}$ .

**Step 2.** Divide the constraint system (3) into  $t$  constraint sub-systems:

$$\begin{aligned} l_j \leq x_j \leq u_j; \quad j = 1, \dots, m; j \neq j_{\max}; \\ l_j^i \leq x_j \leq u_j^i; \quad j = j_{\max}; i = 1, \dots, t. \end{aligned} \quad (14)$$

Each sub-region is defined by the constraint systems (2), (4) and by one constraint sub-system from (14).

**Step 3.** Perform *diversification* of the search process by going from one sub-region to another, generating the initial population at random with uniform distribution around the Tchebicheff center of the current sub-region. Then perform the *Intensification phase* described in *Wave – spreading strategy* (Step 4.) in each sub-region.

C. Hybrid strategy

This strategy consists in slicing the feasible domain in the way described in *Slicing strategy*. After that the search procedure performs a *Wave – spreading strategy* in each sub-region.

III. ILLUSTRATIVE EXAMPLE

Let us consider the following two-dimensional example.

Five sub-areas in the feasible domain are defined:

$$A1 = \{0 \leq x_1, 0 \leq x_2, 21x_1 + 20x_2 - 84000 \leq 0\}$$

$$A2 = \{0 \leq x_1, x_2 \leq 10000, 0 < 21x_1 + 20x_2 - 84000, 7x_1 - 5x_2 + 15000 \leq 0\}$$

$$A3 = \{x_1 \leq 7200, 0 \leq x_2, x_2 \leq 10000, 0 < 21x_1 + 20x_2 - 84000, 0 < 7x_1 - 5x_2 + 15000\}$$

$$A4 = \{7200 < x_1, x_1 \leq 10000, x_2 \leq 5900, 0 \leq x_2\}$$

$$A5 = \{7200 < x_1, x_1 \leq 10000, 5900 < x_2, x_2 \leq 10000\}$$

The optimization problem is:

$$\begin{aligned} \text{Min } F(x) = & \begin{cases} 10 + (x_1 - 2500)^2 + (x_2 - 1000)^2 & \text{if } (x_1, x_2) \in A1; \\ 7 + (x_1 - 1500)^2 + (x_2 - 7000)^2 & \text{if } (x_1, x_2) \in A2; \\ 12 + (x_1 - 6100)^2 + (x_2 - 3400)^2 & \text{if } (x_1, x_2) \in A3; \\ 11 + (x_1 - 9800)^2 + (x_2 - 2100)^2 & \text{if } (x_1, x_2) \in A4; \\ 3 + (x_1 - 8100)^2 + (x_2 - 9700)^2 & \text{if } (x_1, x_2) \in A5; \end{cases} \end{aligned}$$

$$\text{subject to:} \quad \begin{aligned} 0 \leq x_1 \leq 10000; \\ 0 \leq x_2 \leq 10000; \end{aligned}$$

This problem has five local optima – one per each sub-area:  $x^{(1*)} = (2500, 1000)$ ,  $x^{(2*)} = (1500, 7000)$ ,  $x^{(3*)} = (6100, 3400)$ ,  $x^{(4*)} = (9800, 2100)$ ,  $x^{(5*)} = (8100, 9700)$ ;

The corresponding objective function values are:  $F(x^{(1*)}) = 10$ ;  $F(x^{(2*)}) = 7$ ;  $F(x^{(3*)}) = 12$ ;  $F(x^{(4*)}) = 11$ ;  $F(x^{(5*)}) = 3$ . Hence the global optimal solution is  $x^{(5*)}$ .

*Wave-spreading strategy*

Starting at the Tchebicheff center  $i_{tch} = (5000, 5000)$  the simplex with vertices (5000, 5000), (5002.588, 5009.659) and (5009.659, 5002.588) is generated. The weight center of the simplex is  $cs = (5004, 5004)$ .

The population  $P^{(0)}$  includes 10 points (individuals):

$$\begin{aligned} x^{(1)} = (4994, 4994), x^{(2)} = (5004, 4994), x^{(3)} = (5014, 4994), \\ x^{(4)} = (4985, 5004), x^{(5)} = (4005, 5004), x^{(6)} = (5005, 5004), \\ x^{(7)} = (5015, 5004), x^{(8)} = (4994, 5014), x^{(9)} = (5004, 5014), \\ x^{(10)} = (5014, 5014); \end{aligned}$$

The corresponding objective function values are:

$$\begin{aligned} F(x^{(1)}) = 3764084, F(x^{(2)}) = 3742064, F(x^{(3)}) = 3720244, \\ F(x^{(4)}) = 3816053, F(x^{(5)}) = 3793853, F(x^{(6)}) = 3771853, \\ F(x^{(7)}) = 3750053, F(x^{(8)}) = 3828244, F(x^{(9)}) = 3806224, \\ F(x^{(10)}) = 3784404; \end{aligned}$$

For  $k = 10\%$  we choose the best individual:  $x^{(3)}$ . The worst individuals are  $x^{(8)}$ ,  $x^{(4)}$  and  $x^{(9)}$ . The worst individuals are reflected towards  $x^{(3)}$ . Three new better individuals are generated and they replace the worst individuals  $x^{(8)}$ ,  $x^{(4)}$  and  $x^{(9)}$ . Proceeding in this way until no better individuals are generated, and then performing a simple local search, the procedure finds out the locally optimal solution  $x^{(3*)} = (6100, 3400)$  with objective function value  $F(x^{(3*)}) = 12$ .

The generated simplex has the following rounded off vertices:  $v^{(0)} = (5000, 5000)$ ,  $v^{(1)} = (5003, 5010)$  and  $v^{(2)} = (5010, 5003)$ . We will consider the performance of this strategy along one exploring ray, say  $(v^{(0)} - cs) = (-4, -4)$ . Proceeding with  $\beta = 250$  at  $v^{(0)}$  the search procedure creates consecutively 5 initial populations around the calculated central solutions  $cs^{(1)} = (4000, 4000)$ ,  $cs^{(2)} = (3000, 3000)$ ,  $cs^{(3)} = (2000, 2000)$ ,  $cs^{(4)} = (1000, 1000)$ ,  $cs^{(5)} = (0, 0)$ . The last population reaches the boundaries of the feasible domain, so that this direction is explored. The third generated population around  $cs^{(3)} = (2000, 2000)$  comes in the sub-area A1, so that the intensification phase finds out the optimum  $F(x^{(1*)})$ . The same is repeated with the fourth and fifth generated population.

Then the procedure explores the opposite direction, creating again five initial populations around the calculated central solutions  $cs^{(6)} = (6000, 6000)$ ,  $cs^{(7)} = (7000, 7000)$ ,  $cs^{(8)} = (8000, 8000)$ ,  $cs^{(9)} = (9000, 9000)$ ,  $cs^{(10)} = (10000, 10000)$ . The last three populations come in the sub-area A5, so that the intensification phase finds out the optimum  $F(x^{(5*)})$ .

Going on along the other two exploring rays in both possible directions the search procedure finds out also the optima  $F(x^{(4*)})$  and  $F(x^{(2*)})$ .

During the exploration of whole feasible domain 31 intensification phases are performed.

*Slicing strategy*

We choose the component  $x_1$  as slicing component. The created sub-areas are: B1:  $0 \leq x_1 < 1000$ ;  $0 \leq x_2 < 10000$ ;

$$B2: 1000 \leq x_1 < 2000; 0 \leq x_2 < 10000; \dots$$

$$B10: 9000 \leq x_1 \leq 10000; 0 \leq x_2 < 10000;$$

In each sub-area is generated an initial population randomly with uniform distribution. Performing the above described *intensification phase* in each sub-area the search procedure finds out all locally optimal solutions. To explore the whole feasible domain 10 intensification phases are performed. Some of them are more time-consuming than the intensification phases performed by *Wave-spreading strategy*.

*Hybrid strategy*

This strategy also finds out all possible local optima. Here the value of parameter  $\beta$  remains the same like in the *Wave-spreading strategy*, because the component  $x_2$  keeps its variation interval unchanged. This leads to great steps along the exploring rays and in some directions already the first generated initial population is infeasible. For this reason  $\beta$  should be reduced in half and this is repeated until the generating a feasible population becomes possible.



#### Comparisons with other algorithms

The described strategies are compared with scatter search and with particle swarm optimization. Starting with the initial population  $P^{(0)}$  these two algorithms are able to find out only one locally optimal solution, and this is  $x^{(3*)}$ . Better performance is achieved in case the initial population is enough dispersed. The scatter search has better chances to find out the global optimal solution in case the limit of iterations is large. In this case it needs more than 1000 generations.

The presented problem may be solved by genetic algorithm, using niches. In this way all optima can be found but this performance will be essentially more time consuming than the proposed solution procedures.

#### IV. CONCLUSIONS

The presented strategies for fast systematic diversification of the search have the following advantages:

- They systematically diversify the search process, avoiding in this manner the trap of local minima.
- They explore roughly the whole feasible domain and have good chances to find out the global optimal solution.
- The applying of local search technique at the end of the search process guarantees the good quality of the obtained solution.
- The proposed strategies have a better convergence to the global optimum in comparison to other global search algorithms, in which the search process does not perform a systematic diversification.
- They are simpler and don't require large computer memory and complex memory organization in comparison to other global search strategies like Tabu search.
- They use populations with relatively small size. This makes them efficient in solving large dimensional problems.
- The proposed strategies are easy for computer programming implementation.

#### 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] Colomi A., M. Dorigo and V. Maniezzo, "Positive Feedback as a Search Strategy", Department of Electronics, Politecnico di Milano, Italy, Working paper 91-16, 1991.
- [2] Colomi A., M. Dorigo and V. Maniezzo, "Distributed Optimization by Ant Colonies", in Proceedings of the *First European Conference on Artificial Life (ECAL-91)*, F. J. Varela and P. Bourguine eds., The MIT Press: Cambridge, MA, 1991, pp. 134-142.
- [3] Colomi A., M. Dorigo and V. Maniezzo, "An Investigation of Some Properties of an Ant Algorithm", in *Parallel Problem Solving from Nature 2*, R. Männer and B. Mandrick eds., North-Holland: Amsterdam, 1992, pp. 509-520.
- [4] 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.
- [5] Eberhart R. C. and J. Kennedy, "A new optimizer using particle swarm theory", *Proceedings of the Sixth International Symposium on Micromachine and Human Science*, Nagoya, Japan, pp. 39-43, 1995.
- [6] Garey M. R. and Johnson D. S., "Computers and Intractability: A Guide to the Theory of NP-Completeness", W. H. Freeman, San Francisco 1979.
- [7] Glover F., "Heuristics for integer programming using surrogate constraints", *Decision Sciences* 8, 1977, pp. 156-166.
- [8] Glover F., "Future path for integer programming and links to artificial intelligence", *Computers & Operations Research* 13, 1986, pp. 533-549.
- [9] Glover F., "Tabu Search for Nonlinear and Parametric Optimization (with Links to Genetic Algorithms)", *Discrete Applied Mathematics*, "Viewpoints on Optimization", 1991.
- [10] Glover F., "Tabu search and adaptive memory programming – advances, applications and challenges", in *Advances in Metaheuristics, Optimization and Stochastic Modeling Technologies*, Barr, Helgason and Kennington eds., Kluwer: Boston, MA., 1997, pp. 1-75.
- [11] Goldberg D. E. *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison Wesley, Reading, Mass, 1989.
- [12] Guliashki, V., (2007) "A Hybrid Population Based Method Solving Convex Integer Optimization Problems", In: *Proceedings of ICEST2007*, (Editor Prof. Dr. Mitrovski, C.), June 24 – 27, 2007, Ohrid, Macedonia, vol. I, pp. 249-252.
- [13] Guliashki V., C. Korsemov, H. Toshev, (2010) "Elitism Based Evolutionary Algorithm for Discrete Optimization Problems", In: *Proceedings of ICEST2010*, (Editor Prof. Dr. Mitrovski, C.), June 23 – 26, 2010, Ohrid, Macedonia, vol. I, ISBN: 978-9989-786-57-0, pp. 281-284.
- [14] 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.
- [15] Holland J. H., "Adaptation in Natural and Artificial Systems", The University of Michigan Press: Ann Arbor, MI, 1975, MIT Press, 1992.
- [16] Kennedy J. and R. C. Eberhart, "Particle Swarm Optimization", *Proceedings of IEEE International Conference on Neural Networks*, Piscataway, N. J., 1995, pp. 1942-1948.
- [17] Krusienski D. J. and Jenkins W. K. "Design and performance of adaptive systems, based on structured stochastic optimization strategies", *IEEE Circuits and Systems*, Vol.5, No.1, first quarter 2005, pp.8-20.
- [18] Nemhauser G. L. and Wolsey L. A. "Integer and Combinatorial Optimization", Wiley, New York. 1988.
- [19] Pirlot M., "Heuristic Search Methods", *Oper. Res. Designing Practical Solutions; Tutorial and Research Review Papers*, Euro XIII/OR 36, The joint EURO/Oper. Res. Society Conference; Univ. of Strathclyde, Glasgow, 19-22, VII. 1994. pp. 180-201.
- [20] Reklaitis G.V., Ravindran A., Ragsdell K. M., *Engineering Optimization. Methods and Applications*, John Wiley & Sons, 1983.
- [21] 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.

# 2D Weather product visualization using Marching Squares algorithm

Igor Antolović<sup>1</sup>, Dejan Rančić<sup>2</sup>, Vladan Mihajlović<sup>3</sup>, Dragan Mihić<sup>4</sup>, Marija Đorđević<sup>5</sup>

**Abstract** – In this paper we present a efficient way for 2d weather product visualization by using an approach which is based on the *Marching Squares* algorithm. We examine the possibilities of extracting both isolines and color filled isoareas as well as improvements in order to avoid the default broken polyline *Marching Squares* result and force more smooth visualization. By using different sampling resolutions and parameters we test and evaluate this approach on real weather temperature products.

**Keywords** – Isoline, Marching squares, besier curves

## I. INTRODUCTION

Modern climatology depends heavily on weather tracking GIS (Geographic Information System) [1] applications and their ability to process, analyze and visualize weather data. This kind of systems range from simple applications for processing of climate data to more complex weather forecasting systems [2]. The nature of weather data can be vector (e.g. wind) or scalar (temperature, pressure, precipitation etc.) thus emphasizing the need for various visualization algorithms aimed towards both 2D and 3D data as well as vector and scalar data types. HAISIS (Integrated Hail Suppression Information System) [3,4,5] is one such system. It is developed at the Faculty of Electronic Engineering in Niš, at the Computer Graphic and Geographic Information Laboratory (CG&GIS Lab) to satisfy requirements of Republic Hydrometeorological Service of Serbia (RHMS). It is operating more that 10 years and it supports LIC (Line Integral Convolution) [6] methods for vector data visualization, 2D isoline tracing, 3D isosurface extraction, 3D wind visualization and various other

visualization algorithms offering climatology experts powerful tools for weather/climate tracking and analysis.

This paper focuses explicitly on visualization of scalar data using the *Marching Squares* algorithm considering it can be directly applied on 2D scalar weather product visualization. Although the most common version of this algorithm is the classic linear segment approach, we will further cover possible besier curve advancements in order to generate smooth isolines as well as the extended *Marching Squares* version suited for extraction of color filled isoareas.

Considering that the isoline extraction and visualization concept represents a scalar field visualization problem in the second part of this paper we give a short review of available techniques for scalar field visualization. The third part gives a detailed description of the *Marching Squares* algorithm followed by the fourth part which presents practical implementation results in combination with mean temperature test data. Finally the fifth part consists of conclusions and notes on future work.

## II. 2D SCALAR FIELD VISUALIZATION CONCEPTS

A discrete 2D scalar field is essentially a (mxn) grid where every grid cell takes a value from a real set of numbers:

$$S : (x_0, \dots, x_m) \times (y_0, \dots, y_n) \rightarrow \mathbb{R} \quad (1)$$

There are basically two types of 2D scalar data visualization techniques: color mapping and isoline extraction as shown on Fig.1.

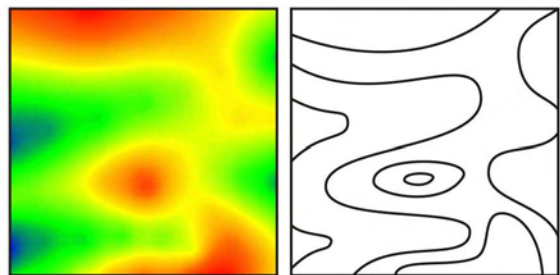


Fig.1. Color mapping (left) and isoline extraction (right)

Color mapping approach can be most easily implemented and uses a quite straightforward approach. The basic idea behind color mapping is to traverse through all scalar field values and map those values into colors. Additionally color values can be interpolated in order to get a smoother output. Usually instead of RGB (Red Green Blue) color model, HSB (Hue Saturation Brightness) is used considering that it defines more practical color scheme. Also an interesting

<sup>1</sup>Igor Antolović is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: igor.antolovic@elfak.ni.ac.rs.

<sup>2</sup>Dejan Rančić is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: dejan.rancic@elfak.ni.ac.rs.

<sup>3</sup>Vladan Mihajlović is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: vladan.mihajlovic@elfak.ni.ac.rs.

<sup>4</sup>Dragan Mihić is with the Republic Hydrometeorological Service of Serbia and South East European Virtual Climate Change Center, E-mail: dragan.mihic@hidmet.gov.rs

<sup>5</sup>Marija Đorđević is with the Republic Hydrometeorological Service of Serbia and South East European Virtual Climate Change Center, E-mail: mapuja@gmail.com

recommendations on using colors with climate data can be found in [7].

On the other hand when isoline extraction is considered it is necessary to start from its very definition. For a given isovalue (c) an isoline consists of a set of points defined by:

$$I = \{(x, y) | S(x, y) = c\} \tag{2}$$

Meaning that isolines consists of all points that share a constant isovalue (c). On the other hand when it comes to practical isoline extraction there are two basic methods:

- Contour tracing
- Isoline segment extraction using *Divide and Concur approach*

Contour tracing is the most intuitive isoline extraction method and it is based on the idea of tracing values on a scalar field grid from a single seed point until a boundary is reached or a contour loop is detected. Recent work on this can be found in [8]. By using this method a continuous poly-line is produced which can be later additionally interpolated and smoothed. The common problem in this approach can be found in the process of seed point determination. Also in the case of multiple contour tracking, multiple track markers must be used. As previously stated, one practical application of this method can be found also in the HASIS 3DI system where this approach is used to extract dBz contours for given isolevels.

The *Divide and Concur* approach relies on the fact that the problem of isoline extraction can be simplified by extracting isoline segments in single scalar field cells and afterwards join those segments into continuous isolines [9]. This method is applicable for 3D as well 2D scalar fields. The 3D case of this method is well known by the name of *Marching Cubes* [10] and it is used for fast extraction of isosurfaces. In this paper we will further explore and describe it's simplified 2D version also known as the *Marching Squares* algorithm.

### III. MARCHING SQUARES

The main idea behind the *Marching Squares* algorithm is to simplify the isoline extraction and visualization procedure by processing scalar field cells one at a time.

It can be easily concluded that the values in the cell corners can be above or below the given iso-value. Considering this fact a set of 16 possible cases can be isolated as shown on Fig. 2. Furthermore a simple scheme can be involved in order to index these cases where 0 means the corner value is below the required isovalue and 1 means the corner value is above the required isovalue.

This coding scheme can be used to precalculate index tables in order to drastically speedup the process of identifying the current cell case.

The steps of processing the current cell can be described as follows:

- Compare the current cell corners against the required isovalue and determine the cell index.

- Use the calculated index in order to extract information of which edges are intersected by the isoline from the precalculated edge table.
- By using linear interpolation determine the exact points of intersection of the isoline and the cell edges.
- Connect the intersection points in order to get the current isoline segment.

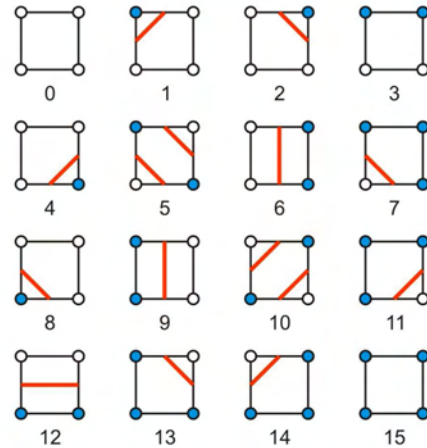


Fig.2. Marching Squares 16 possible cases

Upon processing all cells and drawing all isoline segments a full isoline set will be created as shown on Fig.3.

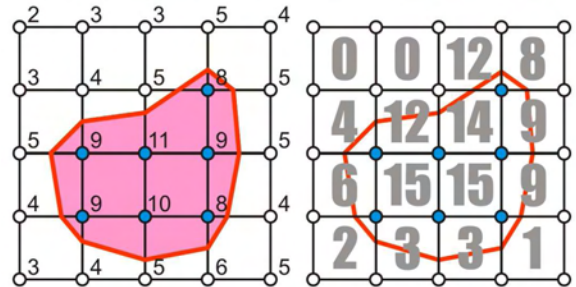


Fig.3. Example of isoline extraction scheme for isovalue (7)

This approach is very fast but it has two main disadvantages:

- The quality of the generated isolines depends on the resolution of the grid. Since the processing of a single cell produces line segments it is obvious that on small resolution grid isolines will be broken polylines.
- The second disadvantage is that there is no information about a continuous isoline.

It is obvious that for fixed edge intersection points there can be lot of solutions for the isoline segments but it is necessary to use the approach which provides the smoothest output.

On order to increase the isoline quality a common method is to take advantage of the gradient property of the scalar field. Basically the gradient is a vector which is pointed towards the greatest differential change in the given point of a scalar field.

Let the 2D scalar field be defined as:

$$S = f(x, y) \tag{3}$$

The gradient can be calculated as follows:

$$\nabla S = i \frac{\partial S}{\partial x} + j \frac{\partial S}{\partial y} \tag{4}$$

In the practical case where a discrete scalar field is used the gradient components can be calculated by using central differences as follows:

$$g(x, y)_x = \frac{S(x+1, y) - S(x-1, y)}{2\Delta x} \tag{5}$$

$$g(x, y)_y = \frac{S(x, y+1) - S(x, y-1)}{2\Delta y} \tag{6}$$

By using these equations we can easily calculate gradient vectors of the cell corners and furthermore calculate gradient vectors on isoline and grid intersection points by using linear interpolation. The most interesting property of the gradient vector is that it is perpendicular to the isoline passing through a given point, meaning that the vector normal to the gradient vector acts as a tangent vector of the isoline. This information can be used to construct smooth isoline segments instead of just using straight lines. In order to accomplish this, besier curves can be used considering the flexible relation between besier control points and besier curve normals.

The process of smooth isoline segment (Fig. 4) construction can be described in several steps as follows:

- Calculate gradient vectors in cell corner points,
- Calculate gradient vectors on isoline intersection points by interpolating between corner gradient vectors,
- Calculate normal vectors of previously calculated isoline gradient vectors,
- Find normal vector intersection M,
- Calculate besier control points as half distances between isoline intersection points and normal intersection point M.

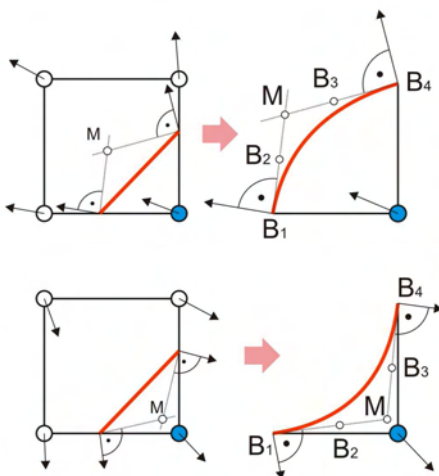


Fig.4. Besier curve construction examples

By processing all cells in this way and generating besier isoline segments, a drastic increase in quality can be noticed as the isolines show smooth flow through the cells.

Finally a simple polygonal scheme can be derived from the original *Marching Squares* set as shown on Fig.5. This new set can be used to generate color filled isoareas. By combining multiple layers of isoareas a multicolor isoarea can be generated as shown on Fig.6. The only problem with this approach is that it is order dependent. It is obvious that the isoareas must be sorted in increasing isovalues going from bottom to top.

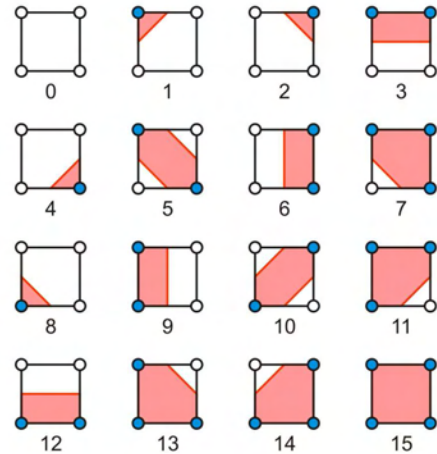


Fig.5. Polygonal Marching Squares scheme

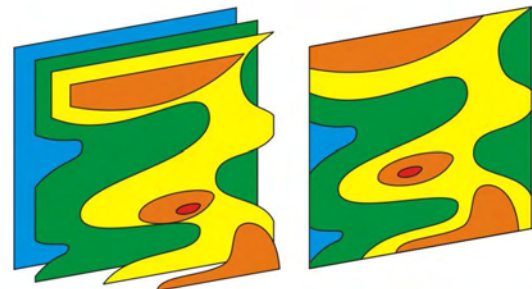


Fig.6. Multilayered polygonal isolevel approach

#### IV. RESULTS

For testing purposes as an input data a average temperature dataset was used which was sampled on a 64x64 grid. In order to visually compare the quality of the generated isolines both classic linear and advanced besier approach was used. The results are shown on Fig. 7. What can easily be noticed is that for higher sampling resolutions as well small cell sizes the difference between the advanced smoothing and classic linear approach becomes indistinguishable. On the other hand for higher zoom levels the broken polyline appearance of the classic linear method becomes more obvious as shown on Fig. 7.c). Beside the classic isoline extraction, the multilayered isoarea approach was also implemented so depending on the need either of this visualization techniques can be used. For example in lots of cases a combination of isolines and terrain map is can be used as shown on Fig.7.b)

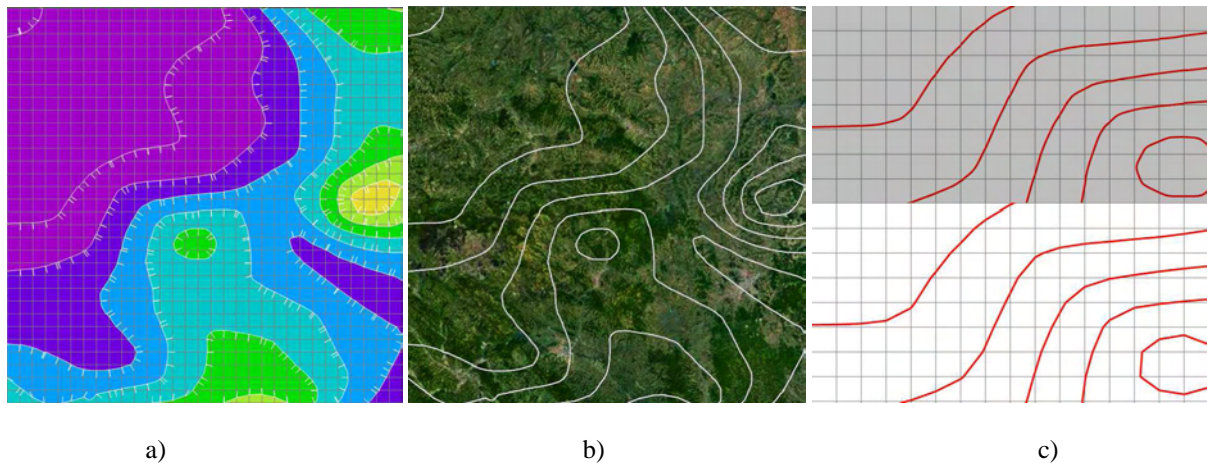


Fig.7. a) Smooth isoareas with normals b) Non-smooth isolines on map c) Comparison of smooth and non-smooth details

## V. CONCLUSION

Efficient visualization of climate data products is a basic requirement in all weather tracking and analysis GIS applications. Taking into account the nature of this data, two groups of climate variables can be distinguished. The first group consists of scalar climate variables (temperature, pressure, precipitation etc.) and the second group consists of vector climate variables (e.g. wind).

In this paper we considered visualization techniques for the first group of variables. This kind of visualization has already been identified as a scalar field visualization problem which is most commonly solved by using isoline extraction approach. Among two basic isoline extraction methods (by contour tracing and by using the *Marching Squares* divide and conquer method) we choose the second method considering that it relies on precalculated tables thus providing great speed as well as multiple isoline extraction.

As a proof of concept we implemented the *Marching Squares* Algorithm and additionally considered isoline segment smoothing by using besier curves. Finally we compared the results of classic linear *Marching Squares* approach with besier curve based results. As it is shown the classic linear approach suffers from a broken polyline look compared to the smooth besier based approach. Furthermore we compared isoareas and isoline visual results as well variants of isoline on map backgrounds. Upon evaluating this results we can conclude that the *Marching Squares* algorithm give a rather elegant, fast, and visually satisfying solution to the isoline extraction problem. What is also important to emphasize is that the final quality of the generated isolines depends directly on the sampling grid resolution thus the future work on this matter will include considerations of techniques for isoline extraction on lower sampling grid resolutions.

## ACKNOWLEDGEMENT

This paper was realized as a part of the project "Studying climate change and its influence on the environment: impacts,

adaptation and mitigation" (43007) financed by the Ministry of Education and Science of the Republic of Serbia within the framework of integrated and interdisciplinary research for the period 2011-2014.

## REFERENCES

- [1] D. Rančić, A. Dimitrijević, V. Mihajlović, "GIS and Virtual Reality Systems Integration", ICEST 2004, Bitola, Macedonia, pp. 313-316, 2004.
- [2] L. A. Treinish, "Visual data fusion for applications of high-resolution numerical weather prediction", *IEEE Proceedings of the conference on Visualization '00*, 2000, pp. 477 - 480.
- [3] V. Mihajlović, S. Đorđević-Kajan, D. Rančić, B. Predić, I. Antolović, P. Eferica, Z. Babić, "Architecture of HASIS-3D System Designed for Hail Suppression Purposes", Proceedings of ICEST 07, Ohrid, jun 2007.
- [4] Đorđević-Kajan, S., Milovanović, B., Rančić, D., Kostić, A., Stanić, Z., Stoimenov, L., "Hail Suppression Information System of Radar Center", GIS/LIS'96, Budapest Hungary, June 1996, pp.102-111.
- [5] Rančić, D., Smiljanić, M., Đorđević-Kajan, S., Kostić, A., Eferica, P., Vuković, P., Vučinić, Z., "Radar Data Processing for Cloud Seeding in Hail Suppression Information System", RADME 98
- [6] M. Kovačević, V. Mihajlović, I. Antolović, D. Rančić, Z. Babić: LIC based Visualization of Air Flow in Clouds, Conference on Computer Science and Information Technologies – YUINFO 2010, Kopaonik, Serbia, 3 – 6 March 2010. (In Serbian).
- [7] American Meteorological Society, Guidelines for using Color to Depict Meteorological Information: IIPS Subcommittee for Color Guidelines. 1993
- [8] Fu Chang, Chun-Jen Chen, Chi-Jen Lu, A linear-time component-labeling algorithm using contour tracing technique, *Computer Vision and Image Understanding*, v.93 n.2, p.206-220, February 2004
- [9] G. Cottafava, G. Le Moli, "Automatic Contour Plotting", *Comm. ACM* 12, 7 (July 1969), pp. 386-391.
- [10] Lorensen W, Cline H. Marching cubes: a high resolution 3D surface construction algorithm. *Computer Graphics* 1987; 21(4):163-9

# Efficient Implementation of BDD Packages on the GPU Platform

Miloš M. Radmanović<sup>1</sup> and Dušan B. Gajić<sup>2</sup>

**Abstract** – Efficient construction and manipulation of Binary Decision Diagrams (BDDs) is an important component of Computer Aided Design (CAD) tasks. One solution to improve the performance of BDD packages is to perform some of the BDD operations in parallel using the Graphics Processing Unit (GPU). The recent evolution of GPU frameworks for programming general purpose computations, such as the OpenCL and Nvidia CUDA, has offered GPUs as a powerful and attractive choice for developing high-performance numerical applications. In this work, we propose an efficient implementation of a BDD package which distributes computations over central processing units (CPUs) and GPUs. The proposed implementation exploits various sources of parallelism that can be found in BDD packages. The experimental results confirm that the application of the proposed solution leads to significant computational speedups in BDD packages.

**Keywords** – Binary decision diagrams, BDD package, parallel implementation, graphics processing unit, GPU computing.

## I. INTRODUCTION

Binary Decision Diagrams (BDDs) are the dominant data structure for representing Boolean functions in CAD applications. The application of BDDs is further extended with their use in various areas of computer science and engineering. In practice, the success of BDD representations depends on the ability to efficiently manipulate large BDDs. Therefore, considerable research has been conducted in order to develop more efficient implementations of BDD algorithms [1-5].

BDD algorithms are usually built on top of BDD packages. Many BDD package implementations have been developed in a variety of programming languages and most of them are freely available as public domain on the Internet. The choice of a BDD package for a certain application is typically guided by the following package characteristics: functionality, software interface, robustness, reliability, portability, support, and performance. In most cases, the performance of a BDD package is of major concern. Parameters which influence the performance of a BDD package include the choice of the programming language and the software and hardware platforms, BDD node structure, type of garbage collection, unique and operation hash table strategies [6-8].

Parallel computing can be used to efficiently solve large-scale problems, either by distributing computational loads among processors or by utilizing the large memory in parallel networked workstations. Parallel processing of BDDs can be used both to reduce the BDD algorithm running time and to extend the memory limitations which exist in the traditional single-processor sequential computing.

In order to increase the performance of BDD packages, the concept of parallelism has been introduced to the BDD representations and algorithms in several papers. In [9], a parallel algorithm for the construction of BDDs is described. The algorithm is motivated by the fact that the construction of a BDD, for certain large or particularly complex Boolean functions, can be a very time-consuming task. In order to overcome limitations of computational resources, research in [10] presents an approach which distributes the BDD data structure across multiple networked workstations. Further, several techniques are introduced which allow parallelization of depth-first search algorithms on a BDD. Reference [11] presents a parallel algorithm for BDD construction targeted at shared memory multiprocessors and distributed shared memory systems. The results obtained using a shared memory multiprocessors system show speedups of over 2×, with four processors, and up to 4×, with eight processors. Alongside the research on parallel BDD construction, various BDD-algorithm parallel implementations were developed for networks of workstations [12-14]. In [15], some key algorithms for performing BDD operations are first described and, afterwards, an approach to their parallelization is described, with a goal to achieve efficient execution of BDD packages on multicore CPUs. The technique of *general purpose computing on the GPU (GPGPU)* enables parallel processing of non-graphics algorithms using graphics hardware. Only recently, the possibility of using GPUs to solve complex problems in logic design has been explored by researchers, for example in [16-20].

Motivated by the existing research on efficient execution of parallel BDD operations on multicore CPUs and possibility of using GPUs, in this paper we propose an efficient implementation of a BDD package using the GPU platform. The proposed implementation exploits the various sources of parallelism that exist in BDD packages. We address several topics considering parallel computations in BDD packages and present their mapping to the GPU architecture. The experimental results confirm that the application of the proposed implementation of a parallelized BDD package leads to significant computational speedups over traditional C/C++ implementations processed on CPUs.

<sup>1</sup>Miloš M. Radmanović is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: milos.radmanovic@gmail.com

<sup>2</sup>Dušan B. Gajić is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: dusan.b.gajic@gmail.com

The paper is organized as follows: Section 2 shortly introduces the *BDD representation of Boolean functions*. Section 3 describes the structure of a *BDD package* and the basic *BDD algorithms*. Section 4 presents the GPU as a computing platform. Section 5 discusses the operations in the BDD package for which we introduce the GPU processing. Section 6 shows experimental results obtained with the proposed implementation. Finally, Section 7 offers some concluding remarks and directions for future work.

## II. BINARY DECISION DIAGRAMS

BDDs consist of *non-terminal (decision) nodes*, *0-edges* and *1-edges* attached to all non-terminal nodes, a *'0' terminal node*, and a *'1' terminal node*, as shown in Fig.1. The non-terminal node with no upper nodes is called a *root node*. As it can be seen from Fig.1, a variable is related to every non-terminal node, such that every path from the root node to one of the terminal nodes respects the same variable ordering.

A Boolean function can be converted into an equivalent function by performing Shannon expansion based on the fixed variable ordering. This new function can be represented by a binary tree. The corresponding BDD is constructed from this binary tree by applying the two reduction rules (*redundant node elimination* and *equivalent sub-graphs sharing*). The Boolean operations such as the logical AND, logical OR, etc., can be achieved by using BDD manipulations, which have an average time complexity propositional to the size of BDDs. It is well known that the size of the BDD for a given *Boolean function* depends on the variable order for the function. The strength of BDDs is that they can represent Boolean function data with high level of redundancy in a compressed form.

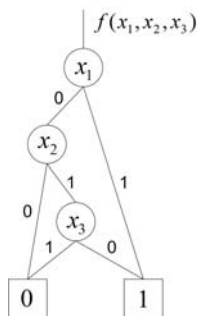


Fig. 1. BDD representation of the function defined by the truth table  $F = [00101111]^T$ .

## III. BDD PACKAGES

BDD packages are deployed in many software tools, particularly in the area of logic design, and they typically deal with the following common implementation features [1]. A BDD package has three main components [21]:

- *The BDD algorithm* component,
- *Dynamic variable reordering* component,
- *Garbage collection* component.

*The BDD algorithm* component builds the result BDDs for various Boolean operations. The implementation of these

algorithms is typically based on the *BDD node data structure*, *unique* and *operation tables*, and *depth-first BDD traversal*.

The decisions made in defining the *BDD node data structure* have impact on memory space requirements for storing node objects. There are many choices for defining a BDD node object, but every node usually contains: an *id*, then *cofactor*, *else cofactor*, *next pointer*, and *reference counter* [22]. The BDD construction is based on applying the traversal in a depth-first manner.

The maintenance of a BDD representation is improved by storing BDD nodes in a dictionary, called the *unique table*. The unique table maps a unique triple of (*v*, *g*, *h*) for a BDD node, where *v* is the variable identifier, *g* is the node connected to the "1" edge, and *h* is the node connected to the "0" edge. The unique table is a hash table with the hash collisions resolved by chaining. A hash function is applied to the triple to obtain the index in the unique table of the start of a collision chain of nodes. Comparing the unique triple against the nodes in the collision chain addresses the look up.

The efficient implementation of almost all recursive BDD manipulation algorithms is made possible by the *operation table*. This table is also implemented as a hash table with the collisions resolved by chaining. The collision lists can be kept sorted to reduce the number of memory accesses required on average for the lookup. Table sizes which are prime numbers require an expensive modulo operation. Table sizes that are a power of 2 are often better handled by memory management.

As the variable ordering can have significant impact on the size of a BDD, *dynamic variable reordering* component is a fundamental part of all modern BDD packages. Dynamic variable reordering algorithms are generally based on the *shifting algorithm* [23]. The BDD variable order changes by exchanging nodes in one level with nodes in the neighbouring levels. Dynamic variable ordering should best be invoked as an asynchronous process that can be activated at any time during the BDD manipulation. Dynamic variable ordering is a complex problem since finding an optimal ordering is NP-hard. Further, small changes in the BDD ordering may have significant impact on both the space and time requirements.

BDD computations are memory intensive, especially when large BDDs are involved. They not only require a lot of memory, but also frequent accesses to many small data structures. Furthermore, many intermediate BDD results are created to arrive at a resulting BDD. These computations may have poor memory handling, as there is not a solution to ensure that the accessed BDD nodes are close in memory. It is important to have a *garbage collector* component [24] to automatically remove BDD nodes that are no longer useful. In modern BDD packages, garbage collector component is based on *reference counting* and the *recycling of nodes* for later reuse. Garbage collection is activated when the percentage of the unusable BDD nodes reaches a threshold. Unusable BDD nodes are nodes with zero reference counts. Some of unusable BDD nodes may become usable again (recycled) if they are obtained as results of new subproblems. Thus, in the case when BDD nodes change state between „usable“ and „unusable“ frequently, garbage collection can reduce the benefit of the operation tables and decrease the overall performance of a BDD package.

#### IV. THE GPU ARCHITECTURE AND GPGPU

Processor frequency progress, which followed the Moore’s law for more than four decades, reached a limit in 2003, mostly due to the inability to further solve the problems of heat dissipation and energy consumption. Since then, there are two approaches in the development of computer architectures. The *multicore approach*, typical for CPUs, seeks to maintain the execution speed of sequential programs while moving into multiple cores. In contrast, the *manycore approach*, found in GPUs, focuses more on the execution throughput of parallel applications. This resulted in a rapid evolution of GPU architectures. The GPU evolution started from fixed-function hardware specialized for rendering computer graphics, which first appeared in 1999, and developed into a massively parallel, scalable, and fully programmable platform characterized by exquisite memory bandwidth and computational power. Due to this, many of the general-purpose applications which were processed on CPUs are now re-implemented in order to efficiently harness the GPU resources. For more details on recent changes that made GPGPU possible, see [25, 26, 27, 28].

The GPU parallel processing model is based on a large number of processor cores which can directly address into a global GPU memory. The GPU architecture follows the *single program, multiple data (SPMD)* paradigm [26, 27], features a *multi-level memory hierarchy* and has simple branching circuits. In SPMD computing, a large number of *threads* execute in parallel the same function, called a *kernel*, over different data.

Application Programming Interfaces (APIs) most often used for the development of GPGPU programs are *Nvidia’s CUDA* and *Open Computing Language - OpenCL*. CUDA is a vendor-specific development framework and only supports execution on Nvidia’s GPU hardware. Therefore, we give advantage to OpenCL which is hardware agnostic. Further, the OpenCL C programming language, included in the framework, allows development of programs that are both accelerated and portable across a wide set of devices (CPUs, GPUs, Field Programmable Gate Arrays (FPGAs), Digital Signal Processors (DSPs), Cell processors, embedded processors) [28].

#### V. GPU ACCELERATION IN THE BDD PACKAGE

Motivated by the existing work on the parallelization of components in BDD packages, described in Section 1, we explored various sources of parallelism that exists within the algorithms included in BDD packages in order to develop an efficient model of parallel BDD operations on GPUs. The components of the BDD package that take advantage of the *GPU processing* in our present solution are the *BDD algorithm* and the *garbage collection* components.

The effectiveness of caching within *unique* and *operation tables* of the BDD algorithm component strongly influences the number of subproblems generated in the BDD algorithm task execution. Thus, the hash tables (unique and operation tables) in a BDD package need to support concurrent execution of the hash operation *lookup\_insert(key)*. This

operation is a crucial component of the *Apply* procedure which is central to the BDD construction and manipulation [1]. The *lookup\_insert* operation returns the key, if it already exists in the hash table, or, otherwise, inserts the key. Reference [15] shows how this operation can be safely executed in parallel on multicore processors. The *lookup\_insert* operation within the BDD algorithm component in our BDD package is, therefore, implemented as an OpenCL kernel which performs the same function over different keys. Since GPUs use hardware multithreading [25, 26, 27], this automatically allows simultaneous execution of as many *lookup\_insert* operations as there are active GPU threads.

The effectiveness of garbage collection component can have significant impact on both space and time requirements of a BDD package. When garbage collector removes unusable BDD nodes, the *unique* and the *operation table* entries that reference these nodes must also be removed to eliminate unusable references. If garbage collection is not invoked frequently enough, the memory usage can be greatly increased. An OpenCL kernel for garbage collection is developed so that each GPU thread removes an entry from the hash tables. Since thousands of GPU threads can be active at the same time, this leads to a massively-parallel GPU garbage collection. The transfer of the garbage collection task to the GPU, also allows the CPU to be free to perform other tasks for which it may be more suitable.

#### VI. EXPERIMENTAL RESULTS

In this section, we compare the performance of our GPU accelerated BDD package implementation, which incorporates the before-mentioned OpenCL kernels, and a single-threaded C/C++ implementation of the BDD package on the CPU. For the comparison, we use a set of well-known standard benchmarks. Table I presents a view on the performance of the BDD package computations performed in the basic BDD construction algorithm on the CPU and the GPU.

TABLE I  
COMPARISON OF THE BDD CONSTRUCTION TIMES FOR THE BDD PACKAGE ON THE CPU AND THE GPU

Benchmark	in / out / cubes	computation time [s]	
		CPU	GPU
<i>alu4</i>	14 / 8 / 1028	0.15	0.08
<i>apex1</i>	45 / 45 / 206	5.18	0.81
<i>apex2</i>	39 / 3 / 1035	3.31	0.62
<i>apex5</i>	117 / 88 / 1227	0.30	0.17
<i>cordic</i>	23 / 2 / 1206	0.06	0.04
<i>cps</i>	24 / 109 / 654	0.15	0.09
<i>misex2</i>	25 / 18 / 29	0.05	0.03
<i>misex3</i>	14 / 14 / 1848	0.03	0.02
<i>table3</i>	14 / 14 / 135	0.02	0.02
<i>table5</i>	17 / 15 / 158	0.01	0.01

The test platform features an *Intel i7-920* quad-core processor, operating at 2.66 GHz, and has 4 GBs of DDR3-2000 RAM. GPU that is used is an *Nvidia GeForce GTX 560Ti* with 1GB of GDDR5 RAM, composed of 384



streaming processors. The OpenCL kernels are developed using the *AMD Accelerated Parallel Programming SDK 2.6*.

The size of the *unique* and the *operation tables* is limited to 8191 entries. The garbage collection is activated if the hash tables exceed the 80%-full marker. All benchmarks are used in the *Espresso-mv* or *pla* format [29] and the computation times are reported in seconds.

As it can be seen from Table I, the addition of the GPU acceleration to the BDD package brings clear performance benefits over the CPU-only solution. The speedup, in terms of the BDD construction algorithm computation time, for most of the cases is substantial and varies from 6.4× to 1.5×. However, it should be noted that the speedup may not be achieved in some cases, e.g., in the case of the benchmark *table3*, because the construction of the BDD in this case is not enough computationally-intensive to benefit from the introduction of the GPU.

## VII. CONCLUSION

This paper proposes an implementation of a BDD package which uses the GPU hardware for the acceleration of certain data-parallel operations. The proposed implementation exploits several sources of parallelism that exist within BDD packages. In particular, we discuss the parallel OpenCL implementation of the *lookup\_insert* hash operation, which is of central importance to the *BDD algorithm component*, and a GPU-accelerated *garbage collection component*. The experimental results confirm that the application of the proposed implementation, which distributes the BDD package operations over the CPU and the GPU, leads to significant computational speedups. The results presented in the paper may also be helpful in the general study on improvement of BDD packages. Since these first research results look promising, further work on this topic will be devoted to the extension of the GPU acceleration method to the implementation of other operations that are common in the components of BDD packages.

## REFERENCES

- [1] K. Brace, R. Rudell, R. Bryant, "Efficient implementation of a BDD package", in *Proc. Design Automation Conf.*, 1990, 40-45.
- [2] J. Sangavi, R. Ranjan, R. Bryton, A. Sangiovanni-Vincentelli, "High performance BDD package based on exploiting memory hierarchy", in *Proc. of the Design Automation Conf.*, 1996.
- [3] D. Long, "The design of cache-friendly BDD library", in *Proc. 1998 IEEE/ACM Intl. Conf. on CAD*, 1998, 639 - 645.
- [4] G. Janssen, "Design of pointerless BDD package", *10<sup>th</sup> Int. Workshop on Logic & Synthesis*, Lake Tahoe, USA, 2001.
- [5] R. Ebendt, F. Gorschwin, R. Drechsler, *Advanced BDD Minimization*, Springer, New York, 2005.
- [6] G. Janssen, "A consumer report on BDD packages", in *Proc. 16<sup>th</sup> Symp. Integrated Circuits and Systems Design*, 2003, 217-223.
- [7] F. Somenzi, "Efficient manipulation of decision diagrams," *Int. J. Software Tools for Technology Transfer (STTT)*, vol. 3, no. 2, 2001, 171-181.
- [8] M. Sentovich, "A brief study of BDD package performance", in *Proc. of the Formal Methods on CAD*, 1996, 389-403.
- [9] S. Kimura, E.M. Clarke, "A parallel algorithm for constructing binary decision diagrams", in *IEEE Intl. Conf. on Computer Design: VLSI in Computers and Processors*, 1990, 220-223.
- [10] T. Stornetta, F. Brewer, "Implementation of an efficient parallel BDD package", in *Proc. of Design Automation Conf.*, Las Vegas, USA, 1996, 641-644.
- [11] B. Yang, D. R. O'Hallaron, "Parallel breadth-first BDD construction", *9<sup>th</sup> ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming*, 1997, 145-156.
- [12] D. Caban, D. Milford, "A parallel BDD engine for logic verification", in *Proc. 5<sup>th</sup> Annual IEEE Int. ASIC Conf. and Exhibit*, 1992, 499-502.
- [13] R.K. Ranjan, J.V. Sanghavi, R.K. Brayton, A. Sangiovanni-Vincentelli, "Binary decision diagrams on network of workstations," *IEEE Int. Conf. on Computer Design: VLSI in Computers and Processors*, 1996, 358-364.
- [14] K. Milvang-Jensen, J. Alan, "BDDNOW: A parallel BDD package", in *Proc. 2<sup>nd</sup> Int. Conf. on Formal Methods in Computer-Aided Design*, 1998, 501-507.
- [15] H. Yuxiong, "Multicore-enabling a binary decision diagram algorithm", Intel Software Network, 2009, <http://software.intel.com/en-us/articles/multicore-enabling-a-binary-decision-diagram-algorithm>.
- [16] K. Gulati, S. Khatri. "Towards acceleration of fault simulation using graphics processing units", in *Proc. 45<sup>th</sup> ACM/IEEE Design Automation Conference*, 2008, 822-827.
- [17] V. Bertacco, D. Chatterjee, "High performance gate-level simulation with GP-GPU computing", *Int. Symp. on VLSI Design, Automation and Test*, 2011, 1-3.
- [18] D. Chatterjee, V. Bertacco, "EQUIPE: parallel equivalence checking with GP-GPUs", *IEEE Int. Conference on Computer Design*, 2010, 486-493.
- [19] D. Gajić, R. Stanković, "GPU accelerated computation of fast spectral transforms", *Facta Universitatis - Series: Electronics and Energetics*, vol. 24, no. 3, University of Niš, Serbia, 2011, 483-499.
- [20] D. Gajić, R. Stanković, M. Radmanović, "Implementation of dyadic correlation and autocorrelation on graphics processors", *Int. J. Reasoning-based Intelligent Systems (IJRIS)*, vol. 4, nos. 1/2, Inderscience, Geneva, Switzerland, 2012, 82-90.
- [21] B. Yang, "Optimizing model checking based on BDD characterization", PhD Dissertation, Carnegie Mellon University Pittsburgh, USA, 1999.
- [22] B. Yang et al., "A study of BDD performance in model checking", in *Proc. Formal Methods in CAD*, 1998, 255-289.
- [23] R. Rudell, "Dynamic variable ordering for ordered binary decision diagrams", in *Proc. Int. Conf. on Computer-Aided Design*, 1993, 139-144.
- [24] N. Klarlund, T. Rauhe, "BDD algorithms and cache misses", BRICS Report Series RS-96-5, Department of Computer Science, University of Aarhus, 1996.
- [25] T. Aamodt, "Architecting graphics processors for non-graphics compute acceleration", in *Proc. IEEE PACRIM Conf.*, 2009, 963-968.
- [26] S. Ryoo, et al., "Optimization principles and application performance evaluation of a multithreaded GPU using CUDA", in *Proc. 13<sup>th</sup> ACM SIGPLAN Symp. on Principles and Practice of Parallel Programming*, 2008, 73-82.
- [27] J. Owens, M. Houston, D. Luebke, S. Green, J. Stone, and J. Phillips, "GPU computing", *Proc. of the IEEE*, vol. 96, no. 5, 2008, 279-299.
- [28] B. Gaster et al., *Heterogeneous Computing with OpenCL*, Elsevier, 2011.
- [29] R. Rudell, Espresso Misc. Reference Manual Pages, <http://embedded.eecs.berkeley.edu/pubs/downloads/espresso/index.html>, 1993.

# Architecture of Distributed Multiplatform GIS for Meteorological Data Analysis and Visualization

Marko Kovačević<sup>1</sup>, Aleksandar Milosavljević<sup>1</sup>, Vladan Mihajlović<sup>1</sup> and Dejan Rančić<sup>1</sup>

**Abstract** – This paper presents a distributed multiplatform architecture of geographic information system (GIS) capable of providing flexible and efficient geospatial and meteorological data analysis and visualization to a broad spectrum of user profiles. Traditionally, geographic information systems were realized as monolithic and platform-dependent applications and used by rather small group of GIS professionals. With the development of computer hardware, Internet and availability of geospatial data in recent years, GIS evolved and adapted to new environments that emerged. Greater availability has led to greater importance and use of GIS in many different domains. In this paper we will focus on architecture specification that allows distribution of geospatial data and functions to different hardware and software platforms. The proposed architecture consists of four subsystems: Desktop GIS, GIS Server, Web GIS and Mobile GIS. Each of the identified subsystems has its different role and intended user groups in the system as a whole. Great diversity of meteorological data applications (weather forecast, hail suppression, climate change monitoring, aviation, agriculture, military etc.) implies variety of different GIS user needs and profiles. This paper will focus on identifying subsystem roles and possible use case scenarios that can generally be applied for all meteorological data applications.

**Keywords** – GIS, Multiplatform Architecture, Meteorology.

## I. INTRODUCTION

A geographic information system (GIS) is a special type of computer-based information system, tailored to store, process, and manipulate geospatial data [1]. The ability of GIS to handle and process both location and attribute data, distinguishes GIS from other information systems. It also establishes GIS as a technology that is important for a wide variety of applications [2]. Traditionally, geographic information systems were built as monolithic and platform-dependent applications [3], but, with the development of computer hardware and availability of geospatial data in recent years, GIS evolved and adapted to new environments that emerged [4]. Greater availability has led to greater importance and use of GIS in many different areas. What was reserved for a small group of professionals in the past now became available worldwide, using Internet, on personal computers and mobile devices.

Modern distributed GIS architectures rely on a client/server model where clients provide user access to geospatial data, while one or more servers provide their sharing. The task of

sharing geospatial data and functions can be done by extending existing traditional desktop GISs with a set of “web interfaces”. This kind of extension forms a GIS Server component. Besides GIS Server, proposed architecture consists of Desktop GIS, Web GIS and Mobile GIS subsystems. Each of the identified subsystems has its different role and intended user groups in the system as a whole. Great diversity of meteorological data applications (weather forecast, hail suppression, climate change monitoring, aviation, agriculture, military etc.) implies variety of different GIS user needs and profiles. This paper will focus on identifying subsystem roles and possible use case scenarios that can generally be applied for all meteorological data applications.

In order to build a distributed GIS that is open for connecting to a variety of different geospatial data sources, commonly acceptable standards for implementation of these web interfaces are needed [5]. Currently, Open Geospatial Consortium, as an international industrial consortium with an aim of developing publicly available standards in the field of GIS, has several implementation specifications that standardize this field. Our functional needs for web interfaces were met by three of these specifications. The first is Web Map Service (WMS) Implementation Specification [6] that describes a web service interface for custom maps retrieval. The second is Web Feature Service (WFS) Implementation Specification [7] which describes a web service interface for manipulating, querying, and retrieval of geospatial entities using both spatial and non-spatial criteria. The third is Web Coverage Service (WCS) Implementation Specification [8] which describes a web interface for retrieval of geospatial data as “coverages” – digital geospatial information representing space-varying phenomena. This paper focuses on a global GIS architecture specification that is based on previously mentioned specifications and that allows distribution of geospatial data and functions to different hardware and software platforms.

The paper is organized as follows: in the second part, we define subsystem-level architecture with emphasis on communication between them. The third part is dedicated to Desktop GIS and GIS Server subsystems. In the fourth part client side subsystems are discussed in more detail. The fifth part presents a case study concerning subsystems’ use in meteorological data applications. Finally, in the conclusion, the achieved results are summarized.

## II. SUBSYSTEM-LEVEL ARCHITECTURE

The architecture of proposed GIS solution consists of four subsystems: Desktop GIS, GIS Server, Web GIS, and Mobile

<sup>1</sup>Marko Kovačević, Aleksandar Milosavljević, Vladan Mihajlović and Dejan Rančić are with the Faculty of Electronic Engineering at University of Niš, Aleksandra Medvedeva 14, Niš 18000, Republic of Serbia, E-mail: marko.kovacevic@elfak.ni.ac.rs, alexm@elfak.ni.ac.rs, wlada@elfak.ni.ac.rs, ranca@elfak.ni.ac.rs.

GIS. Like all modern distributed GIS architectures, proposed solution architecture relies on a client/server model.

Communication between server side subsystems is depicted by the component UML diagram shown in Fig. 1. The responsibility of Desktop GIS component, in the server side of the system, is creation and preparation of a content that can be distributed via GIS Server. GIS Server implements OGC WMS, WFS and WCS services in order to distribute geospatial data and functions to its clients. For preparation of geospatial content, Desktop GIS relies on use of a project file. The project file is an XML document that contains specification of all coordinate systems, geodata services, coverage data, layers, and styles that are used for geospatial and coverage data organizing and presentation. While Desktop GIS allows creation and manipulation of GIS projects, GIS Server only use previously prepared projects for geospatial content distribution.

Component UML diagram in Fig. 2 shows client side subsystems of proposed GIS solution. Mobile GIS and Web GIS rely on web interfaces provided by the GIS Server (WMS, WFS and WCS) in order to provide their own functionalities. Desktop GIS also has the ability of utilizing web interfaces provided by the GIS Server. GIS Server is intentionally left out in Fig. 2 because, from the client standpoint, it can be any GIS server (or servers) that implements OGC standardized WMS, WFS and WCS.

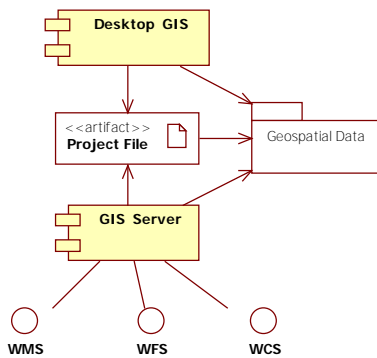


Fig. 1. Server side subsystems

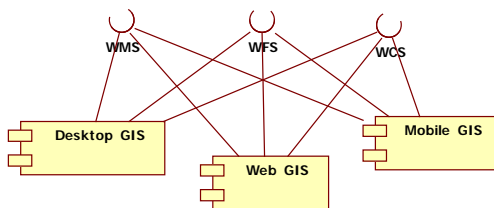


Fig. 2. Client side subsystems

### III. DESKTOP GIS AND GIS SERVER

The described relation between Desktop GIS and GIS Server implies tightly coupled implementation of these two components. Core of Desktop GIS application and GIS Server

system is GIS framework – a set of C++ class libraries developed in Laboratory for Computer Graphics and GIS (CG&GIS Lab), on Faculty of Electronic Engineering in Niš. GIS framework implements common data model and most of the functionalities. Desktop GIS application and GIS Server system service can be seen as a kind of an interface toward them. The main advantage of sharing the same core between these two subsystems is in maintainability. When some new feature (or change) is added to the framework (e.g. support for new data format) it automatically becomes present in both Desktop GIS application and GIS Server. A simplified overview of the implementation architecture of GIS framework along with Desktop GIS and GIS Server is shown in Fig. 3. There are two packages (Visual C++ projects) responsible for production of executable components of Desktop GIS application and GIS Server system service, while the other packages (i.e. projects) represents previous introduced GIS framework.

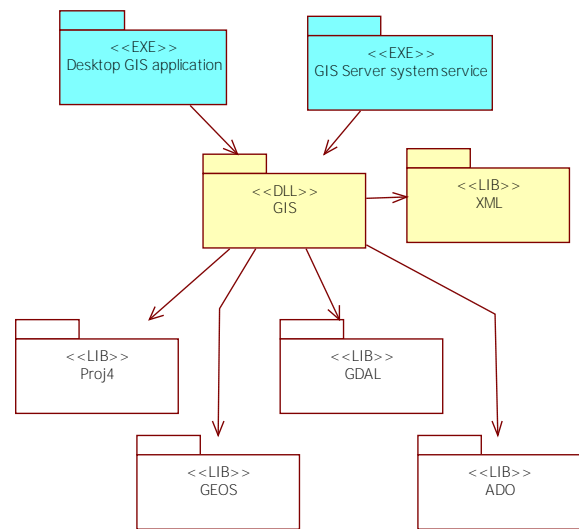


Fig. 3. Simplified implementation architecture of Desktop GIS and GIS Server

The GIS framework is implemented with support of four open source libraries: Proj4, GDAL, GEOS, and ADO. Proj4 [9] is a library that implements mechanism for definition of various coordinate reference systems (i.e. projections) and reprojection of coordinates between them. GDAL [10] is a library for raster geospatial data formats management, while GEOS [11] library enables support for spatial querying and geoprocessing. Finally, ADO [12] library contains a set of classes that simplifies database accessing and management.

The majority of the framework implementation (e.g. data model) is contained within GIS dynamic library classes, while XML is only helper library for accessing, managing, and creating XML documents. Since Desktop GIS and GIS Server share the same implementation framework, both of them are tied to the same platform. In our development we relied on Microsoft Visual C++ 2008 and Microsoft Foundation Classes (MFC). As a result, current Desktop GIS and GIS Server platform is limited to Microsoft Windows and Microsoft Windows Server series of operating systems.

IV. CLIENT SIDE SUBSYSTEMS

As we already stated, client side of the proposed GIS solution includes Web GIS, Mobile GIS and Desktop GIS. Web GIS and Mobile GIS completely depend on services and geospatial data supplied by GIS Server. Desktop GIS has the capability of utilizing web interfaces provided by the GIS Server, but also has other responsibilities and functionalities, as described in section 3.

Web GIS's primarily use is geospatial data visualization and querying. GIS Server's WMS service is used for retrieval of custom maps in form of raster images. WFS service enables query execution and results retrieval in form of GML [13] encoded geographic features. WCS service enables coverage data retrieval for defined spatial and temporal constraints. The implementation of Web GIS heavily relies on OpenLayers [14], open source JavaScript library developed by the Open Source Geospatial Foundation. The implementation of Web GIS is based on several web technologies such as: PHP, JavaScript, AJAX, and .NET.

Mobile GIS has similar background architecture, i.e. it also use WMS, WFS and WCS services of GIS Server, but the implementation and intended use is quite different. As it is intended to run on hand-held devices, as "on field" tool for data acquisition and situation awareness, the platform of choice was Microsoft Windows Mobile. Mobile GIS subsystem is implemented using C# programming language and .NET Compact Framework. Unlike Web GIS that drives huge support form OpenLayers library, Mobile GIS's communication with WMS, WFS and WCS services is implemented from scratch. Since Mobile GIS, besides map visualization and coverage data querying, also deals with geospatial data acquisition, GIS Server's WFS service has to be extended with transactional part of the specification. Transactional WFS, in addition to querying, allows manipulation of geographic features using insert, update and delete operations.

As previously stated, Desktop GIS is able to use web interfaces provided by GIS Server: WMS, Transactional WFS and WCS. Classes that implement communication between Desktop GIS and above listed web interfaces are contained within GIS library (Fig. 3). Since Desktop GIS and GIS Server share the same GIS framework, GIS Server can also act as a client of some other GIS server, thus enabling the creation of cascade connection between servers. Being the most powerful platform of the three, Desktop GIS provides additional functionalities which are not implemented in Web GIS and Mobile GIS: various tools for WCS coverage data analysis and flexible style management for visual representation of WCS coverage data, WMS raster images and WFS geographic features.

V. CASE STUDY

Using GIS in meteorological domain involves variety of different user needs and profiles. Meteorological data applications are vastly diverse: weather forecast, hail suppression, climate change monitoring, aviation, agriculture, military etc. From the GIS user perspective, each

meteorological data application requires different user profiles that differ in:

- Education level,
- Responsibility for decision making,
- Working environments,
- Experience with GIS tools, etc.

For all numbered reasons, using single platform solution (one GIS application) can be inappropriate for wider user audience. In order to analyze the use of proposed GIS subsystems, we should identify several typical GIS user profiles of meteorological data application in general:

- Meteorological data analyzers
- Decision makers
- Simple viewers
- Users involved in "on field" operational tasks, and
- Applications administrators.

Meteorological data analyzers are the most complex group of users that includes experts in particular meteorological (or some related) domain which are also specialized for GIS tools use. For that reason, powerful Desktop GIS applications are what this spectrum of users needs. Furthermore, these users often need specialized tools for geospatial and meteorological data analysis. Our solution includes development and delivery of such specialized tools within Desktop GIS. These tools are developed as plugins – DLLs that conform to a predefined interface used for interaction with the application and underlying geospatial data. When present, a plugin is automatically loaded into Desktop GIS application and can be accessed from appropriate menu item. A sample screen shoot of Desktop GIS application is shown in Fig. 4.

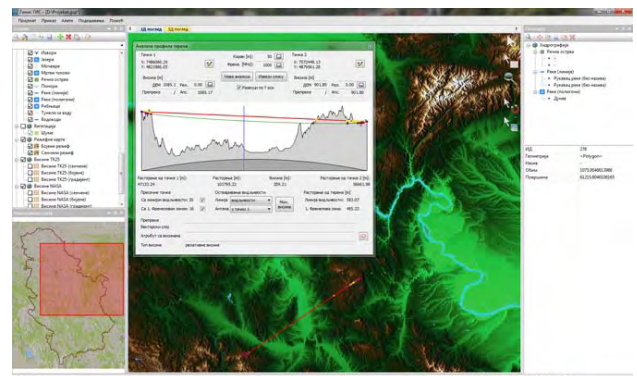


Fig. 4. Screen shot of Desktop GIS application

Decision makers are users that have authority to make decisions (e.g. hail suppression activity initiation, aviation flight plan regulation, taking appropriate actions in agriculture, military etc.). Although usually experts in particular domain, decision makers are characterized as casual GIS users. Their primarily tasks are not GIS related, i.e. they (can) only use GIS as another information source for decision making. More often, decision making must be fast, so this category of users must be quickly presented with the exact as possible information they need. By our opinion, simplified applications with easy-to-use user interfaces, along with adequate geospatial and meteorological data content are what decision makers need. Client side subsystems, primarily Web

GIS and Mobile GIS applications, are what we see as the most appropriate solution for this category. A sample screen shots of Web GIS and Mobile GIS are shown in Fig. 5.

The third group, simple viewers, includes broader audience of generally non-specialized people interested in some aspect of particular meteorological domain (e.g. weather forecast). Web GIS is the most suitable tool for this user group.

Many meteorological applications include “on field” operational tasks. In terms of GIS functions we recognized two major categories: simple meteorological data collecting and support for enhanced situation awareness. On field operations, regarding the environment, require specialized computing hardware and the running software. Mobile GIS, as the proposed tools of choice, basically satisfies listed requirements for this user group.

Finally, the last identified category of GIS users is involved in applications administration. The personnel appointed to these tasks must be well educated and especially trained to support maintenance of the proposed GIS system. Although they directly do not use any of the presented subsystems they must have good knowledge of all of them.

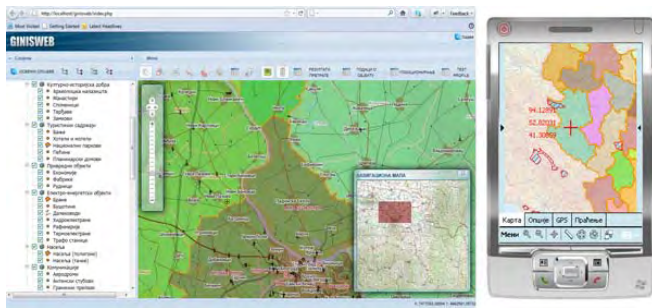


Fig. 5. Screen shot of Web GIS (left) and Mobile GIS (right)

## VI. CONCLUSION

Great diversity of meteorological data applications implies large spectrum of potential GIS users. In this paper we focused on two main issues. Firstly, we presented distributed multiplatform architecture of GIS applicable in meteorological applications that consists of four subsystems: Desktop GIS, GIS Server, Web GIS and Mobile GIS. Second issue was identifying subsystem roles and possible use case scenarios that can generally be applied for all meteorological data applications.

Proposed GIS solution architecture relies on a client/server model. Server side subsystems are GIS Server and Desktop GIS. In this context, the main role of Desktop GIS application is content preparation in terms of data acquisition and processing. In order to distribute prepared data, we rely on well-known concept of project file. Once the project file and corresponding geospatial data are prepared, they can be shared using WMS, WFS and WCS services of GIS Server subsystem. Client side subsystems are Web GIS, Mobile GIS and Desktop GIS. Web GIS and Mobile GIS completely depend on services and geospatial data supplied by GIS Server, while utilization of web interfaces provided by GIS

Server is only part of the responsibilities and functionalities of Desktop GIS. Primary use of Web GIS is geospatial data visualization and querying, and it is intended for broader spectrum of less specialized users. Mobile GIS, as it runs on hand-held devices, is basically intended for “on field” data acquisition and situation awareness. As the most powerful client side subsystem that provides various specialized analysis tools along with basic GIS functionalities, Desktop GIS is primarily indented for specialized users.

## ACKNOWLEDGEMENT

Research presented in this paper is funded by Ministry of Education and Science, Republic of Serbia as part of the project “Environmental Protection and Climate Change Monitoring and Adaptation”, Nr. III-43007.

## REFERENCES

- [1] Worboys, M., Duckham, M., GIS: A Computing Perspective. Second Edition. CRC Press, Boca Raton, FL, 2004.
- [2] Chang, K., Introduction to Geographic Information Systems, Third Edition. McGraw-Hill, New York, NY, 2005.
- [3] Wong, S. H., Swartz, S. L., Sarkar, D., “A Middleware Architecture for Open and Interoperable GISs”, IEEE MultiMedia, 9(2)(2002) pp. 62-76.
- [4] Shekhar, S., Vatsavai, R. R., Sahay, N., Lime, S., “WMS and GML based Interoperable Web Mapping System”, Proceedings of the 9th ACM international symposium on Advances in geographic information systems, Atlanta, Georgia, USA, 2001, pp. 106-111.
- [5] Milosavljević, A., Đorđević-Kajan, S., Stoimenov, L., “An Application Framework for Rapid Development of Web based GIS: GinisWeb”, in: J.T. Sample, K. Shaw, S. Tu, M. Abdelguerfim eds. Geospatial Services and Applications for the Internet, Springer, New York, 2008, pp. 49-72.
- [6] La Beaujardière, J. ed., Web Map Service Implementation Specification (Ver. 1.1.1), document 01-068r3, Open Geospatial Consortium Inc., 2002, <http://www.opengeospatial.org/standards/wms>.
- [7] Vretanos, P. A., ed., Web Feature Service Implementation Specification (Ver. 1.0.0), document 02-058, Open Geospatial Consortium Inc., 2002, <http://www.opengeospatial.org/standards/wfs>.
- [8] Whiteside, A., Evans, J. D., ed., Web Coverage Service (WCS), Implementation Specification (Ver. 1.1.0), document 06-083r8, Open Geospatial Consortium Inc., 2006, <http://www.opengeospatial.org/standards/wcs>.
- [9] PROJ.4 – Cartographic Projections Library, <http://proj.osgeo.org>.
- [10] GDAL – Geospatial Data Abstraction Library, <http://www.gdal.org>.
- [11] GEOS – Geometry Engine, Open Source, <http://trac.osgeo.org/geos>.
- [12] Antollini, C., “A set of ADO classes - version 2.20”, The Code Project, 6 Sep 2005, <http://www.codeproject.com/KB/database/aaadoclass1.aspx>.
- [13] Cox, S., Cuthbert, A., Lake, R., Martell, R. eds., OpenGIS Geography Markup Language Implementation Specification (Ver. 2.1.2), document 02-069, Open Geospatial Consortium, 2002, <http://www.opengeospatial.org/standards/gml>.
- [14] OpenLayers – Free Maps for the Web, <http://www.openlayers.org>.

# Tracking Analogue to Digital Converter Modelling using VHDL-AMS

Marieta Kovacheva<sup>1</sup> and Ivailo Pandiev<sup>2</sup>

**Abstract** – This paper describes a VHDL-AMS implementation of a behavioural model for tracking analogue to digital converter (ADC). For creating the model, simplification and build-up techniques known from modelling analogue integrated circuits have been adapted. This generalized model is independent from actual technical realizations and is based upon compromises regarding the representation of exact circuit structure in the model. The proposed tracking ADC model accurately predicts the circuit behaviour for dc and transient responses. Its behaviour is created consequently using tracking ADC basic structure. The modelling of the ADC behaviour is implemented and confirms to the format of the simulation program System Vision 5.5 (from Mentor Graphics). The simulation results show accurate agreement with the theoretical predictions.

**Keywords** – Mixed-signal circuits, Analogue to digital converters, Counting method, Tracking ADCs, VHDL-AMS, Mixed-signal simulation.

## I. INTRODUCTION

An analogue to digital converter (ADC) is a device that converts a continuous quantity to a discrete time digital representation. Typically, an ADC is an electronic device that converts an input analogue voltage or current to a digital number proportional to the magnitude of the voltage or the current. The digital output may use different coding schemes. Basically there are three different conversion methods: (1) parallel method (convert a word at a time), (2) weighting method (convert a digit at a time) and (3) counting method (convert a level at a time). In Table 1, is shown a comparison between three basic methods of conversion [1-4]. The benefits of the tracking converters are that (1) their output is continuously available and (2) their accuracy is very high. The main drawback of tracking ADC is their slow step response, however, this is compensated with their possibility of large resolution (12-,14-,16- bits).

After analyze of the existing model libraries in OrCAD PSpice A/D [5], SystemVision (from Mentor Graphics) [6] and Matlab Simulink [7] some conclusions are made. In System Vision libraries an ADC behavioural model can be found, but it works by using successive-approximation method and its resolution is fixed to 10 bits, others types of converters are not included. In OrCAD PSpice A/D libraries exists simulation model of ADC – called ADC8, ADC10 or

ADC12. Its resolution can be set as 8 -, 10 - or 12-bit. In the PSpice based model the voltage at *in* mode and *ref* mode with respect to *ground* mode is sampled starting at the rising edge of the *convert* signal and ending when the status goes high. These models with a suitable choice of parameters and elements can be used of a mixed-signal circuit simulations, but not confirm to the architectures of a broad class of the monolithic ADCs. Transformation from OrCAD PSpice A/D to System Vision libraries can be done, but it's quite complicated, requires a lot of resources and additional processing. In Matlab Simulink exists one ADC model, called idealized ADC quantizer with the following changeable parameters: (1) number of converted bits, (2) minimum and maximum output voltage and (3) output data type. This model is generalized structure and cannot be used for several specific applications. Also an investigation in the published articles was made and the conclusions are following – found one article describing a VHDL implementation of a behavioural model for pipeline ADCs by using VHDL description to facilitate the synthesis of the digital part [8], another one for simulation tools in designing sigma delta ADC and its efficiency by using both Simulink and VHDL-AMS [9] and the last one presents behavioural VHDL-AMS model for monolithic Half-Flash (Two Step) 10-bit ADC [10]. The goal of this paper is to develop a behavioural VHDL-AMS model that accurately simulates the basic characteristics of tracking ADCs.

TABLE 1 COMPARISON OF CONVERSION METHODS

ADC types	Number of clocks	Number of reference voltages	Characteristics
Parallel (Flash) method	1	$2^n - 1$	very fast; needs many parts ( $2^n - 1$ comparators); lower resolution, expensive; large power consumption
Successive-approximation method	$n$	$n$	capable of high speed; medium accuracy; good tradeoff between speed and cost; speed limited ~ 5MSPs; require S/H circuit
Counting method	max $2^n$	1	input signal is averaged; greater noise immunity; high accuracy; slow; require S/H circuit

<sup>1</sup>Marieta Kovacheva is with the Faculty of Electronics at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: m\_kovacheva@tu-sofia.bg.

<sup>2</sup>Ivailo Pandiev is with the Faculty of Electronics at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria. E-mail: ipandiev@tu-sofia.bg.

## II. PRINCIPLE OF TRACKING CONVERTERS OPERATION

The counting method requires the least circuit complexity, but the conversion time is considerably longer than with the

other methods – generally between 1ms and 1s. This is a great plus for slowly changing signals. The counting method can be implemented in various ways; one of them is the tracking conversion. The basic structure of tracking ADC is given on Fig. 1.

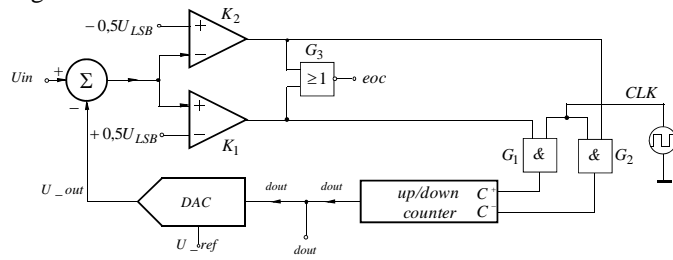


Fig. 1. Tracking ADC basic structure.

The operation of tracking ADC is the following. We subtract compensating voltage  $v_{out}$  from input voltage  $v_{in}$ . The voltage  $v_{out}$  is the DAC output voltage. The result of subtracting the two voltages is passed to two analogue comparators  $K1$  and  $K2$ . It is compared with reference voltages  $v_{ctrl\_pos}$  and  $v_{ctrl\_neg}$  with values  $\pm 0,5U_{LSB}$  ( $U_{LSB}$  is the voltage unit for the least significant bit; that is the voltage for resulting number  $dout = 1$ ). If the voltage difference is bigger than  $+0,5U_{LSB}$ , the comparator  $K1$  jumps to '1' ( $K2$  output is '0') and enables the logic gate  $G1$ . This gate passes the impulses from clock generator to the summing input of the up/down counter. The counter increases its value and DAC output voltage tracks the input voltage. If the voltage difference is smaller than  $-0,5U_{LSB}$ , the comparator  $K2$  jumps to '1' ( $K1$  output is '0') and enables the logic gate  $G2$ . This gate passes the impulses from clock generator to the subtracting input of the up/down counter. The counter decreases its value. The up and down changes in the digital code of the DAC allow its output voltage always to tracks the variation in the input voltage. If the difference between  $v_{in}$  and  $v_{out}$  is in range of  $\pm 0,5U_{LSB}$ , no clock impulses are passed to the counter and this is the end of conversion mode (DAC has reached the input value, and it's digital code corresponds to it). The output comparator values are both '0' and the value of logic element  $G3$  is '1'. The input voltage, the compensating DAC output voltage and end of conversion signal  $eoc$  versus time are given on Fig. 2.

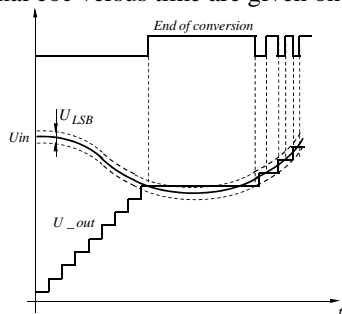


Fig. 2. Tracking process at AC input voltage for the tracking converter on Fig. 1.

### III. MODELLING OF TRACKING ADC WITH VHDL-AMS

The created behavioural model of tracking ADC is developed by using a style combining structural and mathematical description. The structural description is the netlist of the model and the behavioural description consists of simultaneous statements to describe the continuous behaviour. The behaviour of the proposed ADC is strictly described using basic structure given on Fig. 1.

#### A. A behavioural language: VHDL-AMS

VHDL-AMS is a comparatively new standard 1076.1 of VHDL that supports hierarchical description and simulation of analogue and mixed-signal applications with conservative and non-conservative equations. On the mixed-signal side a variety of abstraction levels is supported. The VHDL-AMS modelling is not restricted to mixed-signal applications but also supports thermal and mechatronic systems.

#### B. A tracking ADC behavioural VHDL-AMS model

The proposed behavioural model of a tracking ADC is developed following the design method based on a Top-Down analysis approach and applying simplification and build-up technique, known from modelling analogue integrated circuits. The process of model building and testing can be broken down into three main steps: 1) *structure the model*; 2) *build the model*; 3) *validate the model*.

The structure of the tracking ADC model is built using the results obtained by analyses of the simulation models for ADCs in [5, 6] and by using the structures presented on Fig. 1. The model includes two generic parameters with numerical values:  $v_{ref} = 5,12V$  – reference voltage and  $Nbits = 12$  – number of bits. These two parameters can be changed according different applications. For the model consecutively are defined the work of the up/down counter, DAC, subtracting the DAC output voltage from input circuit voltage, analogue comparators with work ranges, logic gates and assuming the results on output signals. Depending on the result, output value of '1' or '0' is assumed.

Fig. 3 shows the behavioural VHDL-AMS model of tracking ADC. Library clause and the use clause make all declarations in the packages `std_logic_1164`, `logic_arith`, `logic_unsigned`, `math_real` and `electrical_systems` visible in the model. This is necessary because the model uses logic types, operation with logic types and nature electrical from packages. The proposed model is composed by an entity and an architecture, where bold text indicates reserved words and upper-case text indicates predefined concepts. The entity declares the generic model parameters and specifies one interface terminal of nature electrical, two input signal terminals – type `std_logic`, one output signal terminal – type `std_logic` and one output signal terminal – type `std_logic_vector`, with a length of 12 bits. The proposed model includes the following terminals: `ain` – input voltage, `clk`, `stop` – clock and stop signal, `dout` – output vector, `eoc` – end of conversion.

```

library IEEE; library ieee_proposed;
use IEEE.std_logic_1164.all; use IEEE.std_logic_arith.all; use IEEE.math_real.all;
use ieee.std_logic_unsigned.all; use ieee_proposed.electrical_systems.all;

entity tracking_adc_12_bits is

    generic ( Nbits : integer := 12; -- number of bits of ADC's output
              v_ref : voltage := 5.12); -- ADC's reference voltage

    port ( terminal ain : electrical; --ADC's analog input terminal
          signal clk, stop : in std_logic; -- Strobe clock and stop signal
          signal dout : out std_logic_vector (0 to 11);-- Digital output
          signal eoc : out std_logic ); -- equilibrium signal

end entity tracking_adc_12_bits;

architecture default of tracking_adc_12_bits is
    signal out_plus, out_minus: std_logic; -- output of comparators
    signal clk_up, clk_down: std_logic; -- up and down input of the counter
    signal s_out: real := 0.0; -- output of dac
    signal synch : std_logic; -- internal synchronization of clock signal
    terminal sum_out: electrical; -- sum of input signal and dac signal
    terminal dac_out : electrical; -- dac output
    quantity v_out across i_out through dac_out to electrical_ref;
    quantity vin across ain to electrical_ref;
    quantity vsum across isum through sum_out to electrical_ref;
    quantity vctrl_pos, vctrl_neg : voltage;

begin

    vctrl_pos == 0.5*(v_ref/2.0**Nbits); -- upper reference voltage
    vctrl_neg == -0.5*(v_ref/2.0**Nbits); -- lower reference voltage

    process(Vin'above(vctrl_pos), Vin'above(vctrl_neg), clk_up, clk_down, synch)

        variable counter : std_logic_vector (0 to 11):="000000000000"; -- init of the counter
        variable sum : natural;

    begin

    if rising_edge(clk_up) and clk_down = '0' then -- define work mode of counter
        counter := counter + '1';
    elsif rising_edge(clk_down) and clk_up = '0' then
        counter := counter - '1';
    end if;
    dout <= counter;
    sum := 0;

    for i in counter'range loop -- modelling DAC
        sum := sum * 2 + boolean'pos(counter(i) = '1' or counter(i) = 'H' );
    end loop;
    s_out <= v_ref * real(sum) / real(2**Nbits);
    -- comparing the sum_out with ranges of comparators
    if vsum'above(vctrl_pos) then
        out_plus <= '1';
        out_minus <= '0';
        eoc <= '0';
    elsif not vsum'above(vctrl_neg) then
        out_plus <= '0';
        out_minus <= '1';
        eoc <= '0';
    elsif not vsum'above(vctrl_pos) and vsum'above(vctrl_neg) then
        out_plus <= '0';
        out_minus <= '0';
        eoc <= '1';
    end if;
    if (stop = '0') -- stop signal, bans the work of counter
        then counter := (others=>'0');
    end if;

end process;

    v_out == s_out'ramp(1.0E-6); -- assuming variables to output signals
    vsum == vin - v_out; -- voltage difference
    clk_up <= out_plus and clk;
    clk_down <= out_minus and clk;
    synch <= clk;
end architecture default;

```

Fig. 3. A 12-bit tracking ADC behavioural VHDL-AMS model.

Furthermore, the model has two inner terminals: sum\_out and dac\_out. They are used to specify the voltages vsum and v\_out respectively. Also six inner signals are defined: out\_plus, out\_minus – outputs of K1 and K2; clk\_up, clk\_down – summing and subtracting inputs of counter; s\_out – DAC output and synch – internal synchronization of the clock signal.

The architecture is subdivided into four main parts according to the functions of the different elements in basic structure: (1) up/down counter, (2) DAC, (3) subtractor and (4) window comparator with additional logic element, generating eoc signal.

The output value of the converter is formed by the formula:

$$Z = \frac{V_I}{V_{LSB}} = 2^n \frac{V_I}{V_{ref}} = Z_{max} \frac{V_I}{V_{ref}} \quad (1)$$

where  $Z$  is the output digital value,  $V_I$  is input analogue voltage,  $V_{LSB} = V_{ref} / 2^n$  is the least significant bit voltage and is equal to division of reference voltage by  $2^n$  bits,  $n$  is the number of bits, and  $Z_{max}$  - is the maximum possible digital value.

The schematic representation of the symbol confirm to the description of Fig. 3 is given on Fig. 4. The generated model has a total number of five ports: ain – input terminal, clk – clock signal, stop - stop signal, eoc – end of conversion signal, dout[0:11] – output logic vector with length of 12 signals. Number of bits in the model can be changed with define a new value for the parameter Nbits, the signal dout and the variable counter. After the change a generation of new symbol is needed.

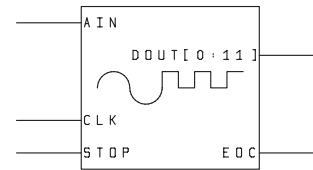


Fig. 4. A schematic symbol of 12 – bit tracking ADC.

#### IV. MODEL PERFORMANCE

Firstly is simulated only the modelled ADC, for the aim of that a piecewise linear voltage source is connected on the input voltage pin. This source can provide a periodic voltage profile, with time period  $50ms$ . The input voltage set values are  $0,5$ ,  $1$ ,  $1,5$  and  $1V$ . The clock signal is with frequency of  $100kHz$ . The stop signal is with period of  $400ms$ . The simulated schematic is given on Fig. 5. The output results are presented on Fig. 6. We can see clearly how the digital output is tracking the variable value of the input voltage. According to the formula the corresponding output values for input voltages  $-0,5$ ,  $1$ ,  $1,5V$  are  $400$ ,  $799,9$  and  $1200$ . Their binary codes are  $110010000$ ,  $1100100000$  and  $10010110000$ . When obtaining the simulation results, we can see the correct work of the converter.

Secondly, to proof the validity of the proposed model, a schematic of resolver to digital converter is created. It's structure is presented on Fig. 7.



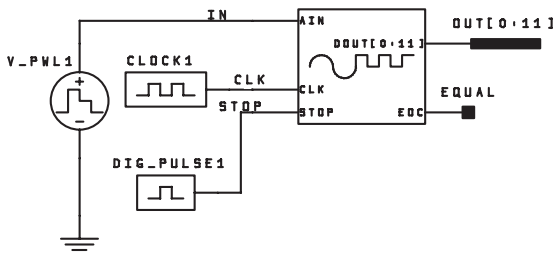


Fig. 5. Tracking ADC simulated schematic.

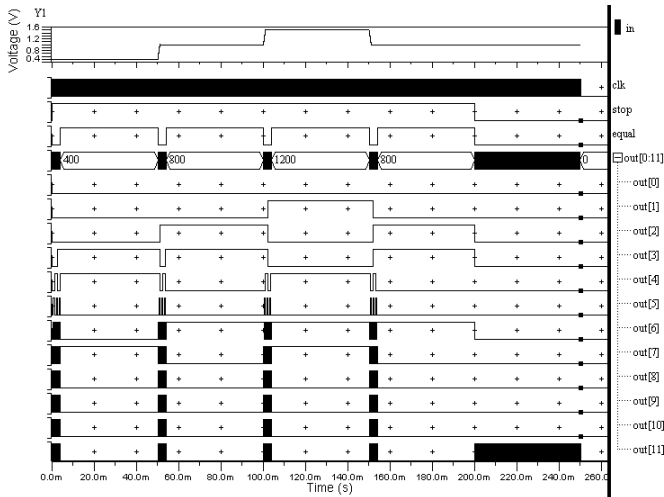


Fig. 6. Simulation results for four values of the input voltage.

The schematic on Fig. 7. includes rotor, primary and secondary windings on the stator and two ADCs. The stator windings are displaced mechanically by  $90^\circ$ . The primary winding is excited with an ac reference. The amplitude of subsequent coupling onto the stator secondary windings is a function of the position of the rotor (shaft) relative to the stator. The resolver, therefore, produces two output voltages ( $S3 - S1, S2 - S4$ ) modulated by the sine and cosine of shaft angle. These output voltages are passed to two ADCs. The ADCs track the change in the rotor position and their digital outputs respond to the respective analogue value. The SystemVision libraries do not include a model of step motor. To realize this structure is used an ac source modelled  $v_p \sin(\omega t)$  and two dc sources modelled  $\sin(\theta)$  and  $\cos(\theta)$ . The dc sources present the sine and cosine of the corresponding angle. Each of the dc sources is multiplied with ac source in order to form output voltages relevant to the stator ones. The simulation is made for six values of the angle –  $15^\circ, 30^\circ, 45^\circ, 60^\circ, 75^\circ$  and  $90^\circ$ . The ac source is with amplitude  $V_p$  equal to  $1V$ , dc offset voltage  $1,5V$  and frequency  $2kHz$ . The simulation results are given in Table 2. As you can see the results confirm to the calculated parameters.

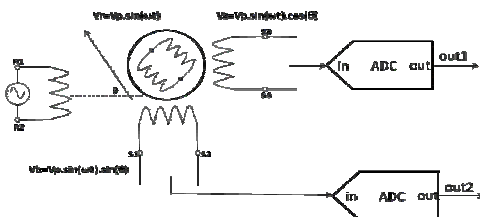


Fig. 7. Schematic of resolver to digital converter.

TABLE 2. SIMULATION RESULTS FOR RESOLVER CONVERTER

shaft angle - $\theta$	$\sin(\theta)$	$\cos(\theta)$	out1	out2
$15^\circ$	0,2588	0,9659	$314_{10}$	$1108_{10}$
$30^\circ$	0,5	0,866	$605_{10}$	$1030_{10}$
$45^\circ$	0,707	0,707	$849_{10}$	$849_{10}$
$60^\circ$	0,866	0,5	$1030_{10}$	$605_{10}$
$75^\circ$	0,9659	0,2588	$1108_{10}$	$314_{10}$
$90^\circ$	1	0	$1135_{10}$	$0_{10}$

V. CONCLUSION

In this paper a generalized behavioural VHDL-AMS model of tracking 12-bit ADC has been presented. The model is implemented and the structure of its description confirm to the format of the simulation program SystemVision 5.5. The proposed model accurately simulates the actual performance of typical tracking ADC. The efficiency of the model was proved by comparison of the simulation results with theoretically calculations for piecewise linear input voltages. Furthermore, the workability of the model was shown by simulation testing of a schematic of a resolver to digital converter. The simulations were performed for several values of the shaft angles of the rotor. The obtained results confirm to the theoretically calculated parameters.

ACKNOWLEDGEMENT

This paper is part of a project, which is sponsored by the research program of the TU-Sofia, Bulgaria.

REFERENCES

- [1] G. Mihov. *Digital Electronics*, Sofia, Technical University – Sofia, 2005 (in Bulgarian).
- [2] M. Seifart. *Analoge Schaltungen*. 6 Auflage. Verlag Technik Berlin, 2003 (in German).
- [3] T. Floyd. *Digital fundamentals*, 10th edition, Prentice-Hall, 2009.
- [4] V. Tietze, Ch. Schenk. *Electronic circuits*. 2nd Edition. Berlin, Heidelberg, New York. Springer-Verlag, 2008.
- [5] *PSpice A/D library list*, Release 9.2, Cadence Design Systems, 2000.
- [6] System Vision mixed-signal model library (ver. 5.5), Mentor Graphics, 2007.
- [7] Matlab Simulink release 2011b libraries, The Mathworks, 2011.
- [8] E. Peralias, A.J. Acosta, A. Rueda, J. Huerdas. VHDL-based behavioural description of pipeline ADC, *Measurement*, vol. 31, pp. 47-60, 2002.
- [9] M. Rizzi, B. Castagnolo. Efficiency in simulations tools in designing sigma-delta ADC. *AEU - International Journal of Electronics and Communications*, vol. 60, pp. 290-298, 2006.
- [10] M. Kovacheva, I. Pandiev. Behavioural VHDL-AMS model for half-flash analog-to digital converter. *Annual journal of electronics*, vol. 5, No 1, pp. 39-42, 2011.

# Development of Parameterized Verilog-AMS Model of Photovoltaic Cells

Elissaveta Gadjeva<sup>1</sup> and Georgi Valkov<sup>2</sup>

**Abstract** – A parameterized Verilog-AMS model of photovoltaic cell, module and panel is developed. The model is available in two variants: one that includes a detailed Spice diode description and the other which is simplified and includes only the diode equations that affect the behavior of the PV elements. Definitions for obtaining the model characteristics from the simulation results are provided. The model is realized in the Dolphin Integration SMASH environment.

**Keywords** – Photovoltaic cells, Parameterized models, Verilog-AMS language.

## I. INTRODUCTION

The design and optimization of photovoltaic systems require development of computer models of PV cells and panels. These models are of significant importance for the simulation of maximum power point tracking (MPPT) control systems and allow the investigation and optimization of the dynamics of PV power systems.

Photovoltaic models of PV cells and modules are developed in [1-3]. Methods for parameter extraction and simulation of photovoltaic modules by taking the manufacturer specified data are developed in [4-6]. A sun-tracking system based on the Verilog HDL and using an FPGA chip as controlling platform is proposed in [7]. Combining all analog and digital blocks of a project, using a behavior HDL like Verilog-AMS and VHDL-AMS [9] allow the entire project to be simulated and verified as a whole. A SoC design of solar irrigation control system based on FPGA is offered in [8].

Parameterized Verilog-AMS macromodel is created in the present paper that can represent a single PV cell, as well as an entire panel consisting of PV cells connected in series and parallel. An additional model using a simplified version of the diode is also described. Definitions for obtaining the model characteristics from the simulation results are provided, according to the pattern file syntax used by SMASH [10].

## II. MODELS OF PV CELL AND PANEL

### A. Model of a single PV cell

The schematic representation of the macromodel for a single PV cell is provided in Fig. 1.  $I_1$  is a DC current source with value  $I_1 = I_{CC} \cdot K_{SH}$ , where  $I_{CC}$  is a function of the solar

irradiation and the shading factor  $K_{SH}$  models the behavior of the cell under reduced light conditions. Its range is [0, 1], where 1 is for unshaded cell.  $R_1$  and  $R_2$  model the shunt and series resistance losses. Their values are  $R_1 = R_{SH}$  and  $R_2 = R_S$ . The diode models the non-linear behavior, its parameters are  $I_S$  and  $N$ , where  $I_S$  is the saturation current and  $N$  is the emission coefficient.

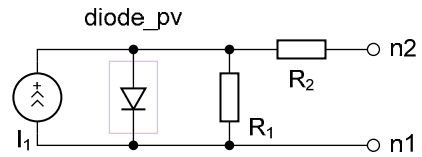


Fig. 1. Parameterized model of a single solar cell.

A generic symbol for the PV cell model is shown in Fig. 2.

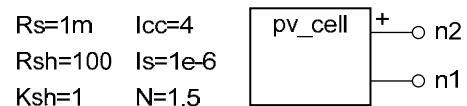


Fig. 2. Generic symbol of PV cell.

### B. Model of PV panel

The model of PV panel uses the same schematic representation as the single PV cell shown in Fig. 1, but with scaled parameters to represent a whole panel. A panel consists of  $N_p$  modules connected in parallel, where each module has  $N_s$  cells connected in series, as shown in Fig. 3.

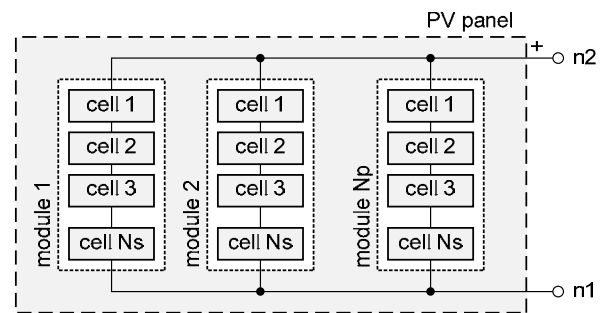


Fig. 3. Structure of PV panel.

In the case when  $N_s=1$  and  $N_p=1$ , the model represents a single PV cell. The values for model parameters are  $I_1 = I_{CC} \cdot K_{SH} \cdot N_p$ ,  $R_1 = R_{SH} \cdot N_p$ ,  $R_2 = R_S \cdot N_s$ ,  $I_S = I_S \cdot N_p$ ,  $N = N \cdot N_s$ .

A generic symbol for the PV panel model is shown in Fig. 4.

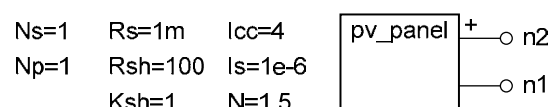


Fig. 4. Generic symbol of PV panel.

<sup>1</sup>Elissaveta Gadjeva is with the Department of Electronics at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: egadjeva@tu-sofia.bg.

<sup>2</sup>Georgi Valkov is with the Department of Electronics at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: gvalkov@abv.bg.

III. VERILOG-AMS IMPLEMENTATION OF PV PANEL

A. Verilog-AMS code of PV panel with full diode description

The Verilog-AMS implementation of the model has the following form:

```

module pv_cell_01_vams(n1, n2);
    inout n1, n2;
    electrical n1, n2, n3;
    parameter real Icc= 4.0;
    parameter real Ksh= 1.0;
    parameter real Rs = 1.0m;
    parameter real Rsh=100.0;
    parameter real Ns = 1.0;
    parameter real Np = 1.0;
    parameter real Is = 1.0u;
    parameter real N = 1.5;
    analog begin
        I(n3, n1) <+ V(n3, n1)*Np/Rsh - Icc*Ksh*Np;
        I(n3, n2) <+ V(n3, n2)*Np/(Ns*Rs);
    end
    diode_pv_sp #(.Is(Is*Np), .N(N*Ns))
        D_pv (.n2(n3), .n1(n1));
endmodule

```

The block has two electrical terminals,  $n_2$  is the positive one and  $n_1$  is the negative. The model parameters are defined. The elements  $I_1$ ,  $R_1$  and  $R_2$  are modeled by their component equations. A diode block is added between nodes  $n_3$  and  $n_1$ , to model the non-linear behavior.

SMASH is a mixed language simulator and allows subcircuits defined using the *Spice* language syntax to be added to Verilog-AMS nets and vice versa. The *diode\_pv\_sp* block wraps a *Spice* diode model inside a subcircuit, using the following *Spice* description:

```

.SUBCKT diode_pv_sp n1 n2 PARAMS: Is=1 N=1
.MODEL diode_pv D (Is=Is N=N)
D1 n2 n1 diode_pv
.ENDS

```

B. Model of PV panel with simplified diode description

For the purpose of simulating PV cells behavior, the diode is only used in forward direction. Additional effects like junction capacitance can also be omitted. Hence it is possible to simplify the model of PV panel with *Spice* diode, by replacing the diode with its equivalent equation. The model is described by the following Verilog-AMS code:

```

module pv_cell_02_vams(n1, n2);
    inout n1, n2;
    electrical n1, n2, n3;
    parameter real Icc = 4.0 from (0: inf);
    parameter real Ksh = 1.0 from [0: 1];
    parameter real Rs = 1.0m from (0: inf);
    parameter real Rsh = 100.0 from (0: inf);
    parameter real Ns = 1.0 from [1: inf];
    parameter real Np = 1.0 from [1: inf];
    parameter real Is = 1.0u from (0: inf);
    parameter real N = 1.5 from (0: inf);
    parameter real Tn = 300.15 from (0: inf);
    parameter real XTI = 3.0 from (0: inf);
    parameter real Eg = 1.11 from (0: inf);
    analog begin
        I(n3, n1) <+ V(n3, n1)*Np/Rsh - Icc*Ksh*Np
        + Np*Is*(exp(V(n3, n1)/(Ns*N*$vt))-1)
        * exp(($temperature/Tn-1)*Eg/(Ns*N*$vt))
        * pow($temperature/Tn, XTI/(Ns*N));
    end
endmodule

```

```

I(n3, n2) <+ V(n3, n2)*Np/(Ns*Rs);
end
endmodule

```

IV. SIMULATED CHARACTERISTICS IN DOLPHIN INTEGRATION SMASH

The characteristics for the Verilog-AMS model of PV panel with full diode description are obtained for a single cell using DC sweep as the primary simulation analysis. To determine the dependence on  $R_s$ ,  $R_{SH}$ ,  $K_{SH}$  and temperature, the following parametric sweeps are run together with the primary analysis:  $R_s$ : value list: 0.1m, 1m, 10m, 100m;  $R_{SH}$ : value list: 1, 10, 100;  $K_{SH}$ : linear from 0.25 to 1, with a step size of 0.25 and  $TEMPER$ : linear from 0 to 75, with a step size of 25. The remaining model parameters are given in Fig. 4.

The current and power characteristics are obtained using the waveforms  $I(V_1)$  and  $IN(V_1,POWER)$ . The following characteristics are also obtained from the simulation results: current and voltage at the maximum power point  $I_{PP}$  and  $V_{PP}$ , fill factor  $FF$ , maximum power  $P_{MAX}$ , short circuit current  $I_{SC}$ , and open circuit voltage  $V_{OC}$ . The parameter  $FF$  is defined as follows:

$$FF = \frac{P_{MAX}}{P_T}, \quad (1)$$

where:  $P_{MAX} = \max(P), \quad (2)$

$$P_T = I_{SC} V_{OC}. \quad (3)$$

The definitions for obtaining  $P_{MAX}$ ,  $I_{SC}$ ,  $V_{OC}$ ,  $P_T$ ,  $FF$ ,  $V_{PP}$ , and  $I_{PP}$  from the simulation results are created using the *MEASURE* command according to the pattern file syntax used by SMASH:

```

.MEASURE ANALYSIS=DC NAME=Pmax
+ WAVEFORM=IN(V1.POWER) EXTRACT=MAX
.MEASURE ANALYSIS=DC NAME=Isc
+ WAVEFORM=I(V1) EXTRACT=MAX
.MEASURE ANALYSIS=DC NAME=Voc
+ WAVEFORM=I(V1) EXTRACT=X ATY=0
+ CROSS=DOWN OCCUR=1
.MEASURE ANALYSIS=DC NAME=PT VALUE={Isc*Voc}
.MEASURE ANALYSIS=DC NAME=FF VALUE={Pmax/PT}
.MEASURE ANALYSIS=DC NAME=Vpp
+ WAVEFORM=IN(V1.POWER) EXTRACT=ATMAX ATY=0
.MEASURE ANALYSIS=DC NAME=Ipp
+ WAVEFORM=I(V1) EXTRACT=Y ATX=Vpp

```

The current and power characteristics as a function of  $R_s$  are shown in Fig. 5.

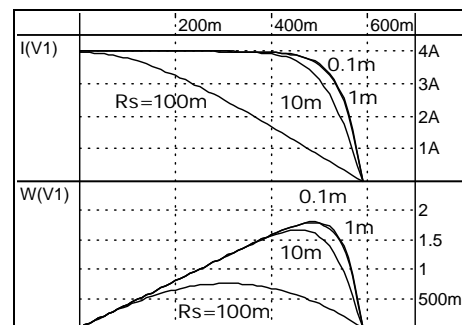


Fig. 5. Current and power characteristics as a function of  $R_s$ :  $I(V_1, R_s)$  and  $P(V_1, R_s)$ .

High values for the series resistance tend to cause a significant drop in the output power of the PV elements as well as reduction of  $I_{PP}$ ,  $V_{PP}$  and  $FF$  (Fig. 6).

The current and power characteristics as a function of  $R_{SH}$  are shown in Fig. 7. High values for the shunt resistance have no impact on the performance, while low values act like an additional consumer, reducing the power at the output of the PV elements and also reducing  $I_{PP}$ ,  $V_{PP}$  and  $FF$  (Fig. 8). The shading factor  $K_{SH}$  indicates what part of the PV cell or panel is not shaded. The output current is proportional to the solar irradiation and hence proportional to  $K_{SH}$ . Lower values for  $K_{SH}$  result in a lower output power (Fig. 9), and also cause a reduction in  $I_{PP}$ ,  $V_{PP}$  and  $FF$ , as shown in Fig. 10.

The temperature dependences of the output current and power characteristics as a function of  $R_S$  are shown in Fig. 11. At lower temperatures  $V_{OC}$  and  $V_{PP}$  are higher. As a result the I-V characteristic is shifted to the right; the output power at the maximum power point and  $FF$  are also higher. There is also a small increase in  $I_{PP}$  (Fig. 12).

The model of simplified PV panel is compared to the PV panel model with full diode description. The temperature dependence of the I-V characteristic is shown in Fig. 13.

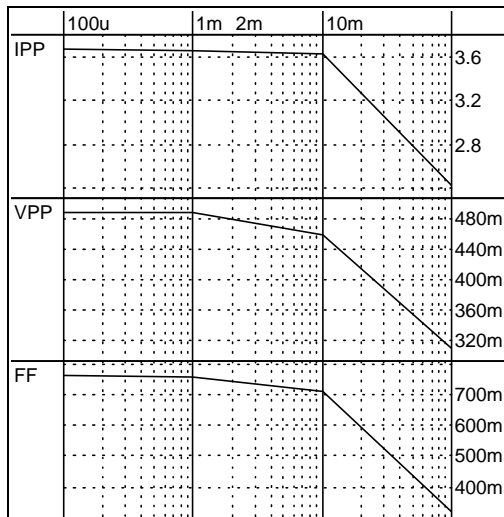


Fig. 6.  $I_{PP}$ ,  $V_{PP}$  and  $FF$  characteristics as a function of  $R_S$ :  $I_{PP}(R_S)$ ;  $V_{PP}(R_S)$  and  $FF(R_S)$ .

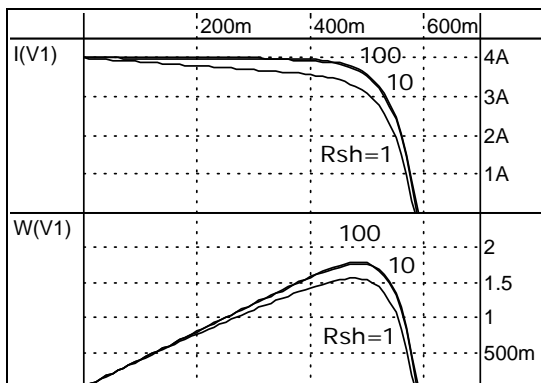


Fig. 7. Current and power characteristics as a function of  $R_{SH}$ :  $I(V_1, R_{SH})$  and  $P(V_1, R_{SH})$ .

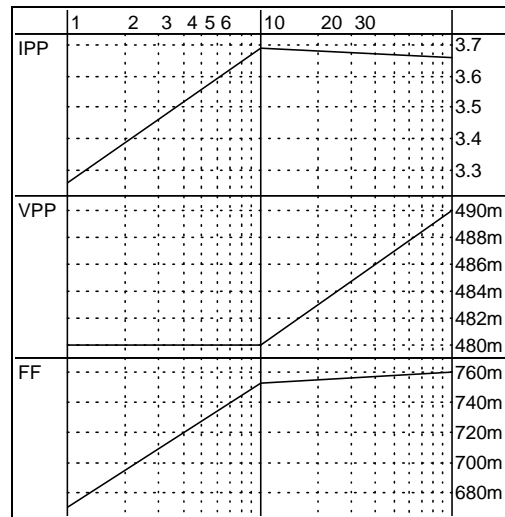


Fig. 8.  $I_{PP}$ ,  $V_{PP}$  and  $FF$  characteristics as a function of  $R_{SH}$ :  $I_{PP}(R_{SH})$ ;  $V_{PP}(R_{SH})$  and  $FF(R_{SH})$ .

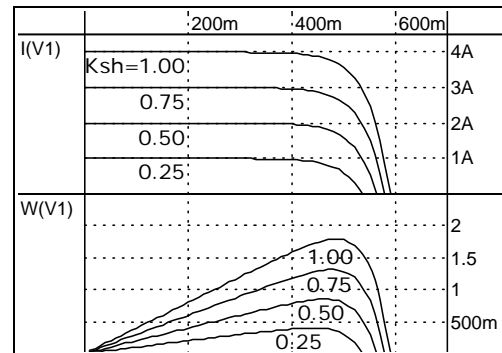


Fig. 9. Current and power characteristics as a function of  $K_{SH}$ :  $I(V_1, K_{SH})$  and  $P(V_1, K_{SH})$ .

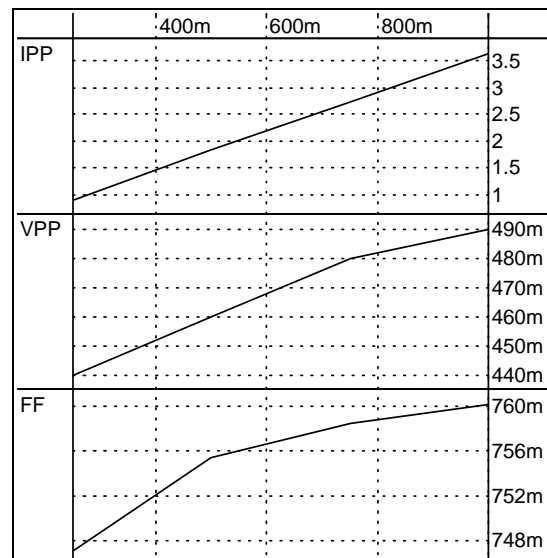


Fig. 10.  $I_{PP}$ ,  $V_{PP}$  and  $FF$  characteristics as a function of  $K_{SH}$ :  $I_{PP}(K_{SH})$ ;  $V_{PP}(K_{SH})$  and  $FF(K_{SH})$ .

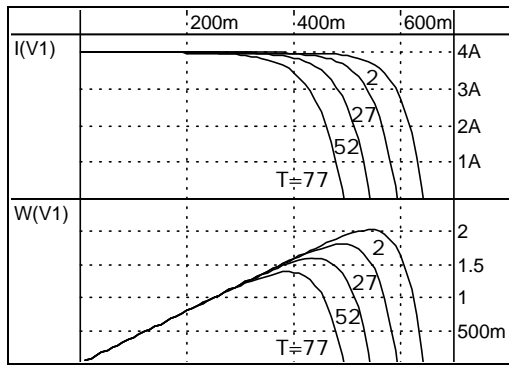


Fig. 11. Current and power characteristics as a function of the temperature:  $I(V_1, T)$  and  $P(V_1, T)$ .

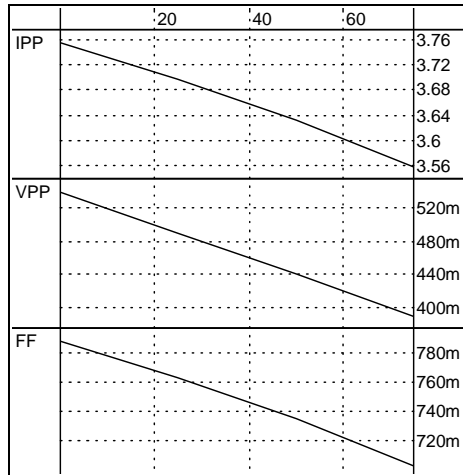


Fig. 12.  $I_{PP}$ ,  $V_{PP}$  and  $FF$  characteristics as a function of the temperature:  $I_{PP}(T)$ ;  $V_{PP}(T)$  and  $FF(T)$ .

The current characteristics match for all of the simulated temperatures. The relative difference is smaller than  $2 \times 10^{-6} \%$ , hence it is advisable to use the simplified model in order to speed the simulation of large circuits, as the accuracy is not sacrificed.

### V. CONCLUSIONS

A behavioral parameterized Verilog-AMS macromodel has been developed for PV cell and PV panel consisting of cells connected in series and in parallel. The model of PV panel can also represent a single cell. The detailed model relies on a *Spice* diode model to describe the non-linear behavior of the PV cell. An alternative simplified model is also provided. It includes only the diode equations that are important for the modeling of the PV behavior. It is a pure Verilog-AMS implementation, and does not depend on mixed language simulators. Both models are simulated in the Dolphin Integration SMASH environment. The basic PV characteristics are obtained from the simulation results.

### ACKNOWLEDGEMENT

The investigations are supported by the project №122PD0026-03.

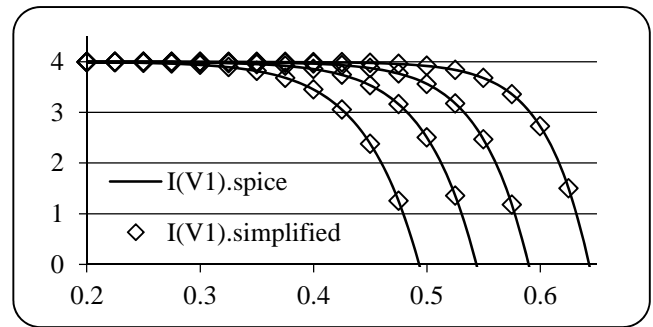


Fig. 13. Comparison between the model with *Spice* diode and the simplified model. Current and power characteristics as a function of the temperature:  $I(V_1, T)$  and  $P(V_1, T)$ .

### REFERENCES

- [1] H-L., Tsai, Ci-Siang Tu, and Yi-Jie Su. "Development of generalized photovoltaic model using MATLAB/SIMULINK", Proceedings of the World Congress on Engineering and Computer Science 2008, WCECS 2008, October 22 - 24, San Francisco, USA, 2008.
- [2] I.H., Altas, A.M., Sharaf. "A photovoltaic array simulation model for Matlab-Simulink GUI environment", International Conference on Clean Electrical Power., ICCEP '07. 21-23 May 2007, pp. 341 – 345, Capri, 2007.
- [3] Hernanz, R., C. Martín, J.J. Belver, L. Lesaka, Z. Guerrero, E. Puelles Pérez."Modelling of photovoltaic module", International Conference on Renewable Energies and Power Quality, (ICREPQ'10) Granada (Spain), 23th to 25th March, 2010.
- [4] D. Sera, R. Teodorescu, P. Rodriguez. "PV panel model based on datasheet values", IEEE International Symposium on Industrial Electronics, ISIE 2007, June 4-7 2007, Vigo, Spain, ISBN: 978-1-4244-0754-5, pp. 2392-2396, 2007.
- [5] G. Farivar and B. Asaei. "Photovoltaic module single diode model parameters extraction based on manufacturer datasheet parameters", IEEE International Conference on Power and Energy, PECon 2010 Nov. 29-Dec. 1 2010, Selangor, Malaysia, ISBN: 978-1-4244-8947-3, pp. 929-934, 2010.
- [6] N.N.B. Ulapane, C.H. Dhanapala, S.M. Wickramasinghe, S.G. Abeyratne, N. Rathnayake and P.J. Binduhewa. "Extraction of parameters for simulating photovoltaic panels", 2011 IEEE 6th International Conference on Industrial and Information Systems (ICIIS), 16-19 Aug. 2011, Peradeniya, Sri Lanka, ISBN: 978-1-4577-0032-3, pp. 539-544, 2011.
- [7] Cheng Sen, Zhao Ping, He Hongkun, Ji Qianqian and Wei Xu. "An improved design of photo-voltaic solar tracking system based on FPGA", AICT'10, International Conference on Artificial Intelligence and Computational Intelligence, 23-24 October 2010, Sanya, China, ISBN: 978-1-4244-8432-4, pp. 267-271. 2010.
- [8] Ze Cheng, Dan Zhong, Baolin Li and Yanli Liu. "The SoC design and realization of small scale solar irrigation control system based on FPGA", 2011 International Conference on Electric Information and Control Engineering (ICEICE), Apr. 15-17, 2011, Wuhan, China, ISBN: 978-1-4244-8036-4, pp. 887-890, 2011.
- [9] F. Pêcheux, C. Lallement, A. Vachoux, "VHDL-AMS and Verilog-AMS as Alternative Hardware Description Languages for Efficient Modeling of Multi-Discipline Systems", Paper #1446, IEEE, ISSN 0278-0070, 2005.
- [10] Dolphin Integration SMASH overview, [http://www.dolphin.fr/medal/smash/smash\\_overview.php](http://www.dolphin.fr/medal/smash/smash_overview.php)

# Optical Control through Stencils Cutting in Surface Mount Technology

Valentin Videkov<sup>1</sup>, Aleksei Stratev<sup>2</sup>, Georgi Furkov<sup>3</sup>

**Abstract** – This paper analyzes cutting techniques to conduct destructive testing in surface mounting. It showcases two methods to control processes during manufacture of technology equipment and process control. Experimental results from a process for manufacture of laser cut stencils are demonstrated.

**Keywords** – SMT, optical control, laser cut stencils,

## I. INTRODUCTION

Surface-mount technology is a major technology used in the manufacture of electronic equipment, from basic LED light sources [1] to complex computer systems and mobile devices [2]. Control over the technological process is crucial due to the widespread use of this technology. Destructive and non-destructive control methods exist. Optical methods for control are applied more frequently, including automated process control methods [3]. In spite of this automation, there are control points where the automated optical control systems are inappropriate to use. Those are processes that are applied relatively rarely and are designed to control individual elements. For example, control over application of solder paste in stencils. Solder paste-laying masks are manufactured based on different technologies [4], and the optical control is applied to assess the manufacturing technology. Both the dimensions and the shape of the resulting holes can be controlled through this method [5 by us]. The standard optical control in stencils sometimes is not able to evaluate the impact of all technological parameters and cannot address all questions. This paper will demonstrate a method for optical control on stencils that allows for a broader assessment of the process for laser cutting of stencils.

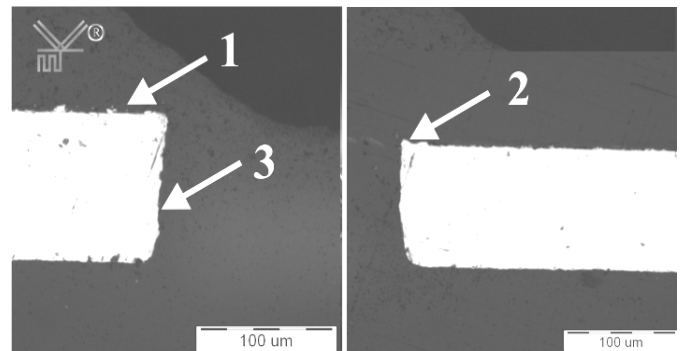
## II. LASER CUT STENCILS

Laser cut stencils feature many advantages compared to chemically pickled ones. Foremost, one can produce large,

in terms of area, masks where the same precision in making the holes on the entire area is applied. At the same time, there are no limitations related to chemical methods' pickling to produce small holes.

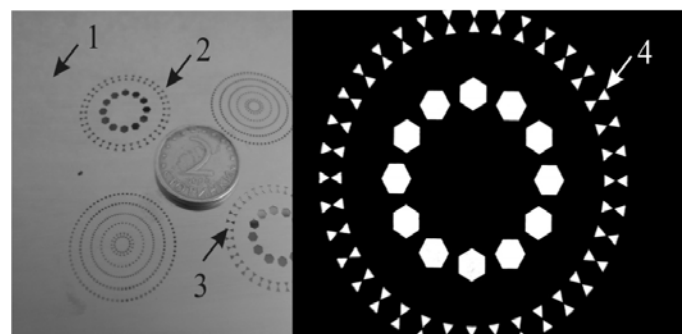
An advantage of the method is also that a lower quantity of consumables is used during the manufacture of the stencils. Laser cutting treats the metal plate directly, as opposed to chemical etching, where the metal plate undergoes a procedure of preliminary treatment through photolithography, in addition to expenditures incurred to pickle and rinse the metal plate.

An advantage of the laser cut stencils is the ability to get relatively vertical walls of the holes – Fig.1.



**Fig. 1** Cutting of a laser cut stencil. 1- plane of the mask, 2 – edge of the hole, 3 – wall of the hole

This verticality of the cutting, combined with the extreme movement precision of the laser beam and the small dimensions of the spot (cut), allow for complex holes featuring different orientations – fig. 2.



**Fig. 2** Shapes of holes in laser cut stencils

<sup>1</sup>Valentin Videkov, Faculty of Electronic Equipment and Technology at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: videkov@tu-sofia.bg.

<sup>2</sup>Aleksei Stratev, Extramural PhD student at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria E-mail: astratev@ivastech.com.

<sup>3</sup>Georgi Furkov, Extramural PhD student at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: georgi\_farkov@festo.com

III. OPTICAL CONTROL

Both destructible and non-destructible optical control can be applied in the case of laser cut stencils. The shape and the dimensions of the holes can be viewed in horizontal plane in the case of non-destructible control. Variants of such an observation in two modes are showed on Fig. 2. In the case of observation in reflected light mode, one can see the surface of a stencil -1, the shape and location of different holes - 2, 3. For the purpose of precise control over the shape and dimensions of the holes, an observation method based on transmitted light is used and then figurations become contrasting - 4.

In the case of these stencils, the quality of the wall itself is crucial, not only the preciseness of the hole and the verticality of the walls. It might feature a different surface as well as defects - Fig. 3.

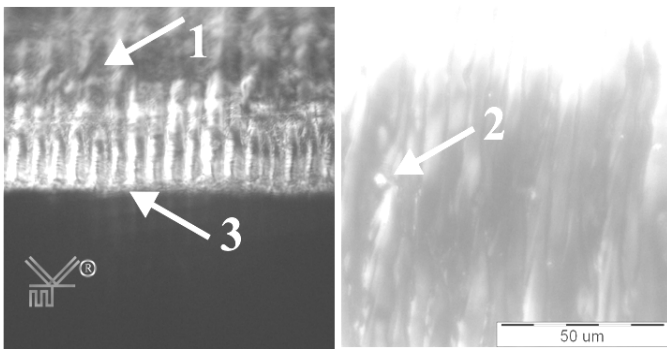


Fig.3 A wall of cut stencil. 1 – wall, 2 – defect, metal droplet, 3 – edge

Samples that have been cut along the holes have been used, allowing for better observation, while enabling observation of edges and walls - Fig. 4.

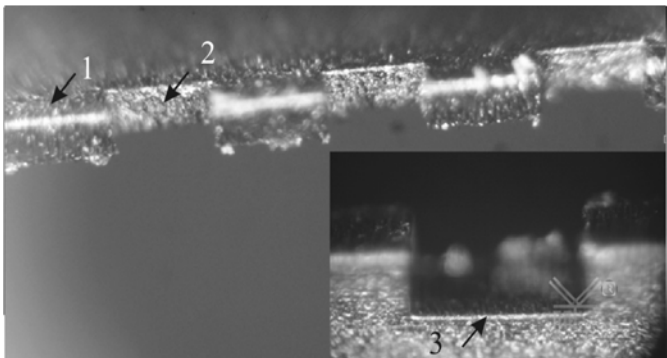


Fig. 4 Samples for observation of cuts

At the above figure, one can see the surface of the stencil - 1, its wall - 2, and the edge of the hole - 3, where angled lighting is used.

In optimizing the modes of operation of the laser (power, impulse frequency, duration, movement speed), the possibility to compare the outcome of cutting against the consequential surface of the wall is essential. It can take the shape of channels, grooves, lusterless, and others - Fig. 5.

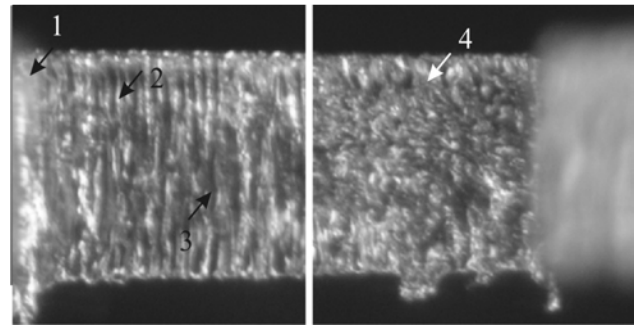


Fig. 5 Different surface in laser cutting

Fig. 5 shows the shapes of the wall at edge (angle of the hole) - 1, top of a groove - 2, pore - 3, lusterless surface - 4. The direct optical observation does not allow to assess the surface of the cut in detail despite that special samples have been made.

The height of the grooves, the presence of sub-surface cavities, etc., should be assessed. The assessment of the surface as a profile of the plane in different levels is of special interest.

IV. EXPERIMENTAL RESULTS

A standard approach to profile a surface involves the use of mechanical profile-measuring devices. Here, there are two restrictions. The wall of the hole is relatively small and narrow, requiring high level of precision during manipulation. On the other hand, the mechanical profile-measuring device is not applicable in the case of complex shapes - for example, curved groove.

Bearing the abovementioned in mind, a decision has been made to make horizontal cross-sections, taking profiles of the hole. For this purpose, epoxy resin has been poured on samples of stencils, which have been placed vertically to the cutting plane - Fig. 6.

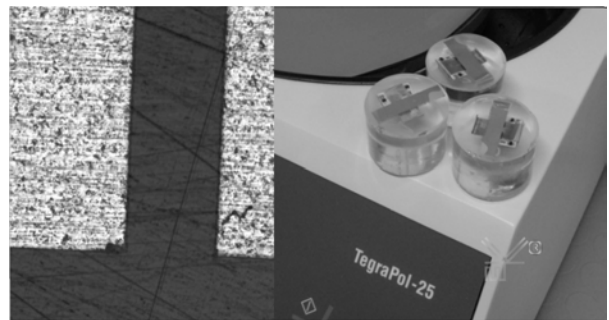


Fig. 6 Samples of cutting in horizontal plane.

The width of the sample has been under continuous control during the cutting process to determine the cutting plane. The cutting has been performed at 250 rpm, pressure of 30 Newtons per sample, and processing period of 30 seconds. Cut levels have been within 30 μm. In the beginning, cuts appeared at the levels of the metal droplets that occurred due to the cut - Fig.7, and then at the surface of the stencil.

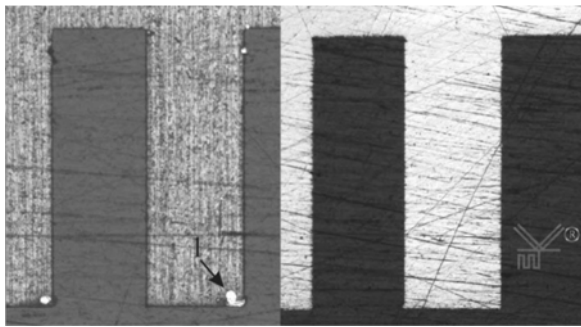


Fig. 7 Horizontal cutting of a stencil. 1 – metal droplets.

When performing cutting at particular distances, the shapes of the wall as resulting from different handling techniques emerged – Fig. 8.

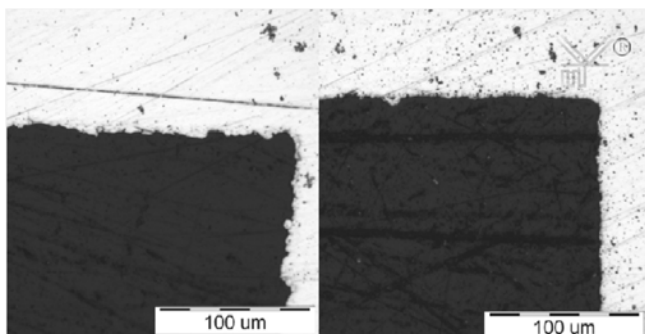


Fig. 8 Different modes of cutting. Cuts after grinding.

Software has been used to process the optical image to increase contrast and get sharp contour – Fig. 9.



Fig. 9 Contrast processing of a cutting.

A diagram of the contour itself is easy to extract, following the application of a respective processing method, where such a contour exists – Fig. 10.

The transformation of the resultant contour into digital data is not hard to realize as well. All and every cut is processed using the demonstrated methodology and is transformed into digital data. Those may be presented graphically and may be used in different types of processing. The below figure 11 shows a sample how the results derived from the wall's contours can be displayed as a 3-D diagram.

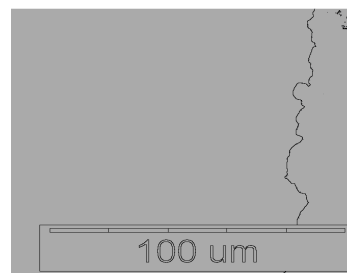


Fig. 10 Contours of the cut surface after extracting

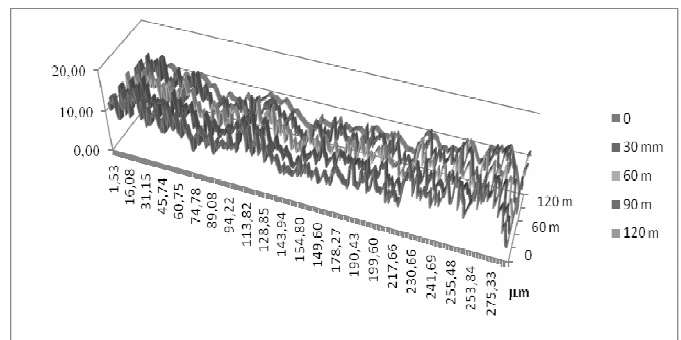


Fig. 11 3D diagram

## V. CONCLUSION

This paper provides a method for destructive control over laser cut stencils designed for surface-mount assembly. This method allows for a wall profile of the hole, measured in absolute measurements, to be achieved and makes it possible to complement the assessment of cutting parameters. Combining profiles that have been accomplished at different cut levels with digital presentation of those profiles allows for creation of a 3-D digital image of the surface.

The efforts to obtain such a digital image using other methods, like mechanical scanning through profile-measuring device, face difficulties.

## REFERENCES

- [1] <http://www.lc-led.com/ecommerce4.html> open 2012-02.
- [2] <http://www.national.com/mpf/LM/LMV243.html#Overview> open .2011-12.
- [3] Tae-Hyoung Park at all, Path Planning of Automated Optical Inspection Machines for PCB Assembly Systems, International Journal of Control, Automation, and Systems, vol. 4, no. 1, pp. 96-104, February 2006
- [4] Steplewski, W., at all, Stencil Design for Lead-Free Reflow Process, Electronics Technology, 30th International Spring Seminar on 9-13 May 2007, ISBN: 987-1-4244-1218-1, pp. 330 - 334
- [5] Stratev A., Farkov G., Videkov V., Optical Control of Laser Cut Stencils, XLVI International Scientific Conference on Information, communication and energy systems and technologies, Proceedings of Papers, V2, Serbia, Nish, 2011, ISBN 978-86-6125-3, pp. 529 – 532



# Power Quality According to EN 50160

Nikolce Acevski<sup>1</sup>, Kire Mijoski<sup>2</sup> and Tomce Mijoski<sup>3</sup>

**Abstract** – Although electricity, because of its possibility for easy conversion into other types of energy is the most used and most abundant form of energy, its features, except the frequency and amplitude of voltage is standardized too late.

Determination of additional criteria for evaluating the quality of electricity has become necessary with the increasing of the use of electric consumers with nonlinear character. Such electrical devices and equipment, on one hand they need quality voltage for their operation, on the other hand inject many disturbances that negatively affect the characteristics of voltage which is connected. These are a kind of customer whose work is mainly based on microprocessor and logic circuits and particularly is important to be exposed to less interruption at work because they can cause major economic damage. With the increasing of the number of nonlinear consumers it is also increased the interest for power quality in recent decades.

**Keywords** – Distribution network, electric power quality, European norm EN 50160, network performance analyser.

## I. INTRODUCTION

Electrical energy is a product and, like any other product, should satisfy the proper quality requirements. If electrical equipment is to operate correctly, it requires electrical energy to be supplied at a voltage that is within a specified range around the rated value. A significant part of the equipment in use today, especially electronic and computer devices requires good power quality (PQ). However, the same equipment often causes distortion of the voltage supply in the installation, because of its non-linear characteristics, i.e. it draws a non-sinusoidal current with a sinusoidal supply voltage. Thus, maintaining satisfactory PQ is a joint responsibility for the supplier and the electricity user. The responsibilities of the suppliers and the consumers of electricity are defined by internationally accepted electrical standards. In our area the most suitable standard for power supplying in distribution networks is the European standard EN 50160.

This paper presents the results of measuring the PQ at several points in DN using auxiliary sophisticated network analyzer.

## II. POWER QUALITY

### A. Term of PQ

Different ways of understanding what PQ is are present worldwide, including Macedonia. For example, power

suppliers equalize the concept of power quality with supply reliability (continuity in delivery), defining the system as 99.98 percent reliable. Manufacturers of electrical equipment can define electrical energy as being quality if their equipment functions properly. Of course, such approach is subjective and open to criticism. But, however, the most reasonable, and in any case there must be a priority, the definition behind which stands the user of electricity, its consumer. The PQ analysis generally includes the following voltage parameters: power frequency, supply voltage variations, rapid voltage changes (and flicker), supply voltage dips, short interruptions, long interruptions, temporary over voltages, supply voltage unbalance, harmonic voltage, mains signalling voltage.

### B. Standard EN 50160

EN 50160 gives the main voltage parameters and their permissible deviation ranges at the customer's point of common coupling in public low voltage (LV) and medium voltage (MV) electricity distribution systems, under normal operating conditions.

Supporting the requirement to define voltage characteristics in terms of frequency, magnitude, waveform and symmetry, EN 50160 provided definitions and in some cases measurement methods and compliance levels for 12 characteristics of the supply voltage:

TABLE I  
EN 50160 COMPLIANCE LIMITS

No	Parameter	Supply voltage characteristics according to EN 50160
1	Power frequency	LV, MV: mean value of fundamental measured over 10 s $\pm 1\%$ (49.5 - 50.5 Hz) for 99.5% of week -6%/+4% (47- 52 Hz) for 100% of week
2	Voltage magnitude variations	LV, MV: $\pm 10\%$ for 95% of week, mean 10 minutes rms values
3	Rapid voltage changes	LV: 5% normal 10% infrequently Plt 1 for 95% of week MV: 4% normal 6% infrequently Plt 1 for 95% of week (formula 1)
4	Supply voltage dips	Majority: duration <1s, depth <60%. Locally limited dips caused by load switching on: LV: 10 - 50%, MV: 10 - 15%
5	Short interruptions of supply voltage	LV, MV: (up to 3 minutes) few tens - few hundreds/year. Duration 70% of them < 1 s
6	Long interruption of supply voltage	LV, MV: (longer than 3 minutes) <10 - 50/year
7	Temporary, power frequency overvoltages	LV: <1.5 kV rms MV: 1.7 Uc (solid or impedance earth) 2.0 Uc (unearthed or resonant earth)
8	Transient overvoltages	LV: generally < 6kV, occasionally higher; rise time: ms - $\mu$ s. MV: not defined
9	Supply voltage unbalance	LV, MV: up to 2% for 95% of week, mean 10 minutes rms values, up to 3% in some locations
10	Harmonic voltage	95% of the time in 1 week, THD <8% (Table 2, formula 2)

<sup>1</sup>Nikolce Acevski is with the Faculty of Technical Sciences, 7000 Bitola, R. Macedonia, e-mail: nikola.acevski@uklo.edu.mk.

<sup>2</sup>Kire Mijoski master student, Faculty of Technical Sciences, 7000 Bitola, R. Macedonia, e-mail: kire.mijoski@yahoo.com

<sup>3</sup>Tomce Mijoski master student, Faculty of Technical Sciences, 7000 Bitola, R. Macedonia, e-mail: tomcemijoski@yahoo.com.

11	Interharmonic voltage	LV, MV: under consideration
12	Mains signaling voltage	9% - 100 Hz; 1% - 100 kHz, 99% of the time in 1 day

$$P_{it} = \sqrt[3]{\sum_{i=1}^{12} \frac{P_{sti}^3}{12}} \quad (1)$$

$$THD = \sqrt{\sum_{h=2}^{40} (U_h)^2} \quad (2)$$

TABLE II

VALUES OF INDIVIDUAL HARMONIC VOLTAGES AT THE SUPPLY TERMINALS FOR ORDERS UP TO 25, GIVEN IN PERCENT OF UN

Odd harmonics				Even harmonics	
Not multiples of 3		Multiples of 3		Not multiples of 3	
Order <i>h</i>	Relative voltage (%)	Order <i>h</i>	Relative voltage (%)	Order <i>h</i>	Relative voltage (%)
5	6,0%	5	6,0%	5	6,0%
7	5,0%	7	5,0%	7	5,0%
11	3,5%	11	3,5%	11	3,5%
13	3,0%	13	3,0%	13	3,0%
17	2,0%	17	2,0%	17	2,0%
19	1,5%	19	1,5%	19	1,5%
23	1,5%	23	1,5%	23	1,5%
25	1,5%	25	1,5%	25	1,5%

### III. SYSTEM ANALYZER OMNI-QUANT

The OMNI-QUANT mobile is preferably used as portable device in changing locations. Four voltage and current measuring inputs each one allows power measurements, fault analyses and recording functions along with the evaluation of the voltage quality.

Complete scanning and calculation of the following values:

- Phase voltage (L-N) and phase-to-phase voltage (L-L)
- Star point voltage and symmetry L1...L3
- Frequency (identical for all channels)
- Current, total current L1...L3, total current L1...L3 + N
- Power (P, Q, S, power factor, distorted reactive power)
- Power of the fundamental (active power, reactive power, apparent power, cos)
- L1...L3 total of the above power variables
- Harmonics 1...50th order
- Intern harmonic of U and I up to 2.5 kHz
- Ripple control level
- Distortion factor (THD) of V and A.

### IV. EXAMPLES OF MEASURING THE PQ ACCORDING TO EN 50150

#### A. Measuring

Practical measurement of power quality using measuring device OMNI-QUANT was made in several times in different points in distribution system.

The first measuring was performed in the period from 01.11.2011 to 01.12.2011 at 20 kV voltage level of the TS Polog 110/20 kV / kV. It is important to note that this is completely automatic TS with automatic voltage regulation and integrated device for continuous control of power quality.

Next measurement is performed in the period from 20.01.2011 to 28.01.2011 at 0,4 kV voltage level. Measurements were performed simultaneously at supply terminal of one consumer and at the LV level of TS Crniliste 20/0,4 which supplies this consumer.

#### B. Showing the results

After finishing the process of measurement, data is transferred to the PS. Then, by using special software package the summary results of measurements are obtained according to standard EN 50160. Because of the size of the number of the data it is not possible to display all the data and charts provided from measurements. It will be shown only those who are most interesting and most important for this paper.

In the graphical results presented in additional in this paper, the limit values according to standard EN50160 are marked by violet color.

Line voltage variations at 20 kV voltage level of TS Polog are given in the first picture. It can be noticed that the line voltages variations throughout the period were within the allowable limits according EN 50160.

On the second and the third picture are given the results for short interruptions of supply voltage and values of flicker severity and total harmonic distortion factor at 20 kV level of TS Polog 110/20 kV/kV.

It can be seen that on 10.11.2011 the THD is more than 7.5 % (very close to 8%) and there is deviation of flicker severity (Plt>1) about several hours in 26.11.2011, but still the values of these parameters also vary within the allowable limits according to EN 50160.

As it was mentioned above that this is completely automatic TS with automatic voltage regulation and integrated device for continuous control of power quality and this significantly affects to the values of all measured parameters to moving in approved limits.

In Fig. 4 and Fig. 5 are presented the results from measurement of PQ at 0.4 kV voltage level of TS Crniliste. Similar as in the previous example (TC Polog) most of the parameters are moving within the limits according to the standard EN 50160. From the shown parameters, the flicker severity, the total harmonic distortion factor and the supply voltage unbalance did not deviate from the allowable limits prescribed by EN 50160 over the entire period. With the exception on 23.01.2011 when it has a deviation of phase voltages above the limits for several hours.

In contrast to the values measured at 0.4 kV level of TS Crniliste, the values measured in the same period at supply terminal of one consumer supplied by TS Crniliste, significantly differ. In fact, with the exception of the total harmonic distortion factor, which moves within the allowable limits, the other parameters significantly deviate outside the limits. The results from this measurement are presented in Fig. 6 and Fig. 7.

### V. CONCLUSION

Today's electronic loads are susceptible to transients, sags, swells, harmonics, momentary interruptions, and other disturbances that historically were not cause for concern. For sensitive loads, the quality of electric service has become as important as its reliability. Power quality is a new phenomenon. Events such as voltage sags, impulses, harmonics, and phase imbalance are now power quality

concerns. Power quality problems have a huge economic impact. As a result, any discussion of power system reliability must also include power quality.

The main document dealing with requirements concerning the supplier's side is standard EN 50160, which characterizes voltage parameters of electrical energy in public distribution systems.

In this paper were presented several measurements of PQ using modern network analyzer OMNI-QUANT. By using a special software package the summary results of measurements are obtained according to standard EN 50160.

From the first measurement can be seen that measured values fully satisfy the limits prescribed in EN 50160. This PQ is due to the fact that this is completely automatic TS, with automatic voltage regulation and integrated device for continuous control of power quality.

Serious deviations from the limits according to EN 50160 are detected during measuring the PQ at supply terminal of one consumer supplied by TS Crniliste. LV lines with great length and TS with outdated equipment are the reason of such results for the PQ. Hence, the investments in distribution systems are necessary to achieve satisfactory results for the PQ.

REFERENCES

- [1] Angelo Baghini. Handbook of Power Quality. University of Bergamo, Italy: 2008.
- [2] Dr. Željko Novine. „Kvaliteta Elektri ne Enrgije“. Priru nik. Sveu ilište J. J. Strossmayera, Osijek:2006.
- [3] . “ ” :2008.
- [4] EN 50160, Voltage characteristics of electricity supplied by public distribution systems, 1999.
- [5] H. Markiewicz, A. Klajn “Voltage Disturbances Standard EN 50160 - Voltage Characteristics in Public Distribution Systems” Wroclaw University of Technology July 2004.
- [6] . . . . . “ ” . : 2010.

VI. ADDITION



Fig. 1. Voltage magnitude variations at 20 kV level of TS Polog 110/20 kV/kV

Page 1

Evaluation period: from 06.11.2011 | 13:12:43.202 to 26.11.2011 | 04:00:23.481  
Limits: c:\haagewa\damon\en50160(230).lim

OQ	Polog ETRI-20kV	Events total: 6	
	0.0000 sec <= Dur < 60.000 sec	:	Num(evtUlow) = 6
	60.000 sec <= Dur < 180.00 sec	:	Num(evtUlow) = 0
	180.00 sec <= Dur < +Inf sec	:	Num(evtUlow) = 0
	Tot	:	Num(evtUlow) = 6
OQ	Polog ETRI-20kV	Events total: 11	
	Tot	:	Num(evtUrel) = 11

Fig. 2. Statistical report for short interruptions of supply voltage in the observed period



Fig. 3. Flicker severity and total harmonic distortion factor at 20 kV level of TS Polog 110/20 kV/kV

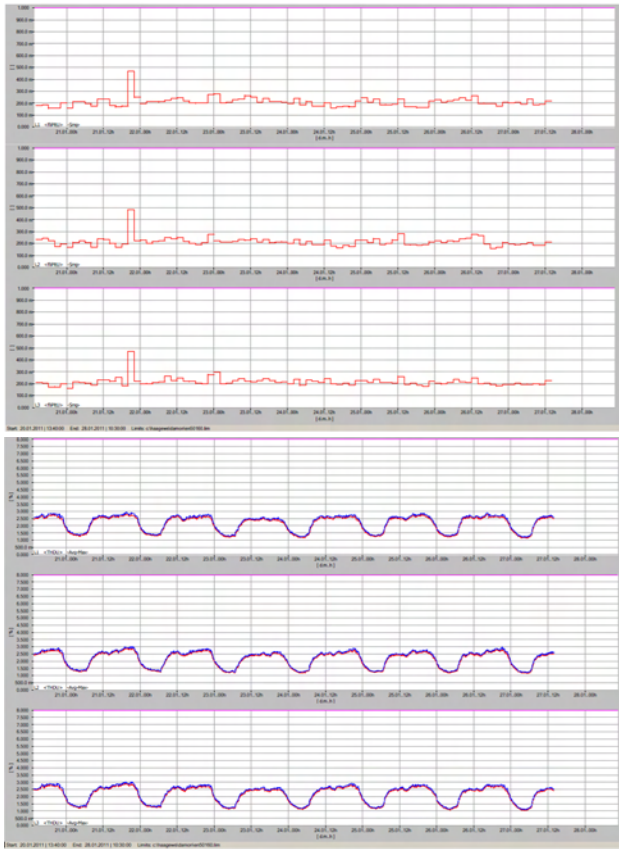


Fig. 4. Flicker severity and total harmonic distortion factor at 0,4 kV level of TS Crniliste 20/0,4 kV/kV

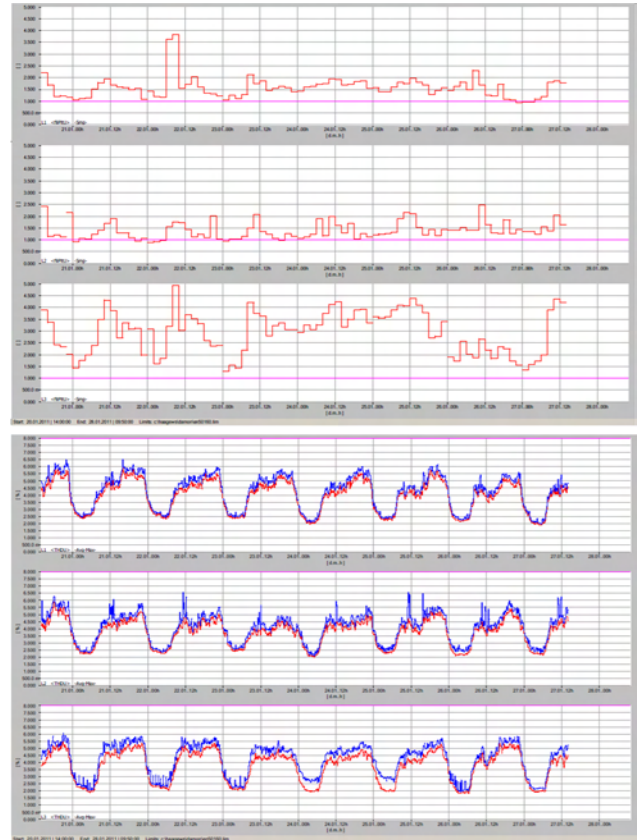


Fig. 6. Flicker severity and total harmonic distortion factor at supply terminal of one consumer supplied by TS Crniliste

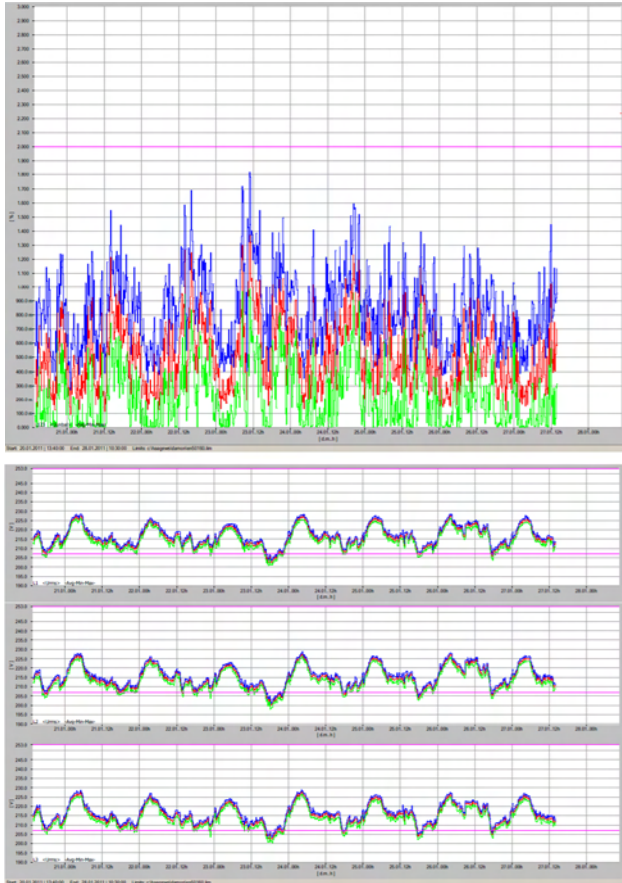


Fig. 5. Supply voltage unbalance and voltage magnitude variations at 0,4 kV level of TS Crniliste 20/0,4 kV/kV

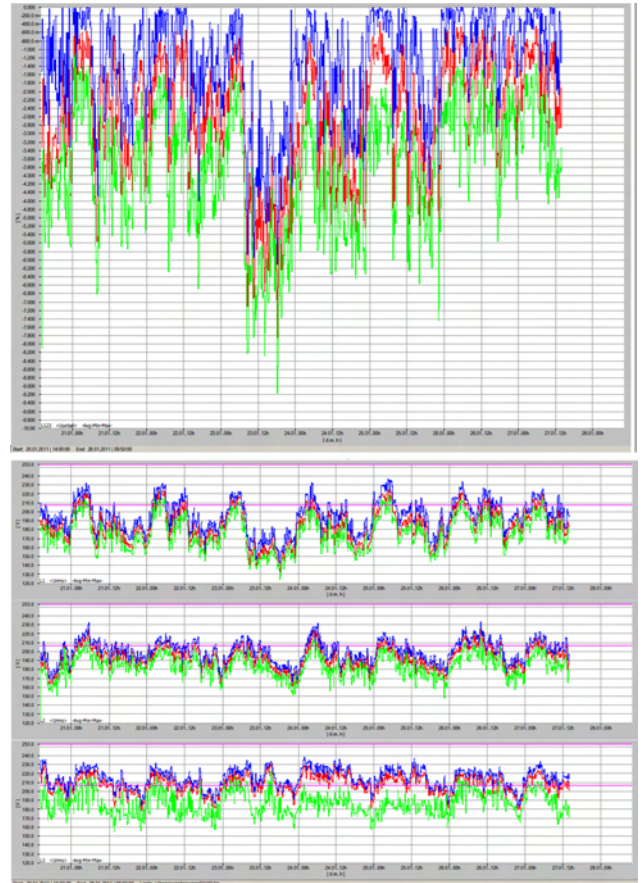


Fig. 7. Supply voltage unbalance and voltage magnitude variations at supply terminal of one consumer supplied by TS Crniliste

# Using $\mathcal{H}_\infty$ synthesis for finding settings of single channel power system stabilizers of synchronous generators

Konstantin Gerasimov<sup>1</sup>, Petko Petkov<sup>2</sup> and Krum Gerasimov<sup>3</sup>

**Abstract** – This paper presents a methodology for finding settings of single channel power system stabilizers by approximation of the frequency response of the synthesized  $\mathcal{H}_\infty$  controller satisfying the requirements for maximal damping of the synchronous generator electromechanical oscillations and minimization of the measurement noise. Test results are presented for a real synchronous generator from the Bulgarian electric power system. The advantages of the proposed methodology are discussed.

**Keywords** –  $\mathcal{H}_\infty$  synthesis, power system stabilizer, tuning

## I. INTRODUCTION

Modern power plants are equipped with power system stabilizers (PSS) for damping the electromechanical oscillations of synchronous units. In accordance to their structure we differentiate them as single- and dual-channel. A typical single channel PSS with rotor speed as input signal (PSS1A) [1] is shown in Figure 1.

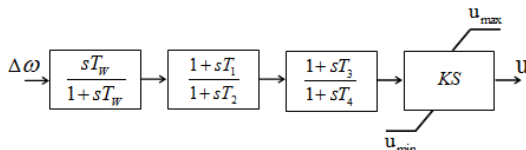


Fig. 1. Block-diagram of PSS1A

In the electric power system (EPS) of Bulgaria mainly single channel PSS are used, with input signal form the equivalent sum of the generator active power ( $P_e$ ) and rotor speed ( $\omega$ ). This equivalent input signal is obtained after the signals of  $P_e$  and  $\omega$  pass through input filters and then once again through a torsion filter which rejects the torsion oscillations originating from the generator rotor. These PSS are classified as type PSS2A and PSS2B.

The general structure of PSS2A of Alstom is shown in Figure 2. The difference between the different manufacturers' modifications is in the number of the phase-shifting blocks included (and in some elements in the input filters). For example, in Bulgaria there are PSS2A from Alstom with

4 phase-shifting blocks and there are as well PSS2A of ABB with 2 blocks.

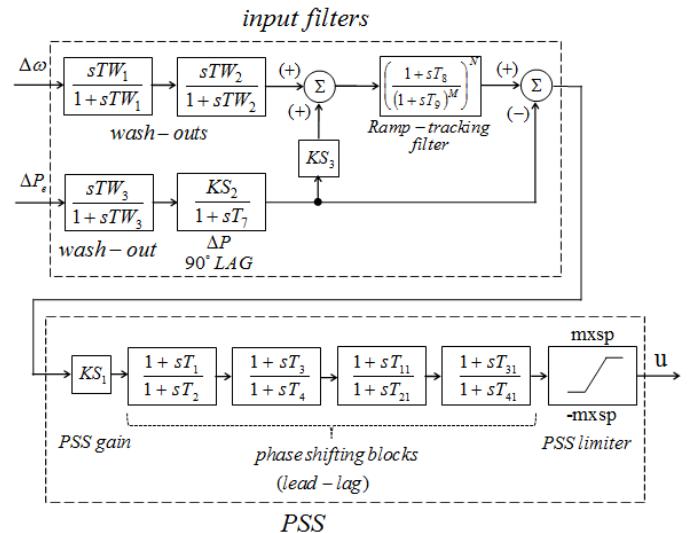


Fig. 2. Block-diagram of Alstom realization of PSS2A

The input filters are tuned in such a way that they do not pass the settled deviations of the regime parameters, and the PSS phase-shifting blocks – to maximally damp the electromechanical oscillations. The settings can be determined by a variety of different methodologies [2-4].

The purpose of this paper is to present a methodology, developed by the authors, for single channel PSS tuning based on  $\mathcal{H}_\infty$  synthesis, and to discuss its advantages.

## II. METHODOLOGY FOR SINGLE CHANNEL PSS TUNING BASED ON $\mathcal{H}_\infty$ SYNTHESIS

### A. Mathematical model

For analysis of the electromechanical oscillations of the motors in EPS a mathematical description linearized around a certain operating point is used. The size of this mathematical description is too big due to the great number of elements in the modern united systems. Because the purpose is to tune a particular PSS of a particular synchronous generator, the authors have developed a methodology for frequency aggregation of the multidimensional EPS mathematical description in respect to the studied generator buses [5,6]. The descriptions results in the following structure:

<sup>1</sup> Konstantin Gerasimov is a PhD student at the Department “Power Engineering”, Technical University of Varna, Studentska Str. 1, Varna, 9010, Bulgaria, E-mail: kkgerasimov@tu-varna.bg

<sup>2</sup> Petko Petkov is a Professor at the Faculty of Automation, Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: php@tu-sofia.bg

<sup>3</sup> Krum Gerasimov is a Professor at the Department “Power Engineering”, Technical University of Varna, Studentska Str. 1, Varna, 9010, Bulgaria, E-mail: k.gerasimov@tu-varna.bg

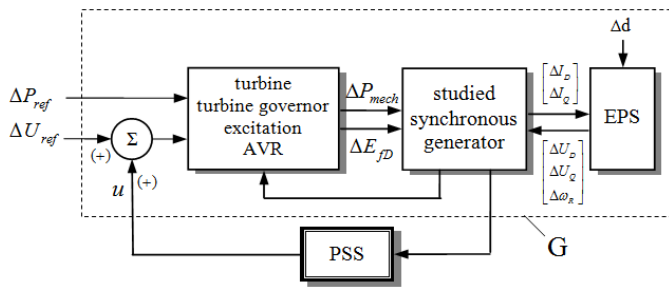


Fig. 3. Block-diagram of the linearized mathematical description for determination of PSS settings of a generator

The mathematical description of the building elements in Figure 3 are obtained according to [2,5,7].

B.  $\mathcal{H}_\infty$  synthesis

The general formulation of the  $\mathcal{H}_\infty$  control problem can be presented by the block-diagram in Figure 4 [8,9].

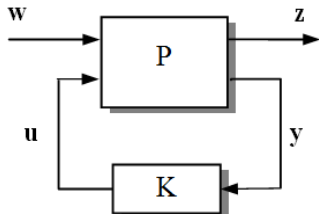


Fig. 4. Block-diagram of  $\mathcal{H}_\infty$  control design

In this form the “external” input  $w$  is the vector of all signals which come into the system and the “error”  $z$  is the vector of all signals which are necessary to describe the behavior of the closed-loop system.  $P$  contains the plant transfer matrix  $G$  and weighting functions which are specific for every synthesis problem.  $K$  is being synthesized (searched) control function. The standard task for  $\mathcal{H}_\infty$  optimal control is to find a stabilizing function  $K$  which minimizes:

$$\|F_l(P, K)\|_\infty = \max_{\omega} \bar{\sigma}(F_l(P, K)(j\omega)) \quad (1)$$

In MATLAB® this task is solved by the hinfsyn function from Robust Control Design® 3 toolbox.

The synthesis of control function based on signals is a common approach to MIMO problems for which simultaneously a few different (and usually controversial) goals are required. In this particular problem the following goals are set:

- maximal damping of the electromechanical oscillations manifested in rapid damping of the rotor speed deviation and the generators active power deviations. Thus the generators influence over the rest of the EPS during transient processes will be minimized;
- maximal filtration of the measurement noise. Passing this noise through PSS leads to high frequency oscillations in the excitation circuit and thus the other generator regime parameters. It is even possible that there may be a 50 Hz component in the noise which can result in very troublesome resonance phenomena.

Having in mind the above, the synthesis model in Figure 4 objectifies to the structure in Figure 5:

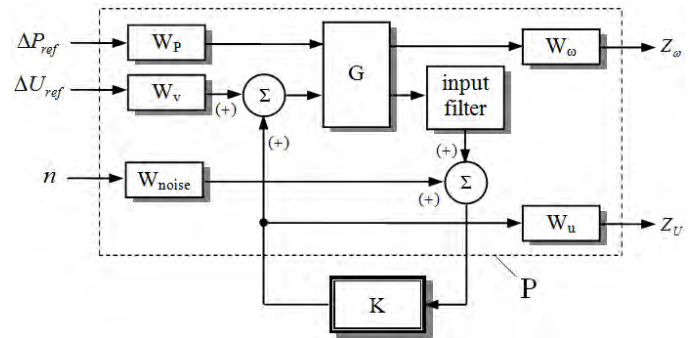


Fig. 5. Block-diagram of the  $\mathcal{H}_\infty$  synthesis, based on signals, of stabilizer  $K$

The weighting functions of reference ( $W_p$ ,  $W_v$ ) and disturbance ( $W_{noise}$ ) can be constant or dynamic and describe the relative importance and/or the frequency contents of the inputs. The weighting function  $W_\omega$  sets the requirements in respect to the degree of damping of the rotor speed, and  $W_u$  – the requirements concerning the limitations of the PSS output signal. As seen in Figures 1 and 2 the PSS output is equipped with non-linear limiters.

C. Algorithm

The algorithm for single channel PSS tuning consists of the following steps:

- 1) Formulation of the mathematical description of the studied generator, as shown in Figure 3;
- 2) Formulation of weighting functions and of transfer matrix  $P$ , as shown on the block-diagram in Figure 5;
- 3) Using the constructed transfer matrix  $P$  an  $\mathcal{H}_\infty$  stabilizer is synthesized by means of the MATLAB® function hinfsyn;
- 4) Tuning of the fixed-structure single channel PSS in Figure 1 or in Figure 2 so that its frequency response is as close as possible to the frequency response of the mathematically synthesized stabilizer. This is achieved by searching for coincidence, of the phases or of the amplitudes of the frequency response in the frequency range of the electromechanical oscillations, using approximating functions which solve non-linear least square problems.
- 5) Analysis of the behavior of the tuned PSS. The step and frequency response are recalculated and the fulfillment of the goals, set during the synthesis, is assessed. If the results are unsatisfactory, first a return to step 4 is made and the structure and parameters of approximation are varied. If even this cannot lead to satisfactory results a return to step 2 is made where the weight functions have to be reconsidered and from there on the process repeats.
- 6) Construction of model with uncertainties for the purpose of the robust analysis. This can be done with the help of the developed by the authors software tool **RobustPSS** [10], allowing modeling of structured uncertainty, presented in state space, and of unstructured

uncertainty, presented in the frequency domain, which are caused by the uncertainties in the generator load and the system operating point.

- 7) Analysis of the robust stability. It is done by means of the MATLAB<sup>®</sup> function robuststab.

### III. TEST RESULTS

The proposed algorithm is tested for tuning Alstom PSS2A of a real 370 MW synchronous generator from a Bulgarian thermal power plant. The frequency equivalentation of the EPS in respect to the generator bus is reduced to order of 20. The weighting functions used in the synthesis are as follows:

$$W_p = 1; \quad W_v = 1;$$

$$W_{noise} = \frac{0.0521 \cdot s^3 + 9.48 \cdot s^2 + 1293 \cdot s + 7.429 \cdot 10^4}{s^3 + 1696 \cdot s^2 + 2.147 \cdot 10^5 \cdot s + 9.018 \cdot 10^7};$$

$$W_u = 0.8; \quad W_w = \frac{46.17 \cdot s^2 + 41.5 \cdot s + 1.442}{s^2 + 0.1198 \cdot s + 0.0003}$$

The  $\mathcal{H}_\infty$  synthesis is carried out under the assumption that the input filters are tuned well because it is a common practice, due to subjective reasons, that it is not allowable to change the settings of the input filters. In this particular case the settings are as follows:  $ks_3 = 1$  p.u.;  $TW_1 = 7$  s;  $TW_2 = 7$  s;  $TW_3 = 7$  s;  $ks_2 = 0.86$  p.u.;  $T_7 = 7$  s;  $T_8 = 0.6$  s;  $T_9 = 0.15$  s;  $M = 4$ ;  $N = 1$ .

Under these conditions the  $\mathcal{H}_\infty$  synthesized controller has frequency response shown in Figures 6 and 7. In Figure 6 it is compared with PSS2A tuned by approximation of the amplitude and in Figure 7 – tuned by approximation of the phase. One should not forget that the fixed-structure PSS settings can vary only in certain ranges. In this case  $T_1, T_3, T_{11},$  and  $T_{31} = 0 \div 10$  s, while  $T_2, T_4, T_{21},$  and  $T_{41} = 0.015 \div 3$  s. Due to subjective reasons, the authors have chosen  $ks_1$  to vary between 5 and 20 p.u.

It is clear that in this particular case the approximation by amplitude gives better results and this is why it will be used. The obtained in this way settings are:  $ks_1 = 5$  p.u.;  $T_1 = 0.0951$  s;  $T_2 = 0.0367$  s;  $T_3 = 0.6686$  s;  $T_4 = 0.0367$  s;  $T_{11} = 0.0967$  s;  $T_{21} = 0.0367$  s;  $T_{31} = 0$  s;  $T_{41} = 0.0664$  s.

The gain  $ks_1$  is relatively small and we can afford to increase it 1.8 times without the PSS output signal to reach the PSS output signal limitation (see Figure 8).

In general we could fine adjust the gain  $ks_1$  because it doesn't affect the PSS phase compensation and exactly it is crucial for the right operation of PSS. But by changing  $ks_1$  we change the degree of damping. For this reason from here on the presented results are for  $ks_1 = 9$  p.u. and the effect can be clearly seen in Figure 9. It shows a significant damping of the rotor speed oscillations when the tuned PSS is switched on. A significant damping of the active power oscillations is achieved as well and this can be seen in Figures 10 and 11.

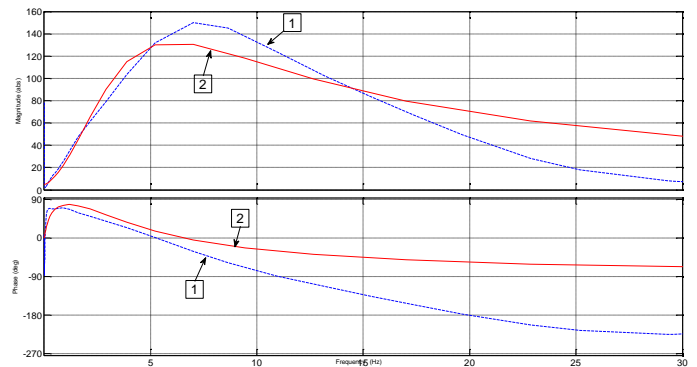


Fig. 6. Frequency response of: 1 – the  $\mathcal{H}_\infty$  synthesized controller 2 – PSS2A, tuned by approximation of the **amplitude**

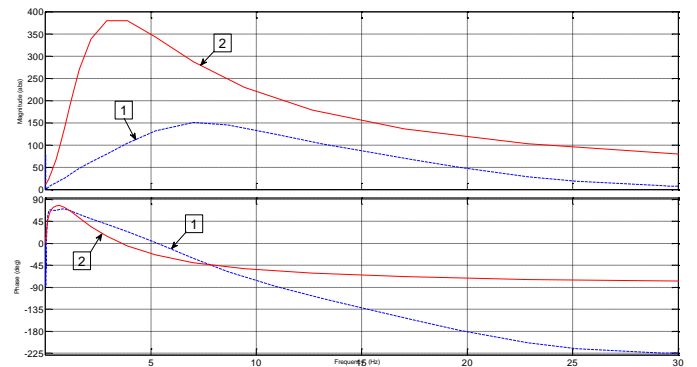


Fig. 7. Frequency response of: 1 – the  $\mathcal{H}_\infty$  synthesized controller 2 – PSS2A, tuned by approximation of the **phase**

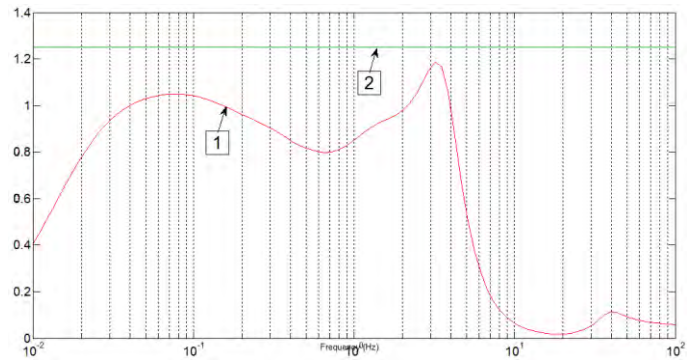


Fig. 8. Frequency response of PSS2A (1), tuned by approximation of the amplitude and  $ks_1 = 9$  p.u., compared to the PSS output signal limitation (2)

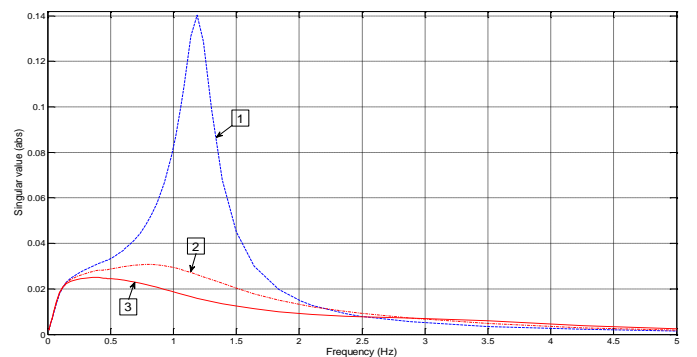


Fig. 9. Maximal singular values of transfer matrix from all inputs to  $\Delta\omega$  of the generator: 1 – without PSS; 2 – with PSS2A ( $ks_1 = 5$  p.u.); 3 – with PSS2A ( $ks_1 = 9$  p.u.)

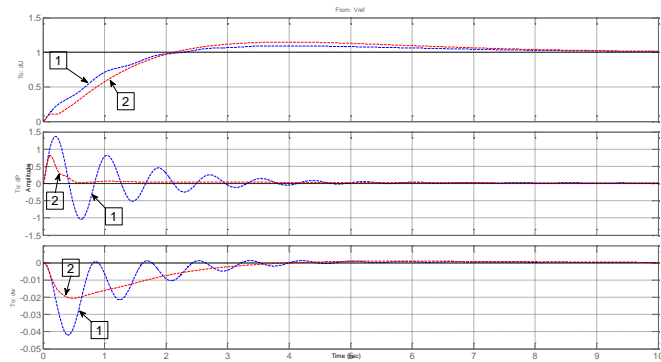


Fig. 10. Step response of the nominal generator model for step change of  $V_{ref}$ : 1 – without PSS; 2 – with PSS2A ( $ks_1 = 9$  p.u.)

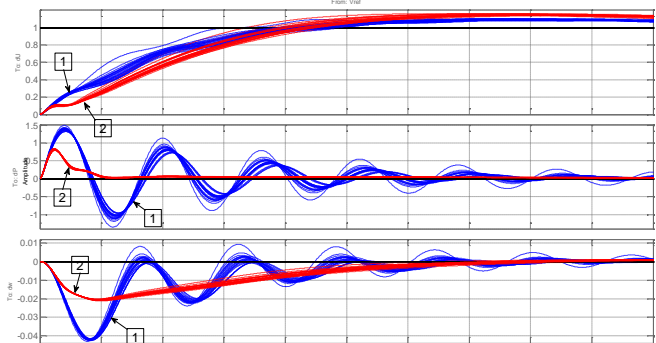


Fig. 11. Step response of the generator, modeled with uncertainties, for step change of  $V_{ref}$ : 1 – without PSS; 2 – with PSS2A ( $ks_1 = 9$  p.u.)

The uncertain model used for robust stability check consists of output unstructured multiplicative uncertainty of 10% in the EPS model and structured uncertainty in the generating unit state space realization describing change of the active power in its whole allowable range. In Figure 12 is shown that a sufficient robust stability margin of 137 % is achieved.

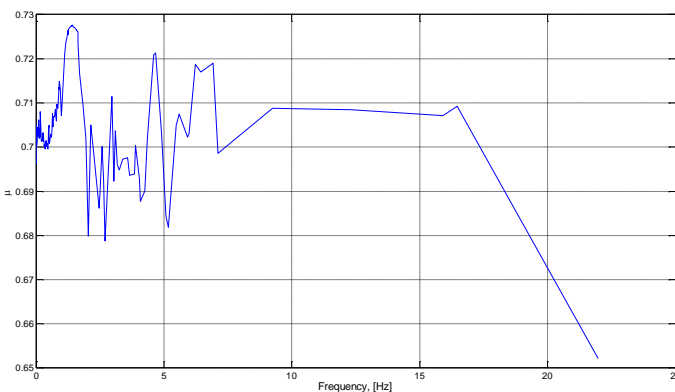


Fig. 12. Upper bound on the mixed structural singular value  $\mu$  of the generator with PSS2A with  $ks_1 = 9$  p.u.

#### IV. CONCLUSION

As a conclusion it can be summarized that the proposed algorithm has as main advantage that given a particular *fixed structure PSS*, one can tune it considering *simultaneously* the following limiting requirements:

- 1) ensure quality of the transient processes (better damping of the generator electromechanical oscillations);
- 2) the PSS output signal doesn't reach its limitation;
- 3) suppresses the measurement noise.

#### ACKNOWLEDGEMENT

The authors would like to express their appreciation of eng. Venci Zahov (head of division "System analyses & Transmission Planning" in National Dispatching Center by the Bulgarian Electricity System Operator) for his generous and active collaboration.

#### REFERENCES

- [1] Rogers Gr., *Power System Oscillations*, Springer, 1999, ISBN: 978-0792377122
- [2] Gerasimov K., *Mathematical modeling of electromechanical transients in electric power systems and algorithms for analysis of stability at small disturbances*, D.Sc. Dissertation, 2006, TU-Varna (in Bulgarian)
- [3] Gerasimov Kr., Y. Rangelov, K. Gerasimov, Y. Kamenov. Methodology for tuning PSS2A system stabilizers, *International Journal of Reasoning-based Intelligent Systems (IJRIS)*, ISSN (Online): 1755-0564, InderScience Publishers
- [4] Murdoch, A., Boukarim, G. Performance criteria and tuning techniques, *IEEE Tutorial Course – Power System Stabilization via Excitation Control*, June 2007, pp. 26-34, ISBN: 978-1424462308
- [5] Kamenov, Y., K. Gerasimov, Y. Rangelov. Structuring the Nominal Mathematical Model of the Electric Power System for the Aims of Robust Analysis, *Proc. of ICEST 2011, Niš, Serbia, June 29 - July 1, 2011, Volume 2, pp.481-484*, ISBN: 978-8661250323
- [6] Gerasimov K., *Frequency method for analysis of small signal stability of large electric power system*, Ph.D. Dissertation, 1983, TU-Sofia (in Bulgarian)
- [7] Notov P., K. Gerasimov, *Transient processes in electric power systems*, TU-Sofia, 1997, ISBN: 9544382054
- [8] Petkov P., M. Konstantinov, *Robust Control Systems*, ABC Tehnika, 2002. (in Bulgarian) ISBN: 9548873516
- [9] Petkov P., G. Lehov, A. Markovski, *Handbook on Robust Control Systems*, ABC Tehnika, 2006. (in Bulgarian) ISBN: 978-9548873772
- [10] Gerasimov K., Y. Kamenov, *Software tool for robustness analysis of PSS in electric power plants*, TU-Varna Yearbook, 2011. (in Bulgarian) ISSN: 1311-896X



# Technical Conditions for PV Plants Connection on the MV Distribution Grids in the Republic of Macedonia

Ljupco Trpezanovski<sup>1</sup>, Metodija Atanasovski<sup>2</sup> and Dimitar Dimitrov<sup>3</sup>

**Abstract** – In this paper the technical conditions, which have to be fulfilled for connection of a Photovoltaic Power Plant (PV plant) on distribution MV grid in the Power System (PS) of the Republic of Macedonia, are given. As the base, the technical conditions from the aspect for obtaining a permission from the distribution system operator for connection of a PV plant, the following conditions are described: nodes voltages changing, flickers appearances, increasing of higher-harmonics currents, increasing of short currents intensity in the grid, changing of feeder load segment distribution and location of energy-meter. The approach and results for checking the technical conditions which have to be fulfilled for a real case 1 MW PV plant connection to a 10 kV distribution grid are presented.

**Keywords** – Renewable energy sources, Photovoltaic system, Photovoltaic plant connection, Dispersed generation.

## I. INTRODUCTION

According to the Base Study on Renewable Energy Resources in the Republic of Macedonia [1] and Strategy for power plants development in the period 2008-2020 in the Republic of Macedonia [2], it is planning to construct 10 – 30 MW PV Power plants till 2020, with total electricity production of 16 – 60 GWh/year. There are two categories of PV plants: 1. with installed power  $\leq 50$  kW and feed-in tariff 30 €/kWh and 2. with installed power between 50 kW and 1 MW and feed-in tariff 26 €/kWh. At the moment, the total allowed capacities which the Government would support by feed-in tariffs are: 2 MW for the first and 8 MW for second category. Recently, the Energy regulatory commission of the Republic of Macedonia issued a register of connected PV plants [6]. Till now, six PV plants from the first (total capacity of 220.765 kW) and two from the second (total capacity of 1246.7 kW) category have been put into operation.

According to the technical recommendation [3], [4] and Distribution grid code [5], there are several technical conditions which have to be fulfilled in order to connect a PV plant on distribution grid. In the following sections the technical conditions which have to be fulfilled by the PV plant, necessary for obtaining permission for connection to the MV grid of the distribution company EVN, are explained.

<sup>1</sup>Ljupco Trpezanovski is with the Faculty of Technical Sciences at University St. Clement Ohridski – Bitola, I. L. Ribar bb. 7000 Bitola, Republic of Macedonia, E-mail: ljupco.trpezanovski@uklo.edu.mk.

<sup>2</sup>Metodija Atanasovski is with the Faculty of Technical Sciences at University St. Clement Ohridski – Bitola, I. L. Ribar bb. 7000 Bitola, Republic of Macedonia.

<sup>3</sup>Dimitar Dimitrov is with the Faculty of Electrical engineering and Information Technologies, University St. Cyril and Methodius – Skopje Karpos II, 1000 Skopje, Republic of Macedonia.

## II. CONDITIONS FOR THE ALLOWED PV PLANT INSTALLED POWER

The installed power of the PV plant has impact to the voltages in the distribution grid during the plant switching-on or switching-off transient period. This voltage impact should not exceed value of  $\Delta u_m = 2\%$  at the plant connection point to the distribution grid. The small PV plant may connect to the distribution grid according to the criterion of allowed installed power, only if the condition given by Eq. 1 is satisfied:

$$S_{nPV} \leq \frac{S_{3pc}}{50 \cdot k} \quad (1)$$

where:

- $S_{nPV}$  is rated installed apparent power of the plant in MVA,
- $S_{3pc}$  is three phase short-circuit power in MVA and
- $k$  is coefficient, which has value 1 for inverters DC/AC.

In the MV distribution grids in the power system of the Republic of Macedonia, there are three standard values for maximal permitted three phase short-circuit currents (powers):

- for 10 kV grid, current of 14.5 kA and power of 250 MVA,
- for 20 kV grid, current of 14.5 kA and power of 500 MVA,
- for 35 kV grid, current of 12 kA and power of 750 MVA.

It is worth to mention, that no matter how big is installed power of the PV plant, the maximum voltage deviation in the connection point in steady-state conditions, should not exceed the interval  $\Delta u_m = \pm 5\%$  from the MV grid rated voltage [3].

## III. CONDITION FOR FLICKERS GENERATING

The flicker criteria can be assessed by the disturbance factor  $A_{fd}$  for the PV plants with flicker duration more than two hours. This factor can be calculated with Eq. 2:

$$A_{fd} = \left( c_f \cdot \frac{S_{nPV}}{S_{3pc}} \right)^3 \quad (2)$$

where:

- $c_f$  is flicker coefficient ( $S_{nPV}$ ,  $S_{3pc}$  are defined in section II).
- With this coefficient is assigned the ability of the PV plant to produce flickers. After PV plant connection to the distribution grid, the flicker coefficient should be measured in real operating steady-state conditions. The value of this coefficient usually is  $c_f > 20$ , but it can reach a value up to 40. For every PV plant there should be issued certificate that condition for long time duration flickers is fulfilled. Calculated value of the disturbance factor with Eq. 2, should be  $A_{fd} \leq 0.1$  as a proof that the PV plant would not generate flickers [3].

#### IV. CONDITION FOR HIGHER-HARMONICS CURRENTS

The criterion of permitted content of higher-harmonics currents can be checked by applying Eq. 3 [3]:

$$I_{hhp} = I_{hhr} \cdot S_{3pc} \quad (3)$$

where:

- $I_{hhp}$  is the permitted value of the higher-harmonic current on the generator voltage level, in A,
- $I_{hhr}$  is the value of the higher-harmonic current in A/MVA (reduced on the  $S_{3pc}$  in the connection point on the grid).

The values of maximum permitted content of higher-harmonics currents reduced on the three phase short-circuit power in the PV plant connection point on the grid, are given in Table 1.

TABLE I  
MAXIMUM PERMITTED HIGHER-HARMONICS CURRENTS

harmonic v	5	7	11	13	17	19	23	25
$I_{hhr}$ A/MVA	0.7	0.6	0.5	0.3	0.3	0.2	0.2	0.2

If the current of any higher-harmonic  $v$  exceeds the permitted value, the voltage for the  $v$ -th harmonic  $U_{hv}$  should be calculated. If the voltage of the 5-th higher-harmonic is  $U_{h5} \leq 0.2\% \cdot U_n$  and for the rest higher-harmonics from Table 1  $U_{hv} \leq 0.1\% \cdot U_n$  (where  $U_n$  is rated voltage of the grid), than the criterion of permitted higher-harmonics currents is fulfilled. On the contrary the owner of the PV plant should undertake measures to eliminate higher-harmonics currents [3].

Another way to check the condition for higher-harmonic currents is to compare total harmonic distortion factor  $THD$  allowed by the distribution company, with the certified  $THD$  factor of the DC/AC inverter [8].

#### V. CONDITION FOR SHORT-CIRCUIT POWER

If the connection of the PV plant causes increasing of the three phase short-circuit power (current) over the allowed values for the equipment, the following measures should be undertaken:

- to limit the short-circuit currents in the PV plant,
- to replace the switching and other equipment which not meet requirements for short-circuit currents,
- to connect the PV plant in other place of the grid.

Because the PV plants with installed power less or equal to 1 MVA don't influence significantly on short-circuit power (current) increasing, it is not necessary to check this criterion for PV plants constructed in the Republic of Macedonia.

The relay protection of the MV feeder will trip the feeder circuit-breaker in cases of short-circuit faults in the feeder. In these cases is not allowed island operation of the PV plant (or PV plants if more then one are connected on the same feeder).

The relay protection of the PV plant should switch of the plant from the grid immediately.

The automation installed in the PV plant will allow switching on the plant to the grid when all three phases in the grid have normal operating voltages. The conditions for synchronized connection are voltage difference  $\Delta U_{pv} < \pm 0.1 U_n$ ,  $\Delta f_{pv} < \pm 0.5 \text{ Hz}$  and phase angle difference  $\Delta \phi < \pm 10\%$  [3], [4].

#### VI. CONDITION FOR ELECTRICITY MEASURING

Despite that in the Grid Code [5] is not clarified, the distribution company obliges the dispersed producers, the measuring of produced and consumed electrical energy should be performed on one indirectly measuring place on 10 (20) kV side, with measuring transformers placed in separate measuring cabinet [8].

The current transformers (CT) in each phase, should have two cores X/5/5 A, 50Hz, 10/25 VA, class 0.5/5P10 ( $F_s < 5$  for the first core and  $F_s > 10$  for the second core). The voltage transformers (VT) in each phase should have the following characteristics:  $10000/\sqrt{3}$  ( $20000/\sqrt{3}$ )/100/ $\sqrt{3}$ /100/ $\sqrt{3}$  V/V, 50 Hz, 25/25 VA, class 0.5/3P.

Electricity meter have to be with two-way measuring system of produced and consumed electrical energy. The type of this meter must be 5(6) A,  $3x(100/\sqrt{3})/100$  V, class 1/2 (A/R) with optical port, internal connecting clock, CS/RS communication and connected with modem to the system of distance reading in the company EVN Macedonia. No other additional equipment is allowed to be connected on the secondary windings of the measuring transformers.

The measuring transformers and the electric energy meter are provided by the distribution operator - company EVN Macedonia and stay in their property.

#### VII. CASE STUDY: 1 MW PV PLANT CONNECTION

In this paper a case study for connection of 1 MW PV plant on 10 kV distribution feeder named "Egri", from substation TS 35/10 kV/kV Bukri, is analyzed. The schematic diagram of the PV plant main components is shown on the Fig. 1.

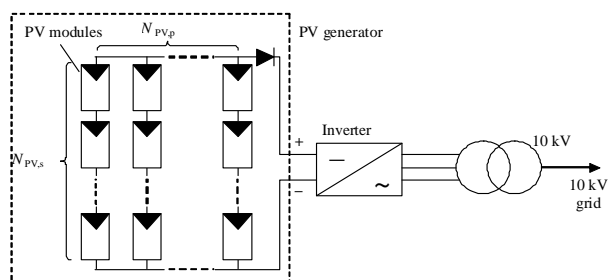


Fig. 1. Schematic diagram of the PV plant main components.

Total number of the modules will be 4320 each per 230 Wp. The nominal power of the DC generator of the PV plant will be 993.6 kWp. Because the power factor for this type of plants is  $\cos \phi \approx 1.0$  and modules power is with positive tolerance in the following calculations can be taken that  $S_{nPV} \approx 1 \text{ MVA}$ .

The DC/AC inverter which is planned to be used is Sunny Central HE 1000MV for direct medium-voltage feed-in, because the MV/LV transformer is build in the same house.

Distribution feeder named “Egri” with all its elements as a part from 10 kV grid is shown on Fig. 2.

All data for the elements connected in and between the nodes are given on the same Fig. 2. A new PV plant will be connected in the node J4.

B. Check for the flickers generating

As it was mentioned in section III., the flicker coefficient  $c_f$  should be measured in real operating steady-state conditions. However, if the maximum possible value of this coefficient  $c_f=40$  is taken into Eq. 2 and  $S_{3pc}=250$  MVA,  $S_{nPV}=1$  MVA, the calculated value of disturbance factor  $A_{fd}=0.0041$ .

In this case  $A_{fd} \ll 0.1$ , so it can be concluded that the new

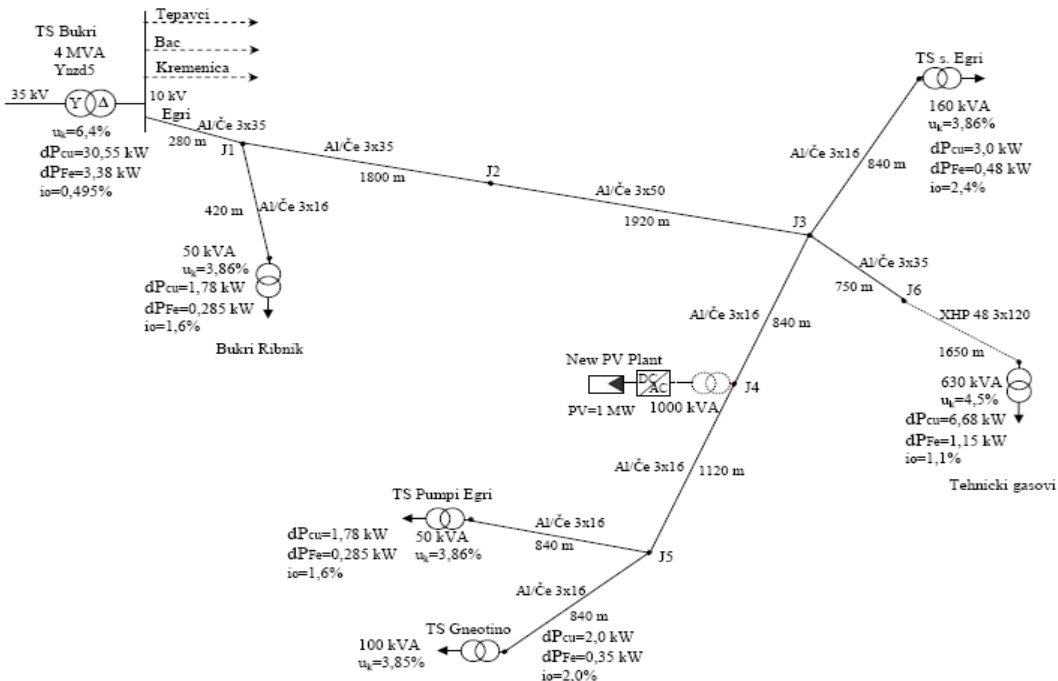


Fig. 2. 10 kV feeder named “Egri” with data for the connected elements and connection point of 1 MW PV plant.

A Load-flow and short-circuit analysis for the 10 kV feeder shown on Fig. 2 are performed with Neplan 5.0 software [9]. Two cases are analyzed. In the first case the new PV plant is not connected and in the second case this plant is connected in node J4. Also the calculations for checking the technical conditions which have to be fulfilled to connect this PV Plant of 1 MW, are done.

A. Check for the allowed PV plant installed power

According to Eq. 1 and taking into account the values for three phase short-circuit power (250 MVA for 10 kV grid) and  $k=1$ , the maximal allowed installed power for the PV plant connected in node J4 can be  $S_{nPV} \leq 5$  MVA. Because the installed power of new PV plant is 1 MVA (actually 1 MW because PV plants work with  $\cos\varphi=1.0$ ) it is obvious that condition for maximal allowed installed power is fulfilled.

The load-flow analysis confirmed that PV plant connection wouldn't change the node voltages and line currents over permitted values. In this case, the voltage drops and energy losses in the lines are smaller than in the first case. Also the power losses in the entire grid are reduced [7].

PV plant will not generate flickers with duration more than two hours. After construction of the PV plant the certificate for measured  $c_f$  should be issued.

D. Check for the higher-harmonics currents

Under the normal operating conditions the total harmonic distortion factor  $THD$  allowed by distribution company EVN Macedonia is 8% [8]. According to the inverter manufacturer data, this factor has value  $THD < 3\%$ . Taking into account these data, it is obvious that this condition is fulfilled and the generated higher-harmonic currents have acceptable values.

E. Check for the short circuit power (current)

The DC generator of the PV plant works as constant current source. For the planned PV modules with power of 230 Wp maximum power point current is  $I_{MPP}=7.8$  A and maximum short circuit current is  $I_{sc}=8.3$  A. Increasing of the short circuit current is only 6.4% over the  $I_{MPP}$ . Because of the presence of the PV plant in the MV grid, the total increasing of the three phase short circuit current and power at the point of

connection will be also 6.4%. This is not significant increase which can take negative influence on the electrical equipment.

### VIII. CONCLUSION

In the last few years, the construction and connection of the PV plants in MV grids in the PS of the Republic of Macedonia are occurred. The technical regulations for PV plants were very poor and lot of these documents were directly accepted from the foreign regulations which sometimes do not coincide with the PS situation and the present legislation. Distribution operator EVN Macedonia introduced additional regulations in this domain.

This paper deals with necessary technical conditions which have to be fulfilled for PV plant connection on the MV grid. The conditions as: the allowed PV plant installed power, the flickers generating, the increasing of the higher-harmonics currents, the increasing of the short circuit currents (power) intensity in the grid, the location and the type of energy-meter are described. These technical conditions are checked for a real PV plant of 1 MW which is planned to be connected in the 10 kV distribution grid in the middle of 2012. All necessary and proper parameters are calculated and compared with allowed values given by the grid code.

From the calculations and values of obtained results, it can be concluded that connection of PV plants with power equal or less than 1 MW will not have negative influence to the normal operation conditions of the MV distribution grid and the appertaining electrical equipment.

### REFERENCES

- [1] Team of ICEIM-MANU. *Base strategy for use of renewable energy sources in R. Macedonia until 2020*, Government of R. Macedonia, 2010.
- [2] Team of ICEIM-MANU. *Strategy for energy sector development in R. Macedonia until 2030*, Government of R. Macedonia, 2010.
- [3] Group of Authors, *Technical Recommendation 16. Base conditions for connection of small power plants on the distribution grid of Serbia (adopted by EVN Macedonia)*, JP Elektroprivreda Srbije, 2003.
- [4] Group of Authors, *Technical Recommendation for application of valid laws and rules for connection the power plants from renewable energy on distribution grid*, HEPODS doo, distribution system operator of Croatia, 2010.
- [5] Management committee of EVN Macedonia, *Grid Code for distribution of electrical energy*, Official gazette of the Republic of Macedonia, no. 83, 2008.
- [6] Energy regulatory commission of the Republic of Macedonia, [www.erc.org.mk](http://www.erc.org.mk)
- [7] Lj. Trpezanovski, M. Atanasovski, *Analyze of 10 kV distribution feeder "Egri" from TS Bukri, before and after connection of PV plant of 1 MW from investor Torpedo-Mobil Bitola*, Faculty of Technical Sciences Bitola, 2010.
- [8] EVN Macedonia, Decision for 1 MW PV plant "Torpedo-Solar GT" connection on the distribution grid, EVN Macedonia, AD Skopje, 2010.
- [9] Neplan 5.0 software, BCP Switzerland (Educational version).

# Control cards. Control cards and control points as part of the manufacturing process.

Violina Georgieva<sup>1</sup> and Alexander Hadjidimitrov<sup>2</sup>

**Abstract** – In process manufacturing the control over the process usually is more important than the control over the product itself. In production departments the registration of the operations and the quality control is done using “Control cards”

The control card is used afterwards in the analyses of defects, claims, the load of machines and employees, etc.

**Keywords** – Manufacturing, Production process, Control card, Control point, Integrated system.

## I. INTRODUCTION

The current paper describes the usage of control cards as a tool for quality control of the production process. The authors share their experience in the creation of a software module for control cards management as a part of an enterprise resource planning (ERP) system. This module is integrated to the existing manufacturing module of the ERP system. The system is implemented at a manufacturing enterprise that produces electronic equipment (cash registers, fiscal printers, scales, etc.). Before the creation of the module the control cards were created manually on paper. The implementation of the module automated the process of Control Cards management and increased the control over it. The paper describes the basic concepts of control cards, the main issues that are met during their creation, the structure and the concepts of the module created. The usage is illustrated with examples from the enterprise, where the system is implemented.

## II. COMPANY STRUCTURE

The main departments in the company are:

- Trade (international and domestic) – manages the sales of production and the purchase of materials and goods.
- Manufacturing of products
- Manufacturing of Printed circuit boards (PCBs) – Semi-finished products
- Software – develop software and firmware for the products produced
- Development – develop new products and modifications of the existing products
- Accounting
- Management

Control cards are used in production departments (products and PCBs.)

The manufacturing process consists of two stages – production of semi-finished products (PCBs) and production of products.

The production of PCBs consists of:

- Delivery of the components needed
- Print of the blank PCB
- SMD assembly
- Visual control
- Conventional assembly
- Visual control
- Control card filling
- Transfer of the produced PCBs to the manufacturing department

The production of the products consists of:

- Preliminary operations
- Test of PCBs, recording of firmware
- Assembly of the components of the product
- Assembly of the final product
- Packaging

The registration of the planning and production processes in the system consists of the following stages:

- Sales order – it is created every time when a new order from a customer is received.
- Main production schedule – this is the planning stage. It is based on the sales orders received. Usually it is prepared once a week. As a result new production orders for the products needed are created.
- Start of the production orders – each order is started when the production of the product in the manufacturing department is started.
- Registration of the production – at this stage are filled up the serial numbers and lot numbers. The control cards are also filled.
- Consumption of materials – the system proposes the materials to be consumed based on the “Bill of materials” of the product. Manual adjustment of the quantities proposed is allowed.
- Finalization of the production order.

## III. APPLICATION OF CONTROL CARDS

The quality control of the products is based on the control of the production process itself. The documentation of the process is done using control cards. Thus is tracked that all

<sup>1</sup> Violina Georgieva – PhD. Student, Technical University Sofia, Faculty of Automatics. Sofia 1000, 8 “Sv. Kliment Ohridski” Blvd., violina\_jg@abv.bg.

<sup>2</sup> Alexander Hadjidimitrov – ERP consultant at Team VISION Bulgaria Ltd. Sofia 1712, 41 “Alexander Malinov” Blvd., Alexander@Hadjidimitrov.com.

operations needed are done and also is tracked who are the employees that have done the operations. This way two important targets are achieved:

- Following the sequence of operations described in the control card is guaranteed that no operations are skipped;
- In case of defective product is easy to find which operation has caused the defect and to trace who is the employee who has done the operation and when is the operation done. Thus the reasons of the defect could be analyzed and actions to be undertaken in order to change the process to avoid future problems of the same type.

The most important thing during the creation of a control card is to define a proper list of control points. Each control point describes one operation or a group of operations that are done in conjunction during the production process. Defining the list of operations should keep in mind that operations should be grouped in control points in a way that no operation is skipped but also the list of control points should be short enough in order to be easy to be documented and followed. It's advisable to group related operations that are done together by the same employee. As an example if during the assembly of a product 5 Integrated Circuits (ICs) should be mounted in the sockets on a PCB by the same employee these could be grouped to one control point ("IC mounting") as during this operation the employee receives a PCB with the sockets mounted, mounts the ICs in the sockets and afterwards transfers the ready PCB to the next stage. In other case if the employee should mount only three of the ICs, a test should be carried out and afterwards the two other ICs should be mounted then maybe it would be better to separate the process in 3 control points ("Mounting of the first group of ICs", "Test", "Mounting of the second group of ICs"). This will help the process to be tracked properly in order to allow future analyses in case of defects or claims.

In order to be possible to use control cards in the production process is required to prepare templates of the control cards for each product (semi-finished or final) that is tracked separately (e.g. for each PCB, each cash register, etc.) Each template contains a list of control points listed chronologically during the production process. If some operation is repeated at different stages during the process (e.g. testing of different modules) then it is documented as several control points at the appropriate positions (one control point for each occurrence) in the list.

A sample template of a control card could look like:

- Main board test
- Recording of firmware
- Operator display test
- Fiscal module test
- Top cover assembly
- Bottom cover assembly
- Display assembly
- Final product test
- Final product assembly
- Packaging

This control card template becomes a part of the product specification. It is an integral part of the production documentation. It is used in the preparation of the current control cards of the products during the production process.

Current control cards are filled up during the production process. Based on the production process organization different approaches are possible:

- A separate control card is filled in for each product (for each instance).
- A separate control card is filled in for each lot of products (e.g. for the products produced by a single work shift).
- A single control card is filled in for the total quantity produced.

In regard of the selected approach some modifications in the control cards are possible. In general it contains the following fields:

- ID and name of the product
- ID of the production order
- Production date
- Serial number / Lot number (or a list of numbers)
- A list of control points
- For each control point – a flag that the operation is done and the ID of the employee (or a list of employee IDs) who has done the operation. If the control card is for a production period longer than one day – also a production date should be added to this list [3].

#### IV. IMPLEMENTATION OF THE CONTROL CARDS IN THE ERP SYSTEM

Depending on the enterprise and the software used for production planning and management different approaches in the implementation of control cards are possible. In the case described in the current paper an ERP system is used. The functionality for control card management is implemented as an additional development especially for the specific needs of the enterprise. It is integrated to the standard functionality of the system for management of items (products) and production orders.

The control points are developed as a separate table. It contains the full list of operations that are tracked separately. These control points are used afterwards to assemble the control cards of all products. The table contains the following fields:

- ID – a unique ID of the point in the system
- Description a description of the tracked operation

The product control cards are developed as a separate table related to the Item table (that contains the list of items in the system). This way for each item is specified the list of the control points that should be tracked. The table contains the list of the control points that form the control card of the item.

The control points are selected from the global list of control points in the system (the table described above.) This table contains the following fields:

- Item ID – contains the unique ID of the item. It is related to the Item table of the system. Usually for a single item a several lines in the table with the same item ID are created – each line contains one control point from the item control card.
- Control point – contains the ID of the control point. The field is related to the table with the global list of control points.

This functionality is used for management of control cards during the production process. As the manufacturing functionality itself is based on production order the corresponding control cards are developed as documents related to the lines of the production orders. The production order itself is a document containing one or several lines. Each line describes the production of a single item. For each item are stored its ID and the quantity for production.

The functionality consists of the following tables:

- Production employee – contains the list of employees in the production departments. Contains the following fields:
  - ID – a unique ID of the employee in the system
  - Name – the name of the employee
- Production order line control card – contains a list of the control points for the item in the production order line. Contains the following fields:
  - Production order ID – the ID of the production order for which the control card is prepared;
  - Production order line ID – the ID of the line in the production order for which the control card is prepared;
  - Control point ID – ID of the control point in the control card;
  - Approved – a flag that means that the operations described in the control point are done successfully;
- Performed By – this table is related to the “Production order line control card” table. For each control point here are listed the employees who have performed the operations. The table contains the following fields:
  - Production order ID
  - Production order line ID
  - Control point ID
  - Employee ID – ID of the employee who has performed the operation

During the initial implementation of the system the global list of control points is set. Periodically in cases of need of new operations the list is updated with the new control points.

Also periodically when new employees are employed the list of production employees is updated.

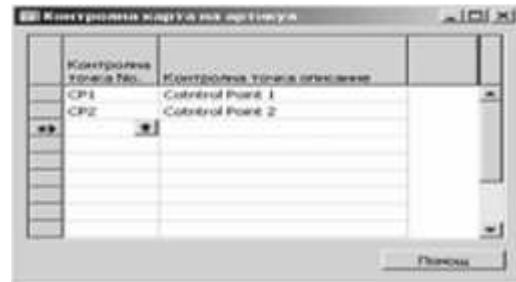


Fig. 1 Product control card template in case of definition of new products a new control card template is created (see Fig. 1).

During the creation of new production orders for each line that contains a product the system automatically creates a blank control card based on the template set for that product (see Fig. 2).

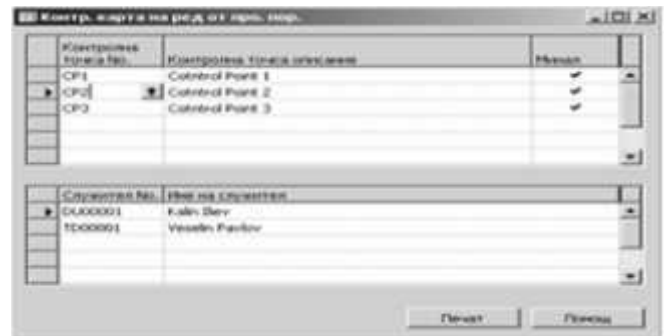


Fig. 2 Production order control card

When an operation is completed the employee in charge sets the “Approved” flag for the corresponding control point and fills the list of employees who have performed the operations.

The control card could be printed on paper at any moment if needed.

## V. BENEFITS FROM CONTROL CARD USAGE

The properly prepared control card template contains all operations that should be performed during the production of an item. This guarantees that during the production process no operation will be skipped (because otherwise the corresponding control point will not be filled in and the control card will be incomplete. This should be noticed when the production order is finished.) The ERP system described in this paper checks the control cards during the production order closure and does not allow the order to be finished if there are control points that are not marked as “Approved”.

The second benefit from control cards usage is the possibility for future analyses and control. They are performed in cases of defective production when is needed to find the possible reasons for the defects. Each produced item has a lot number or a serial number. Based on this number the system can identify the production order the item was produced by. In the production order control card could be identified the employees who have performed the operations. This allows to trace the production process for the specific item and to give

hints for the possible reasons for the defect [3].

## VI. RISKS

The successful usage of control cards could be compromised if some risks are neglected.

The first risk is the improper definition of the control points. As the control card is a list of the operations that should be tracked and controlled, the proper definition of this list is the foundation for the successful usage of control cards.

Possible problems include:

- Skipping important operations – in case a control point is not created for an important operation than that operation will not be tracked by the control card. As a result it could be skipped during the production and also it will not be documented and future control and analyses will be impossible.
- A very detail list of the operations – in case the operations are not grouped in a proper way to control points a very long list of control points might be produced. This could make the process of filling control cards very tough and the employees to stop to pay attention to it. As a results they could mark operations as “Approved” without checking their actual status and thus to compromise the process.

The second big risk is the human factor. In general the control cards are created by people and are used to document the work of people.

That’s why many errors and misuse could occur. For this reason it is very important during the implementation of this tool to guarantee the involvement of the team who is supposed to fill and use them in order to guarantee that the information filled is correct. Otherwise they could not pay enough attention to it and to feel it like an additional useless time-consuming documentation that should be filled. This also will compromise the process.

## VII.

## CONCLUSION

Control cards are a tool that allows guaranteeing of the proper flow of the production process and as a result to guarantee the quality in the terms of not skipping operations due to errors. For the proper usage of the tool however it is very important the proper implementation in the organization to be done and to guarantee the involvement of the team who is supposed to use them.

## REFERENCES

- [1] G. Tsvetkov, “Production management”, Sofia, 2006, “Softrade”.
- [2] I. Dakov, “Production engineering”, Sofia, 2003, “Luren” publishing house - “Luren commerce” Ltd.
- [3] E. Andronov, M. Aleksandrova, “Operations management”, „Stopanstvo”, Sofia, 2005.



# Computer Simulation and Analysis of Two-Coordinate Position Electric Drive Systems

Mikho Mikhov<sup>1</sup> and Marin Zhilevski<sup>2</sup>

**Abstract** – Analysis of some algorithms for position control of two-coordinate electric drive systems is presented in this paper. Models for computer simulation with various types of motors have been developed. Detailed studies by means of computer simulation and experimental research have been carried out. The results obtained can be used in the design and tuning of such types of drive systems with position control.

**Keywords** – Two-coordinate electric drive, Position control, Computer simulation.

## I. INTRODUCTION

Two-coordinate electric drive systems are widely used in many industrial applications.

Generally, motions in these systems are formed by the respective trajectories along both coordinate axes. Control algorithms affect the performance, productivity and energy consumption [2], [3], [4].

Time shortening of the transient regimes at positioning is essential for mass production of parts, because it increases the respective machine effectiveness.

Mathematical modeling and computer simulation provide very good opportunities to explore different control algorithms aiming at optimizing of motion trajectories [2], [3], [5].

With respect to modernization of a class of machine tools some two-coordinate electric drives have been analyzed, allowing the choice of the appropriate once meeting the required performance. Models of such drives have been developed used for studying of different position control algorithms for the respective dynamic and static regimes at different operation modes.

## II. FEATURES OF THE DRIVE SYSTEM

The simplified block diagram of the system under consideration is represented in Fig. 1, where the notations are as follows: CP – control panel; CD – control device; PC1 and PC2 – position controllers; SC1 and SC2 – speed controllers; CC1 and CC2 – current controllers; C1 and C2 – power converters; M1 and M2 – DC motors; SS1 and SS2 – speed sensors; DM1 and DM2 – driven mechanisms; CF1 and CF2 – current feedback blocks; SF1 and SF2 – speed feedback blocks; PF1 and PF2 – position feedback blocks;  $V_{pr1}$  and

$V_{pr2}$  – position reference signals;  $V_{sr1}$  and  $V_{sr2}$  – speed reference signals;  $V_{cr1}$  and  $V_{cr2}$  – current reference signals;  $V_{pf1}$  and  $V_{pf2}$  – position feedback signals;  $V_{sf1}$  and  $V_{sf2}$  – speed feedback signals;  $V_{cf1}$  and  $V_{cf2}$  – current feedback signals;  $\theta_1$  and  $\theta_2$  – angular positions;  $S_1$  and  $S_2$  – linear displacements.

The set of achievements required for the drive system can be formulated as follows:

- forming the necessary motion trajectories at given position cycles;
- maximum starting torque to ensure good dynamics;
- reversible speed and torque control;
- compensation of the disturbances.

## III. MODELLING OF THE DRIVE SYSTEM

The vector-matrix model of the DC motor drive under consideration is as follows:

$$\begin{bmatrix} \frac{d\theta_i}{dt} \\ \frac{d\omega_i}{dt} \\ \frac{di_{a_i}}{dt} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & \frac{K_{t_i}}{J_{\Sigma_i}} \\ 0 & -\frac{K_{e_i}}{L_{a_i}} & -\frac{R_{a_i}}{L_{a_i}} \end{bmatrix} \begin{bmatrix} \theta_i \\ \omega_i \\ i_{a_i} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \frac{K_{c_i}}{L_{a_i}} \end{bmatrix} v_i + \begin{bmatrix} 0 \\ -\frac{1}{J_{\Sigma_i}} \\ 0 \end{bmatrix} i_{l_i}, \quad (1)$$

where:  $\theta_i$  is angular position;  $\omega_i$  – motor speed;  $i_{a_i}$  – armature current;  $K_{e_i}$  – back EMF voltage coefficient;  $K_{t_i}$  – torque coefficient;  $R_{a_i}$  – armature circuit resistance;  $L_{a_i}$  – armature inductance;  $K_{c_i}$  – amplifier gain of the chopper;  $v_i$  – input control signal of the power converter;  $J_{\Sigma_i}$  – total inertia referred to the motor shaft;  $i_{l_i}$  – armature current which is determined by the respective load torque;  $i = 1, 2$  – number of the coordinate axes.

Both subsystems have identical cascade structures with subordinate regulation of currents, speeds and positions. Control loops optimization and tuning of the respective controllers have been done sequentially, starting from the innermost ones [1].

For the used mechanical gear maximum operating speed of the motor for the respective coordinate axis is chosen to be equal to the nominal value:

$$\omega_{\max_i} \leq \omega_{\text{nom}_i}. \quad (2)$$

<sup>1</sup>Mikho Mikhov is with the Faculty of Automatics at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: mikhov@tu-sofia.bg.

<sup>2</sup>Marin Zhilevski is with the Faculty of Automatics at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: electric\_zhilevski@abv.bg.

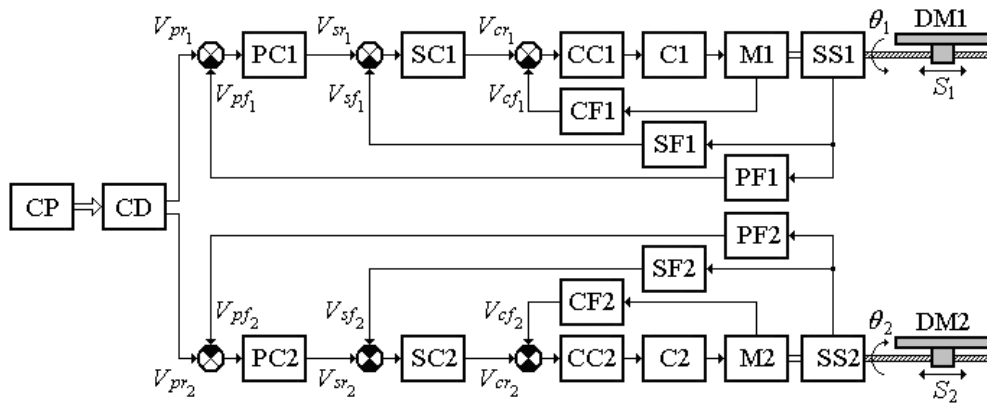


Fig. 1. Block diagram of the two-coordinate drive system.

The maximum rate of speed change in the respective axis can be determined from the following equation:

$$\varepsilon_{\max_i} = M_{\max_i} / J_{\Sigma_i}, \quad (3)$$

where  $M_{\max_i}$  is the maximum torque, which the respective motor can develop along this coordinate axis;  $J_{\Sigma_i}$  – total inertia referred to the motor shaft.

For the deceleration motion in this case the following relationship is valid:

$$\Delta\theta_{d\max_i} = \omega_{\text{nom}_i}^2 / 2\varepsilon_{d\max_i}. \quad (4)$$

Because subordinate regulation of coordinates is applied, the output voltage of the respective position controller is the assigning speed signal.

$$U_{sr_i} = K_{pc_i}(V_{pr_i} - V_{pf_i}) = K_{pc_i}K_{pf_i}(\theta_{r_i} - \theta_i) = K_{sf_i}\omega_{r_i}. \quad (5)$$

Equation (4) for  $\Delta\theta_{d\max_i}$  and  $\omega_{\text{nom}_i}$  takes the following form:

$$K_{pc_i}K_{pf_i}\Delta\theta_{d\max_i} = K_{sf_i}\omega_{\text{nom}_i}. \quad (6)$$

After substituting (4) to (6) and solving with respect to the position controller coefficient, the equation becomes:

$$K_{pc_i} = \frac{K_{sf_i}\omega_{\text{nom}_i}}{K_{pf_i}\left(\omega_{\text{nom}_i}^2 / 2\varepsilon_{d\max_i}\right)} = \frac{2K_{sf_i}\varepsilon_{d\max_i}}{K_{pf_i}\omega_{\text{nom}_i}}. \quad (7)$$

For the corresponding mechanical gear the linear speed and linear position can be determined as follows:

$$V_i = \omega_i / K_{gi}; S_i = \theta_i / K_{gi}, \quad (8)$$

where  $K_{gi}$  is the respective gear coefficient.

In general, when two-coordinate systems with position control are used, the motion trajectories are formed by the respective displacements of both axes.

Motion trajectories for the studied position control

algorithms are presented in Fig. 2, where the symbols used are: O (0, 0) – initial position; A ( $S_{1f}, S_{2f}$ ) – final position of the specified cycle.

Fig. 2a shows a trajectory obtained by successive movement along the coordinate axes. The total time for positioning is as follows:

$$t_p = t_{p1} + t_{p2}, \quad (9)$$

where:  $t_{p1}$  is the motion time along the coordinate axis  $x$ ;

$t_{p2}$  – the motion time along axis  $y$ .

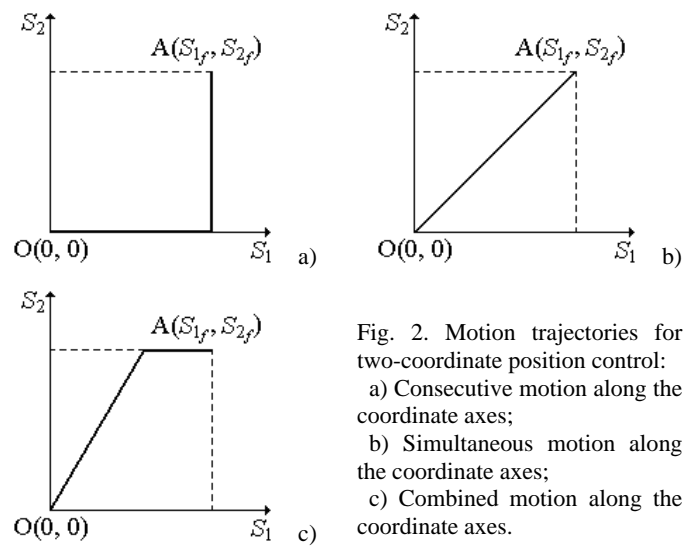


Fig. 2. Motion trajectories for two-coordinate position control:  
 a) Consecutive motion along the coordinate axes;  
 b) Simultaneous motion along the coordinate axes;  
 c) Combined motion along the coordinate axes.

Fig 2b represents a trajectory obtained by simultaneous movement along both coordinate axes. In such way of control position time is:

$$t_p = t_{p1} = t_{p2}. \quad (10)$$

Fig. 2c shows a trajectory obtained at combined motion along the coordinate axes. If both drives work at the same speeds, the total time of positioning is equal to the time necessary for the drive with longer displacement time set.

$$t_p = t_{p1}. \quad (11)$$

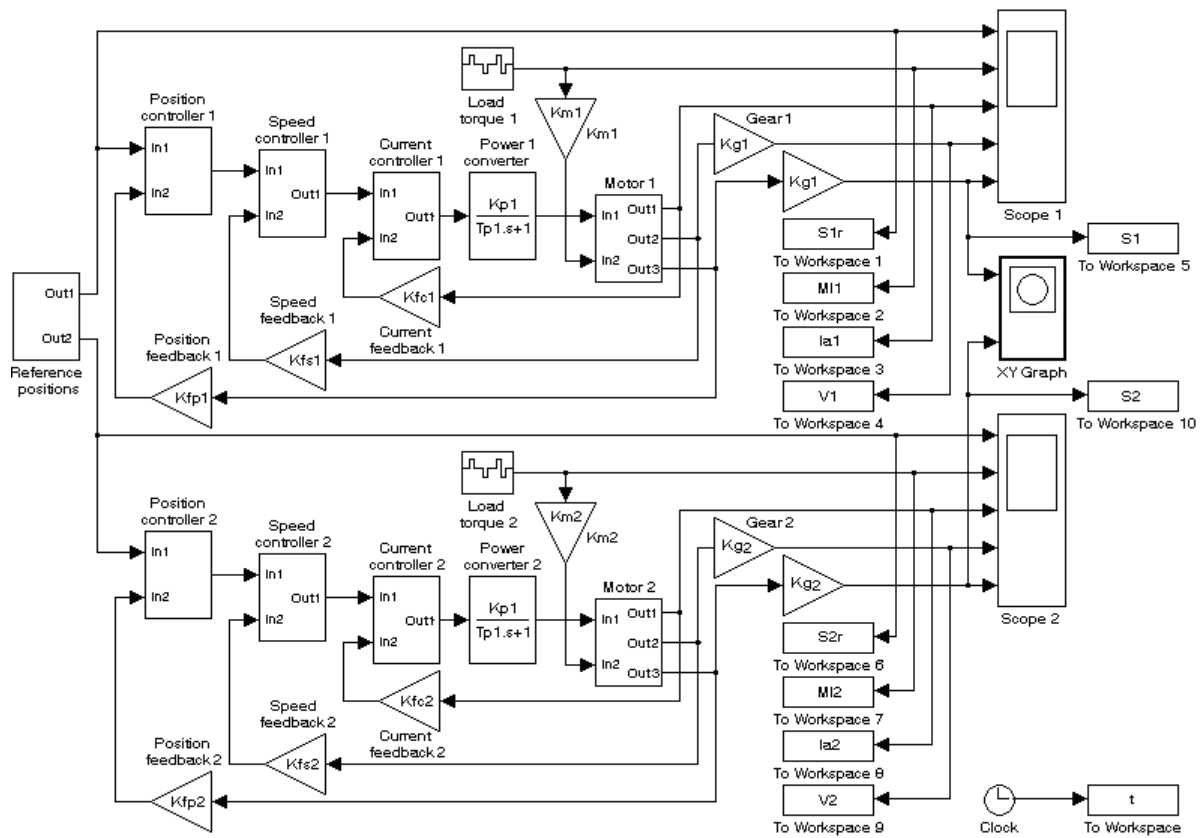


Fig. 3. Simulation model of the two-coordinate drive system.

In the MATLAB/SIMULINK environment some models of systems for two-coordinate electric drives have been developed with different types of motors. They allow for detailed studies of the respective static and dynamic regimes and analyses of performance. A simplified block diagram of one of the models is presented in Fig. 3.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

The electric drives for both coordinate axes are identical. The DC motors used for modeling and computer simulation have the following parameters:

$$V_{nom_i} = 30 \text{ V}; I_{anom_i} = 15.7 \text{ A}; \omega_{nom_i} = 115.19 \text{ rad/s}.$$

Fig. 4 shows the time-diagrams obtained by computer simulation during processing a position cycle along the *x* coordinate. The set displacement  $S_{1r}$ , the load torque  $T_{l1}$ , the armature current  $I_{a1}$ , the displacement speed  $V_1$  and the linear position  $S_1$  along this coordinate axis are shown. Armature current is limited to the maximum admissible value  $I_{a1max}$ . The respective control loops setting provide compensation of the disturbances and smooth positioning without overshooting. The disturbances applied sequentially on the electric drive are  $\Delta T_{l1} = \pm 0.25T_{l1nom}$

In Fig. 5 the trajectories of movement to the same final position have been shown, obtained through different algorithms of movement on both axes.

Fig. 5a represents a consecutive motion along the *x* and *y*

coordinates.

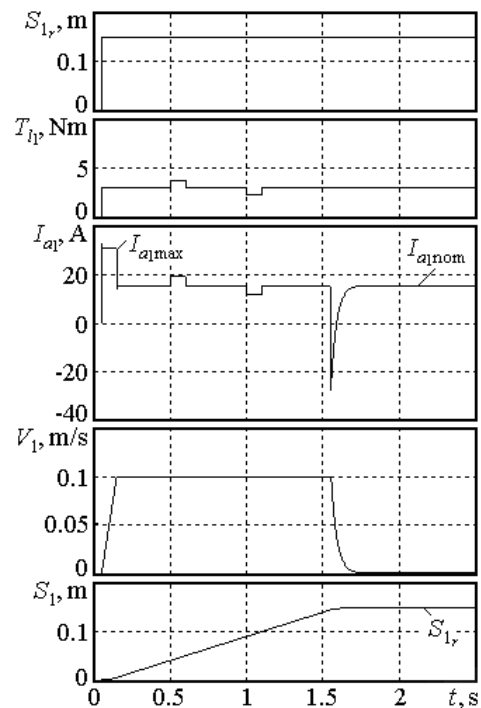


Fig. 4. Time-diagrams for a set position cycle along one axis.

Fig 5b shows the trajectory when simultaneous motion is performed along both coordinate axes.

Fig. 5b illustrates combined motion along the coordinate axes.

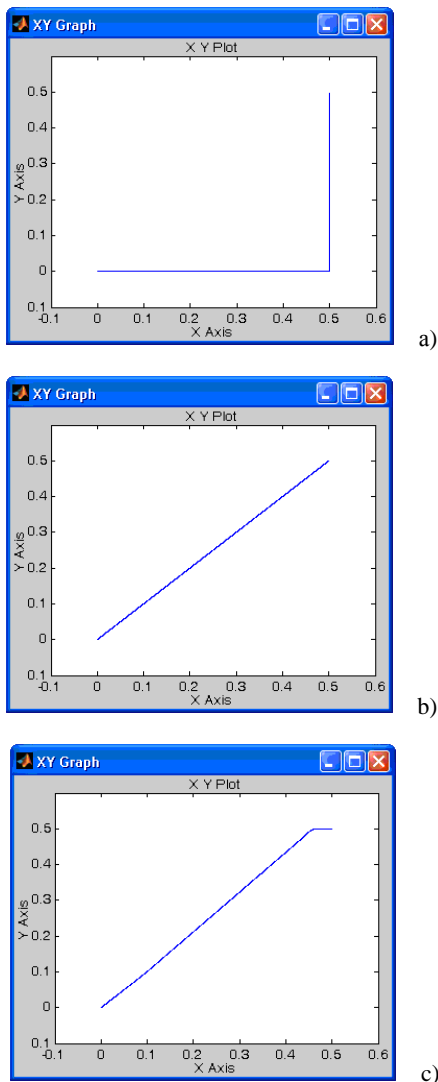


Fig. 5. Trajectories for different algorithms of control.

Detailed experimental studies have been carried out for different versions of controllers' tuning and operation regimes. Some time-diagrams are presented in Fig. 6, 7 and 8.

Fig. 6 shows a linear speed trajectory, obtained experimentally for displacements of 0.62 m along the coordinate axis x.

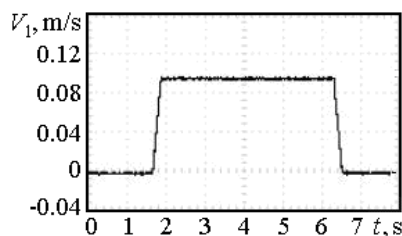


Fig. 6. Time-diagram of motion along the x coordinate.

A linear speed diagram, obtained for motion along the

coordinate axis y is represented in Fig. 7. The set distance in this case is 0.32 m.

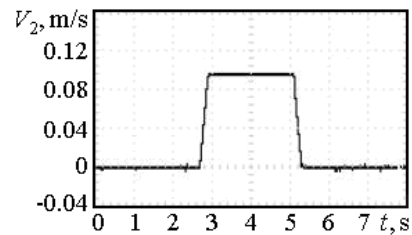


Fig. 7. Time-diagram for motion along the coordinate axis y.

Fig. 8 illustrates reverse control. The set distance in both directions is 0.32 m.

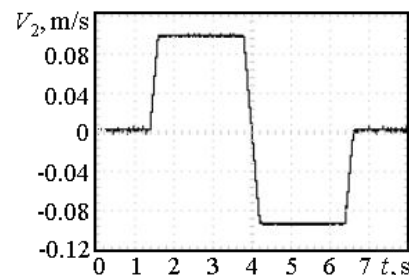


Fig. 8. Time-diagram illustrating reverse control along the axis y.

The behavior analysis shows that the presented position control algorithms provide for good performance suitable for practical applications.

### V. CONCLUSION

Models for computer simulation of two-coordinate electric drive systems with various algorithms for position control have been developed.

On the basis of computer simulations and experimental studies the presented algorithms for position control have been analyzed.

This research and the results obtained can be used in the design and tuning up of such two-coordinate systems of electric drives.

### REFERENCES

- [1] M. Mikhov, *Electric Drives Control Systems*, Technical University of Sofia, Sofia, 2009, ISBN 978-954-438-628-3.
- [2] M. Mikhov, I. Tatarov, Analysis of a Two-Coordinate Driving System Aiming at Performance Improvement. *Bulletin of the PGU*, Vol. 61, No. 3, pp. 105-110, Ploiesti, Romania, 2009, ISSN 1224-8495.
- [3] M. Mikhov, G. Mitrinski, Analysis of Some Algorithms for Positioning Control of Two-Coordinate Driving Systems. *Proceedings of the ICRDMI*, Vol. 1, pp. 638-643, Donji Milanovac, Serbia, 2010, ISBN 978-86-6075-017-6.
- [4] G. Blažiunas, Accuracy Investigations of Multifunctional Two-coordinate Drive System, *Electronics and Electrical Engineering*, Vol. 51, No. 2, pp. 8-14, 2004, ISSN 1392 – 1215.
- [5] C. Ong, *Dynamic simulation of electric machinery using MATLAB/SIMULINK*, New Jersey, Prentice Hall, 1998, ISBN 0-13-723785-5.

# Automated multichannel broadband spectrum analysis of fiber-optic grating sensors

Plamen Balzhiev<sup>1</sup>, Wojtek Bock<sup>2</sup>, Tinko Eftimov<sup>3</sup> and Rumen Arnaudov<sup>4</sup>

**Abstract** – In this paper we report methods for spectrum analysis and data processing algorithms. An automated multi-channel spectrum measurement system is introduced with controlled fiber-optic signal switching and spectra analysis with linear CCD photodiode array, diffraction grating and precise stepper motor. The designed system demands advanced measurement and data processing techniques. The paper reports the implemented methods for automated multi-channel measurements, accuracy improvement, noise cancellation techniques and fiber-optic grating sensor measurements

**Keywords** – long-period grating sensors, fiber-optic sensor interrogation and multi-channel spectrum measurement.

## I. INTRODUCTION

Over the past decade optical fiber-based sensors are gaining significant progress and popularity. Optical fiber gratings are often classified as Fiber Bragg gratings (FBGs) and Long Period gratings (LPGs) [5]. Because of the large periodicity, LPGs are usually easier to fabricate in mass production in comparison with FBGs. LPGs have found applications in various devices like equalizers for erbium doped fiber amplifiers, band-rejection filters and sensors for strain and temperature [3]. Because of their great width, spectral multiplexing is limited; the absence of a reflected signal demands detection of center wavelength shifts in a noisy minimum; and since resonance coupling in LPGs is to a cladding mode, the fiber typically has to be stripped, which creates challenges for long-term reliability and packaging.

On the other hand, LPGs are sensitive to a number of physical quantities such as surrounding refractive index, hydrostatic pressure, bending and twisting. They therefore offer significant application opportunities. However, in order to reduce the effective price per sensor, simple and efficient multiplexing systems must be developed. Wavelength and time-division multiplexing are well advanced with FBG sensor networks [1,2], but comparatively little has been reported on the multiplexing of LPGs [10]. Also a precise stepper motor is introduced to extend measured spectrum range and resolution by rotating the diffraction grating. Precise stepping drives are presented in [12, 13], where

<sup>1</sup>Plamen Balzhiev is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia, Bulgaria, E-mail: [baljiev@gmail.com](mailto:baljiev@gmail.com)

<sup>2</sup>Wojtek Bock is with the Canada Research Chair in Photonics, Université du Québec en Outaouais, Québec, Canada, E-mail: [Wojtek.Bock@uqo.ca](mailto:Wojtek.Bock@uqo.ca)

<sup>3</sup>Tinko Eftimov is with the Faculty of Physics, Plovdiv University, Bulgaria, E-mail: [tinkoeftimov@yahoo.ca](mailto:tinkoeftimov@yahoo.ca)

<sup>4</sup>Rumen Arnaudov is with the Faculty of Telecommunications, Technical University of Sofia, Bulgaria, E-mail: [ra@tu-sofia.bg](mailto:ra@tu-sofia.bg)

improvement in position accuracy and micro-stepping control is applied.

In this paper we report on the further development of a previously proposed spectrally and spatially multiplexed sensor network using an InGaAs CCD photodiode array and opto-mechanic switches. We also present results on the implemented methods for automated multi-channel measurements, accuracy improvement and noise cancellation techniques and fiber-optic grating sensor measurements..

## II. MULTI-CHANNEL SPECTRUM MEASUREMENT SYSTEM

### A. System description

The basic scheme of the multi-channel spectral measurement system is shown in Fig. 1. The radiation of a C+L band amplified spontaneous emission (ASE) broadband source (Joinwit) is coupled to port 1 of a 3-port optical circulator. Light reflected from the end of each channel is redirected from port 2 to port 3 and then to the diffraction-grating-based spectrometer and the CCD photodiode array detector [4].

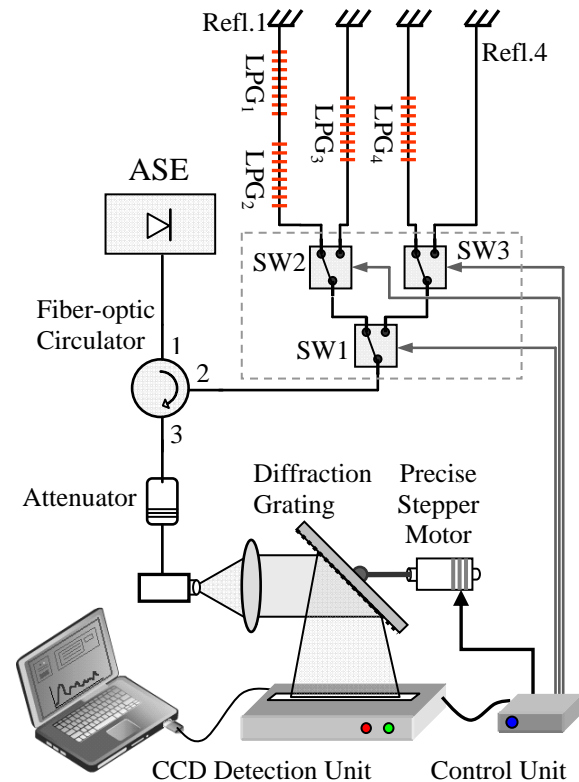


Fig. 1. Multi-channel spectrum measurement system

At port 2 there is an arrangement of three electrically controlled 1x2 fiber optic switches that allow an arbitrary access to four sensing channels, which can accommodate up to four LPGs depending on their bandwidth and sensitivity to a particular physical quantity [9, 14]. At the end of each channel there is a tunable reflector which returns light back to the sensing channel and through port 3 the light is collimated onto a 600 lines/mm diffraction grating so the spectrum is observed by a CCD photodiode array.

The four measured fiber-optic channels are set in the following configuration – in Ch.1 two LPG sensors are placed, Ch.2 and Ch.3 investigate single LPG and Ch.4 is utilized to perform reference signal measurement and system calibration with the ASE light source.

*B. Detection and control devices*

The detection unit is based on a 512-pixel InGaAs CCD (G9204-512D – Hamamatsu Photonics) linear array with an integrated low-noise charge-amplifier featuring high sensitivity, a low dark current and high stability in the 800-1750 nm spectral range. Two high-speed capacitive-based analog-digital converters (ADCs) transform the analog data from the CCD sensor into 16-bit corresponding digital values.

The obtained data from the CCD array is filtered and further transmitted via USB interface to a personal computer. The interrogation system is operated and configured by an application using a Lab-View programming environment [8] which allows an individual settings for the parameters - integration time ( $\tau$ ), sensor sensitivity ( $s$ ), conversion speed, data communication speed ( $r$ ), has start, stop and pause functions to be manually configured. Dark current can be subtracted after averaging, reference and the current signal can be read, and the signal-to-reference ratio will be presented in dB on the screen [11].

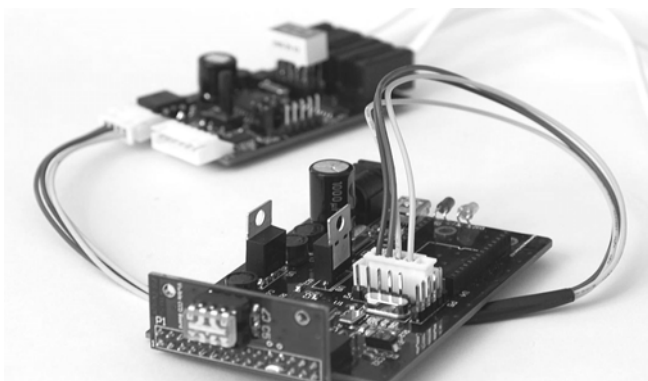


Fig. 2. Photographs of CCD Detection Unit and control device

The communication protocol between personal computer and devices is command-based – via the LabView application a command is transmitted to the CCD detection unit and it responds with corresponding packet of data or device status. If command is intended to adjust or acquire status for fiber-optic switches' position or stepper motor position then the CCD detection unit retransmits the command to the control device.

The LabView application automatically configures diffraction grating angle position via the precise stepper motor and the measured channel. After the current channel is measured and visualized, the program automatically configures the next channel for measurement.

*C. Correlation analysis for accuracy improvements and noise cancellation*

Correlation analysis in two separate spatially shifted signals is introduced to increase accuracy of spectral measurement and to reduce signal noises. The measured broadband signals from fiber-optic sensors are spectrally resolved on the linear CCD photodiode array. To perform a spatial shift of the signal a precise stepper motor is implemented. It rotates the diffraction grating with 0.1deg accuracy.

$$R(\tau) = \sum_{\tau} f_1(n)f_2(n + \tau) \tag{1}$$

$$R(\tau_{\max}) = \max$$

With cross-correlation function (1) the exact signal shift is calculated and any difference in signals is analyzed [6, 15]. In this way the multiple spatially shifted measurement of an identical signal may result in increased spectral measurement resolution.

On Fig.3a two measurements with spatial shift are presented. The cross-correlation function ( $R(\tau)$ ) and exact shift are calculated. The optimal signal match is achieved at maximum of  $R(\tau)$  and the exact shift ( $\tau$ ) is calculated.

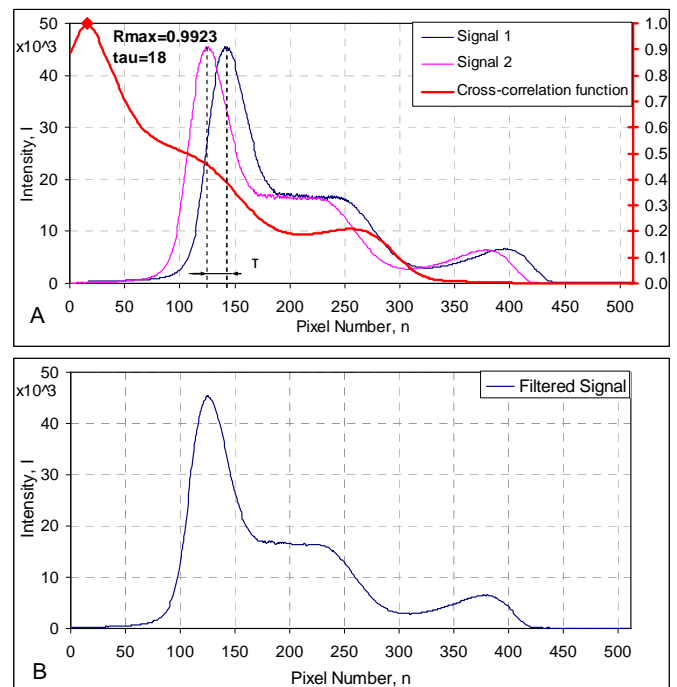


Fig.3. Correlation analysis of two shifted spectrum measurements-A, Noise suppression with spatially shifted measurements-B

To filter any noise resulting from signal conversion or in photodiode array and channel inequalities an averaging low-pass filter with respect to the spatial signal shift is designed (2). It averages N-shifted spectral measurements with the initial signal [7]. Since the correlation function is preliminarily calculated and the exact shift is acknowledged, the average signal is calculated and noises are filtered

$$S(n) = \frac{1}{N+1} \sum_{k=0}^N S_k(n - \tau_k) \quad (2)$$

$$\tau_{k=0} = 0$$

By introducing this filtering scheme the acquired spectrum preserves any narrow minima in measured grating sensors but also suppresses noises due to conversion or channel inequalities in CCD array. The resulting filtered signal is presented on Fig.3b.

### III. EXPERIMENTAL SET-UP AND MEASUREMENT RESULTS

The designed multi-channel broadband spectrum measurement system is tested by analysing four different long period grating sensors arranged in three channels. On Ch.4 only a tuneable reflector was connected and this channel was utilized as reference signal. A joint multiple channel graphic is presented on Fig.4 with relative measurement to the reference signal.

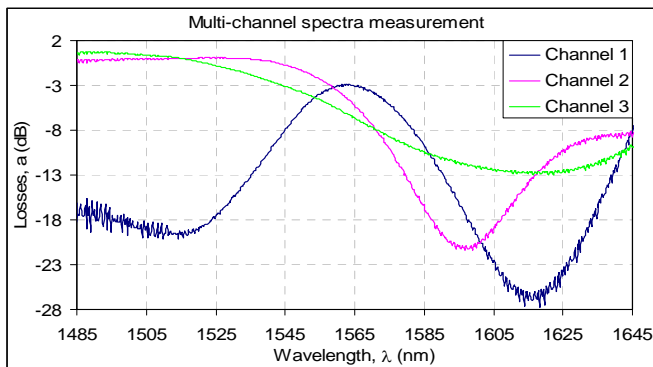


Fig. 4. Joint multi-channel spectral response measurement

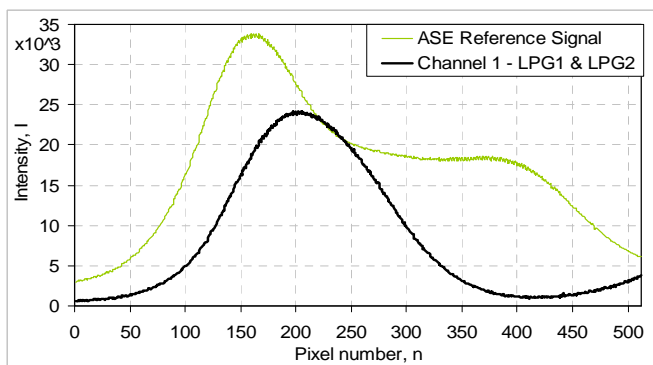


Fig. 5. Channel 1 LPGs' signal compared to reference spectrum

The results from measurements of the transmission spectra in Channel 1 are shown in Fig.5. Fig.6 presents linear change of spectrum change with two LPG sensors compared to the reference ASE broadband light source. The same results are shown for Ch.2 and Ch.3 respectively in Fig.7-8 and Fig.9-10. The spectrum change in Ch.1 when sensors are under stress (bending) is presented on Fig.6, where two LPGs are simultaneously measured.

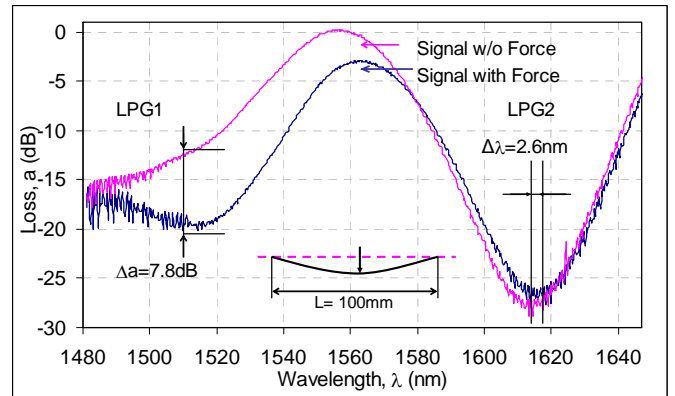


Fig. 6. Channel 1 relative measurement with LPG sensor under stress

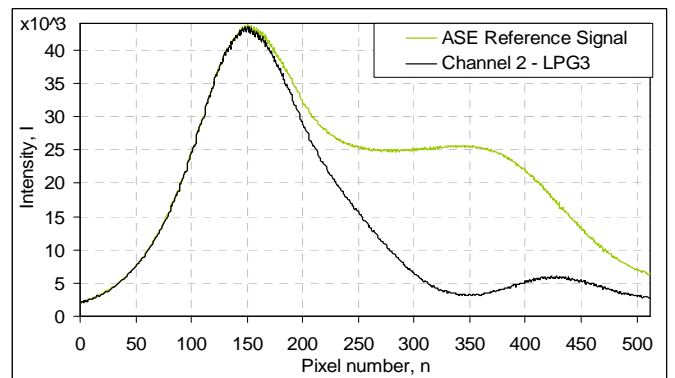


Fig. 7. Channel 2 LPGs' signal compared to reference spectrum

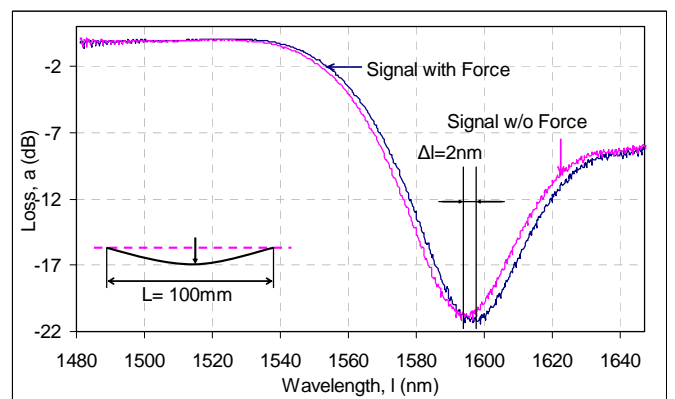


Fig. 8. Channel 2 relative measurement with LPG sensor under stress

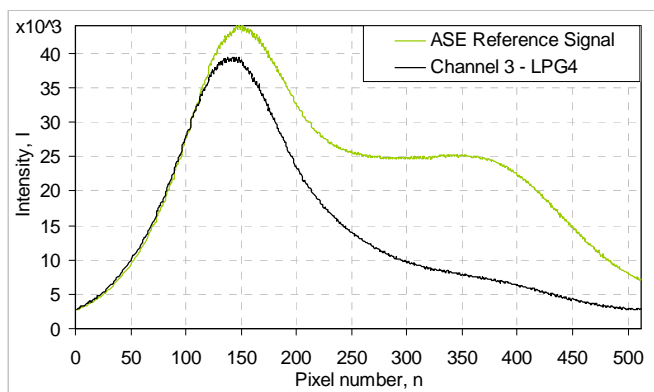


Fig. 9. Channel 3 LPGs' signal compared to reference spectrum

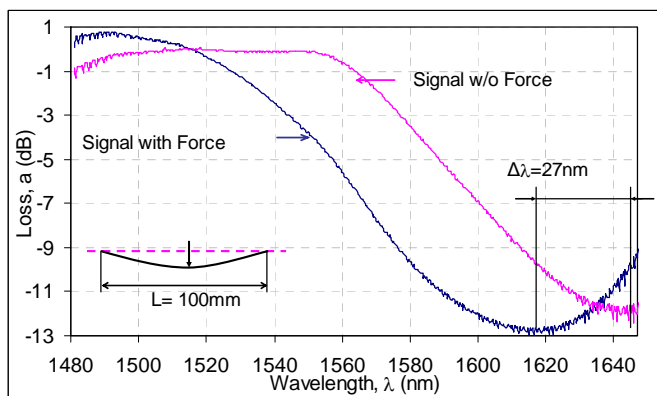


Fig. 10. Channel 3 relative measurements with LPG sensor under stress

They are particularly selected not to interfere with each others' spectra.  $LPG_1$  shows significant depth change in its transmission minimum at  $\lambda=1515\text{nm}$ , while  $LPG_2$  results in minimum shift of  $\Delta\lambda=2.6\text{nm}$ .

Corresponding transmission minimum shifts for Ch.2 and Ch.3 are presented on Fig.8 and Fig.10. The latter sensor is particularly sensitive resulting in a larger minimum shift of  $\Delta\lambda=27\text{nm}$ .

#### IV. CONCLUSION

The reported automated multi-channel broadband analysis system for measurement of fiber-optic grating sensors is capable of simultaneously monitoring up to 12 long period gratings in groups of three spectrally multiplexed sensors per channel. An advanced signal processing and analysis is demonstrated using cross-correlation function and adaptive filtering techniques which effectively suppress noises while preserving the spectral resolution.

It is also possible to extend significantly the measured spectrum range by rotating the diffraction grating with the precise stepper motor and the wide sensitivity range of the linear CCD photodiode array (800-1750 nm).

The measurement results demonstrate the high sensitivity of implemented long period grating sensors to external strain

forces on the fibers and also the ability of the demonstrated system to detect and analyze those spectral changes.

#### ACKNOWLEDGEMENT

The authors acknowledge the support if the Ministry of Education of Bulgaria under research project VU EES 303/07, the Natural Sciences and Engineering Research Council of Canada, the Canada Research Chairs Program, and NanoQuébec.

#### REFERENCES

- [1] Murphy, D. F., Flavin, D. A., McBride, R., and Jones, J. D. C., Interferometric Interrogation of In-Fiber Bragg Grating Sensors Without Mechanical Path Length Scanning, *J. Lightwave Techn.*, 19, 1004-1009 (2001).
- [2] Alves, J., Santos, J.L., Carvalho, A., and Lage, A. Fiber Bragg Sensor Interrogation System Based on a CCD Spectrometer, *Proceedings of IEEE, Sensors*, 1, 909-913 (2003)..
- [3] J. Choi, Adaptive and Iterative Signal Processing in Communications, Cambridge Press, 2006.
- [4] C.G. Askins, M.A. Putnam E.J. Friebele, Instrumentation for Interrogating Many-element Fiber Bragg Grating Arrays, *SPIE vol.2444 p.257-266*, 1995.
- [5] K. Srimannarayana1, M. Shankar,R.S. Prasa, Fiber Bragg Grating And Long Period Grating Sensor For Simultaneous Measurement And Discrimination Of Strain And Temperature Effects, *Optica Applicata*, Vol. 38,No. 3, 2008
- [6] S. Vaseghi, Advanced Digital Signal Processing and Noise Reduction, 4th Edition, Wiley, 2009.
- [7] Farhang-Boroujeny, B., Adaptive Filters – Theory and Applications, Wiley, 1998.
- [8] Bitter, R., Mohiuddin, T., Nawrocki, M., LabVIEW – Advanced Programming Techniques, 2<sup>nd</sup> Edition, CRC Press, 2007.
- [9] Zahariev P., Hristov G., Iliev M., Tsvetkova I., "Extending the lifetime of wireless sensor networks by mechanisms for routing the information in more energy efficient ways", *Mosharaka International Conference on Communications, Computers and Applications MIC-CCA*, Istanbul, Turkey, pp. 73 – 78
- [10] Guan, Z.G., Zhang, A.P., Liao, R., "Wavelength Detection of Coherence-Multiplexed Fiber-Optic Sensors Based on Long-Period Grating Pairs", *IEEE Sensors Journ.*, 7, 36-37, 2007.
- [11] Tamhane, A. C., "Statistical Analysis of Designed Experiments: Theory and Applications", Wiley, 2009.
- [12] M. Mikhov, P. Nakov,"A Hybrid Step Motor Drive System with High Positioning Accuracy", *Proceedings of the International Conference on Electronic Devices and Systems*, pp. 59-64, Brno, Czech Republic, 2005, ISBN 80-214-2990-9.
- [13] M. Mikhov, P. Nakov, "Stepping Motor Drive for Precise Positioning Applications", *Proceedings of the International Scientific Conference on Information, Communication and Energy Systems and Technologies*, Vol. 1, pp. 227-230, Nish, Serbia, 2008, ISBN 978-86-85195-59-4.
- [14] Zahariev P., Hristov G., "Performance evaluation of data delivery approaches for wireless sensor networks", *World Conference on Information Technology WCIT 2010*, Istanbul, Turkey; *Procedia Computer Science*, Volume 3, 2011, Elsevier, pp. 714 – 720
- [15] Tsvetkova I., Aleksandrov Y., Hristov G., Zahariev P., Iliev M., "Comparison of target tracking algorithms in hierarchical WSNs", *World Conference on Information Technology WCIT 2011*, Antalya, Turkey



# Measurement of the Position by Using Hybrid Pseudorandom Encoder

Dragan Denić<sup>1</sup>, Goran Miljković<sup>2</sup>, Jelena Lukić<sup>3</sup>, Miodrag Arsić<sup>4</sup> and Milan Simić<sup>5</sup>

**Abstract** – This paper proposes the position measuring method conducted with a hybrid pseudorandom encoder designed by combining the pseudorandom and the incremental code tracks. This combination of code tracks increases the measurement resolution. The proposed hybrid pseudorandom encoder contains the detector of the pseudorandom code reading errors, which helps to reduce the frequency of the autocalibration and eliminates the possibility for the coarse error. The paper also describes the testing of the proposed encoder electronic block in the presence of incremental and code track contaminations.

**Keywords** – position measurement, hybrid pseudorandom encoder, automatically guided vehicles (AGV)

## I. INTRODUCTION

The high accuracy measurement of the lengths and the angles is often requested in modern industrial production processes. The accuracy of 0.1  $\mu\text{m}$  is no rare and it is more often requested. Such a request is a result of the high level of production automation and strict requirements regarding the positioning of the movable parts of the manufacturing units. At the same time, the requirements for the position measuring of the automatically guided vehicles (AGV), which move independently on the factory floor as a production platforms or transport units, are stricter. These vehicles have the increasing tendency towards the flexible movable systems which represent the combination of vehicles and robots [1]. These automatically guided vehicles can often be in the interaction with the other production units, such as robots, machine tools, charging and discharging units and etc. Therefore, the precise and highly reliable information about the position is necessary. For this reason the utilization of the classical absolute measuring systems is unacceptable for economic reasons, because of the large number of code tracks. Today, most of the systems rely on the incremental methods for the position determination, which in turn have the known

drawback of error accumulation. There were attempts to compensate these drawbacks by using the optical calibration methods or special code marking of the specified locations [2]. Unfortunately, the usage of these solutions is limited with the number of reference points, i.e. for economic reasons.

The result of the latest research in the field of digital position transducers for AGVs is the pseudorandom coding method, which, for the absolute position determination, requires only one bit per quantization step, [3], [4]. The longitudinal coding method was developed with the help of an useful pseudorandom code property that the first  $n-1$  digits of a pseudorandom code word are identical with the last  $n-1$  digits of a previous pseudorandom code word. Therefore, unlike transverse coding technique, which requires that for each coding device sector a digital code is written in the transverse direction, this technique allows the absolute position measurement with the use of just one code track. It is based on the „window property“ [3] of the pseudorandom binary sequences (PRBS)  $\{S(p)/p=0,1,\dots,2^n-2\}$ . Accordingly, any  $n$ -bit long code word  $\{S(p+n-k)/k=n,\dots,1\}$  provided by the scanning of the PRBS with the window of width  $n$ ,  $\{x(k)/k=n,\dots,1\}$ , is unique and may fully identify the absolute position of the window  $p$  with regard to the sequence start.

In the case when the high resolution of the measurement is required, the PRBS of the greater length need to be applied because the operating range of the AGV can be 300-400 m. Therefore, due to the code conversion problem [4], for a given maximum velocity of the AGV, the overall measurement resolution is limited. Even if there is no such a problem, the code reading uncertainty limits the overall measurement resolution of the pseudorandom position encoder. However, the convenience of the pseudorandom coding application can be utilized for the case when the high precision of positioning is required. The basic idea, which will be presented in this paper, was derived from the fact that the pseudorandom coding is very suitable for use with incremental position measurement methods. PRBS-s can be used for encoding of the reference markers used for the measuring system autocalibration. This approach leads to the realization of the hybrid measuring system. The aim of this paper is to emphasize the great advantage of the hybrid method for the measuring of the AGV position.

From the measurement and the position determination points of view, the hybrid measuring system uses a measurement method that combines the absolute and the incremental methods. The hybrid measuring method includes the functional elements of both methods. The intention is to retain the good qualities of these methods, and to eliminate their disadvantages. In 1983, Whitwell proposed [5] that the track with conventional markers (measurement grid) needs to be placed along the several code tracks of some absolute scale. Using  $n$  code tracks for each sector of the width  $d$  the

<sup>1</sup>Dragan Denić is with the Faculty of Electronic Engineering at University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: dragan.denic@elfak.ni.ac.rs

<sup>2</sup>Goran Miljković is with the Faculty of Electronic Engineering at University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: goran.miljkovic@elfak.ni.ac.rs

<sup>3</sup>Jelena Lukić is with the Faculty of Electronic Engineering at University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: jelenalukicpk@gmail.com

<sup>4</sup>Miodrag Arsić is with the Faculty of Electronic Engineering at University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: miodrag.arsic@elfak.ni.ac.rs

<sup>5</sup>Milan Simić is with the Faculty of Electronic Engineering at University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: milan.simic@elfak.ni.ac.rs

digital code is written. At the same time, the conventional incremental encoder with minimum two detectors is applied. Two signals phase shifted for  $\pi/2$  are needed for the AGV motion direction determination.

Whenever the determination of the absolute position is carried out (by direct reading of the digital code in the transverse direction) the accumulated encoder errors are automatically eliminated. Further incremental position determining is performed with respect to the absolute value of the determined position.

Any error in the absolute value of the determined position would represent a systematic error which can be eliminated only during the next absolute position determination process. However, this systematic error would be enormously large in relation to the accuracy and resolution of the applied incremental encoder. Except this drawback, the disadvantage of a large number of code tracks also remains. For the case of movable systems with a wide range of movement, even at relatively low resolution of the absolute position determination the number of code tracks is so large, that the proposed hybrid system becomes economically unacceptable for the implementation.

## II. THE HYBRID PSEUDORANDOM ENCODER

The main drawback of the previously described "classical" hybrid position encoder is that it has a high number of code tracks applied for writing the digital code in the transverse direction. This drawback can be eliminated with pseudorandom position encoder employment for determination of the AGV absolute position. The benefit of combining pseudorandom and incremental coding methods was indicated in [3]. This is done in the terms of resolution increase by introducing a greater number of sensor heads arranged according to the "vernier" method. This opened the possibility for measuring system functioning according to the incremental method. However, in this measuring system are not included the main advantages of the incremental method, such as: simplicity, small number of connecting wires and the high density of measurement range partitioning. As noted above, the hybrid encoder represents the direct combination of the absolute and the incremental encoder [6]. The solution proposed in this paper is shown in the Fig. 1. Along the synchronization and the code track the measurement grid with the conventional markers is added. The whole electronic block, required for obtaining the information on the relative position of the AGV in binary code, is included. It may be noted that the pseudorandom position encoder is simply extended by adding a complete two-phase incremental encoder. The number of pulses is quadrupled by combining A and B signals using the EXOR circuit and with edge detection of thus obtained signal. To determine the direction of movement one of the solutions described in [7] can be applied. Otherwise, depending on the detection technique and specific detectors applied, the distance between  $y(1)$  and  $y(2)$  can be  $r=(e+1/4)d/g=(e+1/4)q$ , where  $e$  can take an arbitrary value from the set  $\{0,1,2,3,\dots\}$ , and  $g$  is the number of markers per one sector width  $d$ . In general, if the resolution increase for  $m$  bits is needed, the requirement for  $g=2^m$  should

be fulfilled. For the case when the number of pulses is quadrupled, as it is proposed here, the number of markers (or reflecting areas, in the case when an optical reflection method is applied) in one sector, for the absolute position determination, is  $g=2^{m-2}$ . The  $(n+m)$ -bit UP/DOWN counter is applied and thus the reached output resolution is  $(n+m)$ -bits. The code reading uncertainty problem is solved by using the external synchronization method. Along the code track an additional synchronization track is placed. The code reading is done at the moment when the sensor head AUT detects the transition between two adjacent sectors, [3], [4].

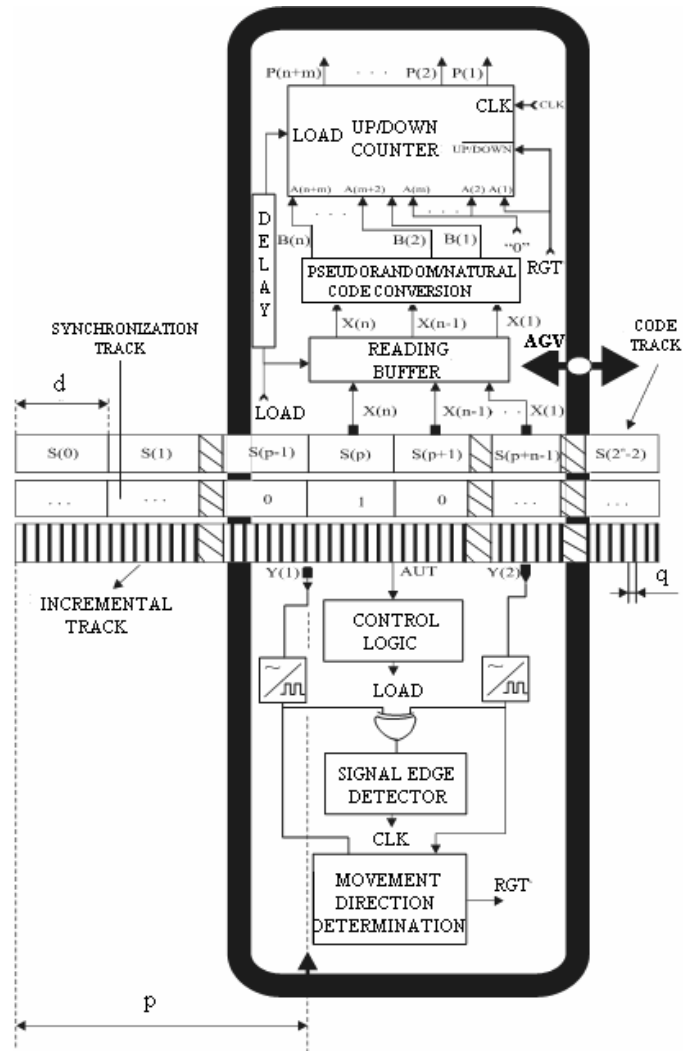


Fig. 1. The hybrid pseudorandom position encoder

Otherwise, the hybrid position encoder from Fig. 1., does not need the additional VER head, which is used at pseudorandom encoders for determining of the movable system movement direction, [3]. The AGV movement direction is now determined with the applied incremental encoder. The marker, in relation to which the position is determined, is placed opposite the detector  $y(1)$ . The  $n$ -bit adder is no more needed as the correction element [3]. The difference in the position of the AGV on the same transition, depending on the AGV movement direction, is now  $q/4$ . This

correction for the least significant bit is automatically performed during the measuring system autocalibration.

The proposed measuring system operates according to the incremental method. At the moment when the synchronization head AUT detects the transition between two sectors, of width  $d$ , the pseudorandom code reading is done. In this case, the control logic for LOAD signal obtaining is very simple and can be realized by using the Schmitt circuit (or the comparator) and edge detector. Converting of the read code into the natural code gives the absolute position of the AGV. At the same time a pulse is obtained at the output of the delay circuit and the absolute position is loaded as a new state of the counter. These  $n$  bits of the AGV current position are respectively entered in the first  $n$  inputs of the UP/DOWN counter starting from the input of the greatest weight. In the next  $m-1$  counter stages the logic zeros are entered. At the least significant input logic one or logic zero is fed depending on the AGV movement direction, Fig. 1. In this way the measuring system autocalibration is performed. It continues to work according to the incremental method and in relation to the new counter state.

This increase of this solution complexity in relation to the pseudorandom position encoder complexity is insignificant comparing to the quality obtained. The proposed hybrid position encoder is a cost-effective solution even for movable systems with the highest requirements. It can be applied in systems with the wide range of movement, which require high precision and high resolution of the position measuring because of the mutual influence with the other production units.

However, the act of autocalibration represents a critical moment in the functioning of the hybrid position encoder. That is the reason why the hybrid position encoder has not found its application in practice until today. The word "calibration" implies the existence of a reference against which the correction is made. The question is whether the measured absolute position satisfies the criteria of one reference. At first instant it does. The code is written on a physical track and it can be read only at precisely determined locations. Such information can be accepted as a reference, but it is necessary to ensure the conditions that will guarantee that no errors will happen during the code reading procedure. Unfortunately, no one can claim that in real industrial conditions such terms will be provided. Approaching to these ideal conditions can be very expensive, but it does not solve the problem. Therefore, in this paper we started with a new approach. The additional information is introduced for indicating the presence of an error, with the probability of one, in the pseudorandom code reading. For the realization of the hybrid position encoder the detector of the pseudorandom code reading errors is required. It confirms that the obtained pseudorandom code word corresponds to a certain physical reference, with which the current AGV position coincides. Only in this case the measuring system autocalibration is done. Otherwise, if such confirmation does not exist, the obtained pseudorandom code word will not be accepted as the reference and the measuring system autocalibration will not occur. In this way the autocalibration frequency becomes lower, but the possibility of the course error is eliminated.

### III. EXPERIMENTAL RESULTS

The development of an experimental system can be of great importance for the research of various solutions of position encoders. The realization of different encoder solutions requires the code tracks manufacturing, which can be expensive and impractical because the obtained solutions need the application of different code reading methods. Therefore, the simulation of the system movement and the testing of the proposed encoder electronic block should be performed with the computer and the corresponding hardware application.

In order to test the electronic block of the position encoder, an experimental system consisted of the computer, the parallel I/O card and the microprocessor development system based on the Intel 8031, is built (Fig. 2.). So, in this experimental system configuration the electronic block of the pseudorandom position encoder is realized by using the microprocessor development system based on the Intel 8031, while the computer is used for the simulation of the movable system movement. By starting the program for the simulation of the movable system movement at the corresponding parallel outputs of the I/O card the signals, equal to those which would be obtained at the output of sensor heads for the given movement route, are generated. These signals are fed to the port 1 of the Intel 8031 microprocessor. After the testing is finished, by using the tabular or graphical presentations the given series of positions and the obtained series of positions can be compared. The tabular view provides the continuous information about the measuring system working regime. Since all the real situations can be simulated, including damaged or contaminated measuring tracks, there is no reason to expect a different behavior of the measuring system in its concrete application. In other words, the test results fully and realistically characterize the proposed position encoder.

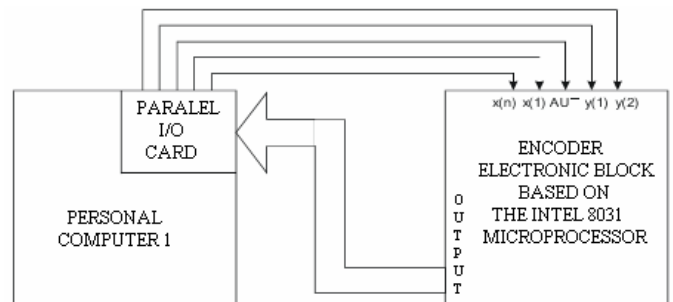


Fig. 2. Testing of the position encoder electronic block

The electronic block has been tested during a period of several months and for over a thousand randomly selected AGV movement routes. After the measuring system entered into the normal working regime, for the ideal measuring tracks located in an ideal environment case, there is no typical AGV movement which would lead to an error in the position determination or which would take the measuring system out of its normal working regime.

The simulation of incremental and code track contamination can be done with the program that has been already developed for the simulation of the movable system

movement. The contamination level is specified in percentage. As a consequence of the code track contamination the frequency of the measuring system autocalibration decreases. Fig. 3 gives the graphical representation of the 16-bit hybrid encoder test results when the incremental track contamination level is 9.4% and an additional code track contamination level is 3%. Obviously, the position measurement error considerably increases with wrong pseudorandom code bit readings. In the best case, the maximal measurement error is about ten times higher. The error is even higher when the code reading synchronization is lost, which often happens. As it is noted above, instead of 64 quantization steps, the next measuring system autocalibration, in the worst case, does not occur until 1280 quantization steps are ran over in the same AGV movement direction. Specifically, in the case of the performed testing, Fig. 3, there was the code reading synchronization loss. The next measuring system autocalibration was performed after 16 absolute quantization steps  $d$  of the AGV movement. As it can be seen from the error graphic, i.e. from the graphic of real and determined position difference, the previous condition caused the maximal error of one hundred increments. Normally, at the moment when the error is detected, the measuring system provides the information that it switched to the incremental working mode. The measuring system continues to function as a classical two-phase incremental position encoder until the next autocalibration of the measuring system is performed. It should be mentioned again that the adopted contamination level of the incremental track is much higher than it is usually reported in the practice, and that the considered case represents the worst possible case. It is important to say that during the long testing period of the encoder electronic block, each simulated code reading error was detected.

#### IV. CONCLUSION

The proposed hybrid pseudorandom position encoder solution represents a new approach in the field of position measurements. The hybrid pseudorandom encoder, which combines the good properties of the incremental and the absolute position measuring methods, is applicable in practice due to the presence of the code reading error detector. For the first time, the reliability of the obtained position information is considered and the reliable detection of the possible errors is provided. The results obtained so far indicate the fully functionality of the proposed encoder. The software realization of the encoder electronic block is also possible by using the microprocessor. The realized encoder electronic block has been tested by using an experimental system developed for the AGV movement simulation, even in the presence of damage or contamination of the measuring tracks. After testing it was shown that the maximal error is one hundred increments when the incremental track contamination level is 9.4%, when the code track contamination level is 3% and in the presence of the code reading synchronization loss.

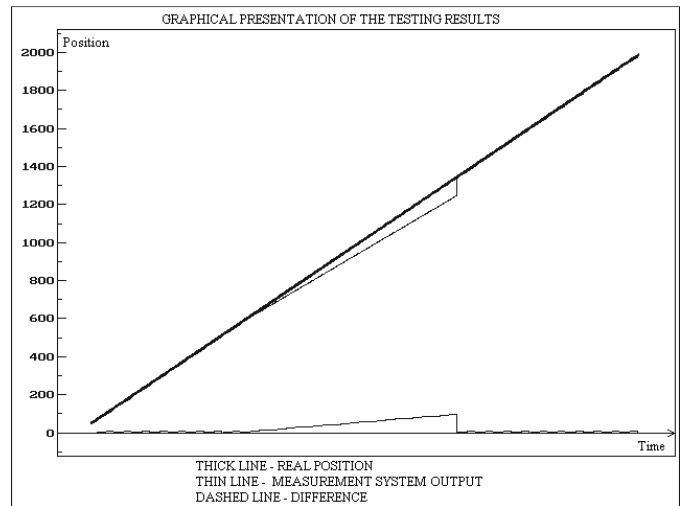


Fig. 3. Testing results of the encoder electronic block when the incremental track contamination level is 9.4% and the code track contamination level is 3%

Of course, it should be noted that the adopted incremental track contamination level during testing was much higher than it has been reported in practice.

#### ACKNOWLEDGEMENT

Research activities presented in this paper, are supported by funds of the Ministry of Education and Science of the Republic of Serbia, having the reference project number TR32045.

#### REFERENCES

- [1] J. Zygmunt, "Guided Vehicles Set Manufacturing in Motion", High Technology, pp. 16-21, December, 1986.
- [2] T. Hongo, "An Automatic Guidance System of a Self-Controlled Vehicles", IEEE Trans., Ind. Electron., vol. IE-34, pp. 5-10, 1987.
- [3] E. M. Petriu, J. S. Basran, "On the Position Measurement of Automated Guided Vehicles Using Pseudorandom Encoding", IEEE Trans., Instrum. and Meas., vol. IM-38, no. 3, pp. 799-803, 1987.
- [4] D. Denić, M. Arsić, D. Živanović and M. Pešić, "Elektronski sistemi za pozicioniranje automatski upravljanih vozila", Zbornik radova SAUM, Kragujevac, Srbija, pp. 503-514, Jun, 1992.
- [5] A. L. Whitwell, "Moire Techniques Ensure Unerring Positional Control", Design Engng., pp. 45-48, November, 1973.
- [6] D. Denić, M. Arsić, "Precizno pozicioniranje automatski upravljanih vozila primenom metoda hibridnog merenja pozicije", Zbornik radova SAUM, Novi Sad, Srbija, pp. 341-345, Oktobar, 1995.
- [7] P. Thajchayapong, M. Chinnakarn, "Further Improvement in a Counting and Direction-Sensing Circuit", International journal of electronics, vol. 66, no. 6, pp. 935-938, 1989.

# Method for calculating the stability at moderate and big heeling angles of a vessel

Mariya Eremieva<sup>1</sup>, Viktoriya Sabeva<sup>2</sup>, Mariya Nikolova<sup>3</sup>, Emiliya Koleva<sup>4</sup>

**Abstract** – Calculating the stability of a vessel at moderate and big heeling angles by building Static Stability Curves. The method introduces the notation Dynamic Height of the Mass Centre (DHMC).

**Keywords** – Static Stability Curves, Dynamic Height of the Mass Centre, Metacentric Height

## I. INTRODUCTION

“Stability” is the ability of the vessel to return to its initial position before capsizing after the inclining forces have ceased to operate. Hence the stability issue is an issue of the ship behavior in transverse and longitudinal angle inclination caused by external effects. The change of the stability is searched by the change of the metacentric height and the corresponding coefficients. The existing dependences for evaluating the stability are based on the criterion of minimal starting metacentric altitude and guarantee basically the starting stability. This report presents a new method for evaluating the stability using “dynamic metacentric height” (DMCH).

## II. METHOD FOR EVALUATING THE DMCH

When sailing each vessel is exposed to external forces like wind, heavy sea, surge and etc. Its safety depends on the seaworthiness quality - stability. In basic aspects this means that the ship must counteract the negative force effects. That is to say not to heel to dangerous angles and to redress its initial balance after the external influences are over. According to the magnitude of the heeling angle  $\theta$  (transverse inclination of the vessel) the stability is classified in two types: initial stability and stability at moderate and big angles of list (heel). In both cases the counteraction is due to the couple forces: buoyancy (Archimedes) force and the weight  $P$ . They form a moment  $M_r$  that is the bases of stability and is opposite to the heeling moment  $M_l$ . At small angles  $\theta$  is defined the notation “initial metacentric height”  $\overline{GM}$ . This altitude is enough to evaluate the initial stability in details. (Figure 1)

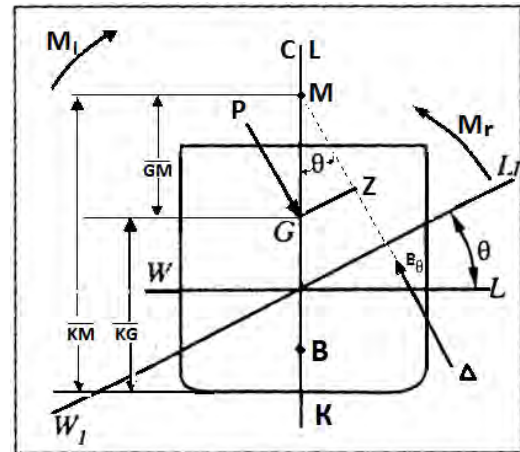


Fig.1. Stability at external effects

Furthermore, some basic assumptions are usually taken into consideration, one of which is the principle of the geometric inverse which states that the vessel is always considered upright and the waterline  $WL$  is inclining at a certain angle ( $\theta$ ). Assume that the point  $M$  is the cross point of the directrix of two infinitely closed buoyancy forces. At moderate and big angles of the heel, after applying the principle of geometric inverse, the point  $M$  doesn't lie down in the diametrical plane  $CL$  of the ship. This assumption makes pointless the using of  $\overline{GM}$  and it's necessary to investigate the change of the arm  $\overline{GZ}$  (the arm of the righting moment) or the righting moment itself ( $M_r$ ). Hence

$$M_r = \Delta \overline{GZ} \quad (1)$$

could be represented as a function of the heeling angle  $\theta$

$$\overline{GZ} = f_1(\theta) \text{ or } M_r = f_2(\theta) \quad (2)$$

denoted as Static Stability Curve (SSC). (Figure 2)

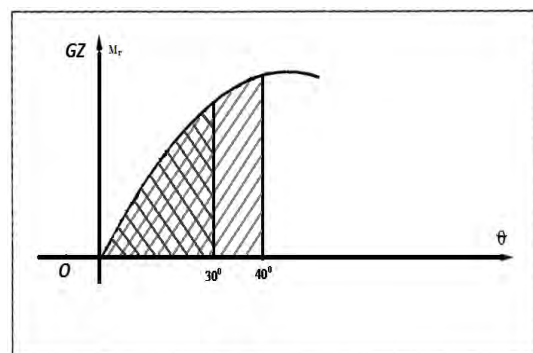


Fig.2 – Dynamic areas in Static Stability Curve

<sup>1</sup>Mariya Eremieva, <sup>2</sup>Viktoriya Sabeva, <sup>3</sup>Mariya Nikolova, <sup>4</sup>Emiliya Koleva are at the Engineering Faculty at Naval Academy “Nikola Yonkov Vaptsarov”, 73 Vasil Drumev str, Varna 9026, Bulgaria,

e-mails respectively:

[eremievam@abv.bg](mailto:eremievam@abv.bg), [viktoriya\\_sabeva@yahoo.com](mailto:viktoriya_sabeva@yahoo.com),  
[mpn@abv.bg](mailto:mpn@abv.bg), [emiliya\\_f@abv.bg](mailto:emiliya_f@abv.bg).

The SSC examination, used to evaluate the stability, is related to the way the external effects are enforced – static or dynamic [1]. In practice dominates the dynamic enforce. That's why the given evaluation is expressed in determining the work of the righting moment i.e. the magnitude of the area under SSC to certain heeling angles. The requirement of the International Maritime Organization (IMO) concerning the minimal values of the areas and angle ranges are lied down in resolution A749(18) as a guarantee for certain aspect of stability. They are the following [3]:

- 1) Area of diagram  $S_1$  for  $\theta = 0^0$  up to  $\theta = 30^0$ ,  $S_1 \geq 0.055$  (but not smaller than 0.055)m.rad
- 2) Area  $S_2 \geq 0.09m.rad$  up to the utmost angle  $\chi$ , which is interpreted as the smallest of the three variants:
  - a) flooding angle  $\theta_f$ ;
  - b) heeling angle corresponding to  $\overline{GZ} \theta_m$ ;
  - c) angle equal to  $40^0$ .
- 3) Area  $S_3$  between  $\theta = 30^0$  and  $\theta = \chi$ ,  $S_3 \geq 0.03m.rad$

The direct building of the diagrams under the terms of IMO is too hard for the command staff because of the many calculations. Therefore the suggestion is to introduce in ship's papers a diagram of the dynamic height of the mass centre (DHMC) of the vessel that meets the required norms of IMO and represents the upper limit for placing the mass centre in height. That way in designing the cargo plan is estimated whether the obtained mass centre satisfies the requirements for dynamic stability.

In creating the shown diagram (fig.2) are examined all the SSC in the range from "empty" to "full" ship with given realistic height tolerance of the mass centre. It means that the arm of the righting moment  $\overline{GZ}$  is considered as a function of the displacement  $\nabla_i$ , the altitude of the mass centre  $\overline{KG}_j$  and the accepted angle interval  $\theta_n$ . (Fig. 3) Since the indicated argument define the arm of the form

$$\overline{KN}_{i,n} = f(\nabla_i, \theta_n) \tag{3}$$

and the arm of the height

$$(l_G)_{j,n} = \overline{KG}_j * \sin \theta_n, \tag{4}$$

hence 
$$(\overline{GZ})_{i,j,n} = \overline{KN}_{i,n} - (\overline{KG})_j * \sin \theta_n. \tag{5}$$

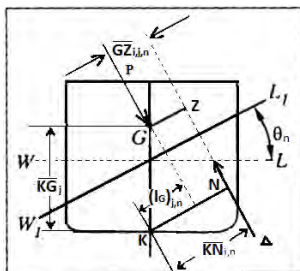


Fig.3 – Arm of the form and arm of the height

For the arm of the form  $\overline{KN}_{i,n}$  are used the  $KN$  - curves for the specific ship for  $\theta = 10^0$ ,  $\theta = 20^0$ ,  $\theta = 30^0$  and often  $\theta = 40^0$ . (Fig. 3) Next step is to calculate the areas  $S_{0^0-30^0}$ ,  $S_{0^0-40^0}$  and  $S_{30^0-40^0}$  (Fig. 2) for each SSC and compare them with the ones required in IMO. If the three areas are the same as the required or one of them is the same and the others are bigger, than the value  $\overline{KG}_{max}$  is maximum acceptable altitude for positioning the mass centre according to this displacement. The graph of all  $\overline{KG}_{max} = f(\nabla_i)$  shows that the mass centre of the ship should not be above it when constructing a cargo plan [2].

The method is applied in the paper "An algorithm and a program module for calculating the border height of the mass centre of a vessel" using documentation from the Naval Academy's training ship "Nikola Yonkov Vaptsarov".

### III. CONCLUSIONS

- 1) The evaluation of the stability using the dynamic height of the mass centre (DHMC)

$$(\overline{KG})_{j_{max}} = f(\nabla_i) \tag{6}$$

is better than the existing similarly dependences

$$\overline{KG} = f(d) = f_1(\nabla), \text{ (d - draught)} \tag{7}$$

or

$$\overline{GM} = f(d) = f_2(\nabla). \tag{8}$$

They are based on the criterion of minimal starting metacentric altitude and guarantee basically the starting stability i.e. they should not be considered as reliable evaluation of stability.

- 2) The suggested DHMC removes the necessity of building SSC when realizing a cargo plan. The results are as authentic as the more complex classical SSC.

### ACKNOWLEDGEMENT

The authors would like to acknowledge Assoc. Prof. P. Metodiev for the special help and support.

### REFERENCES

- [1] M. Zhelyazkov, P. Kaloyanchev, "Book of problems based on ships theory", Military publisher, 2001.
- [2] P. Petkov, V. Petkova, I. Draganov, "Theory of the ships and ships construction", Steno, 2008.
- [3] Resolution A 749 (18) of IMO-requirements for SSC.
- [4] STCW Module 17-Introduction to Ships.
- [5] STCW Module 7-Ship Construction, Stability and Damage Control.

# Design of a high – sensitive capacitive sensor for wireless monitoring of bulk material's level

Teodora Trifonova<sup>1</sup>, Valentina Markova<sup>2</sup>, Valentin Todorov<sup>3</sup> and Ventseslav Draganov<sup>4</sup>

**Abstract** - The design of high sensitive capacitive sensor achieved by reducing the destabilizing impact of temperature and supply voltage is reported. The possible influence of the generators' frequencies on the capacitance measurement is analyzed. Stability and sensitivity of several types of stabilized quartz generators is also investigated. Based on results a new intelligent capacitive sensor for measuring the extremes level of bulk materials is created. The developed sensor is suitable for use in wireless sensor network.

**Keywords** – Capacitive sensor, sensor network, quartz generator.

## I. INTRODUCTION

In recent years there has been an explosion in sensor technology. The selection of sensor for a given application depends on the nature of physical objects that must be observed, such as temperature, pressure, humidity or level of materials [1].

Traditionally, environmental monitoring is achieved through expensive sensors with high accuracy. Creating a wireless sensor network provides an alternative solution by deploying a larger number of sensor nodes with less precision. Network as a whole, however, provides better spatial resolution of the area and users can have immediate access to data [2].

The capacitive sensor that controls the extremes of bulk materials with very low permittivity has been developed by authors in [3-5]. The sensor provides high sensitivity in spite of the influence of destabilizing factors like variation of temperature and supply. Changing the capacity of the sensor leads to a change in the frequency of quartz stabilized generator to which the sensor is plugged.

The aim of this work is to further enhance the sensitivity of the sensor by choosing appropriate generators and quartz resonators, and its adaptation for inclusion in the wireless sensor network.

## II. SCHEMATIC DIAGRAM OF THE CAPACITIVE CONVERTER

The block diagram of developed capacitive sensor is shown on fig.1. A capacitive converter S with capacity  $C_x$  is connected to quartz stabilized generator G1. The frequency of the generator is varying within a certain range in capacity adjustment  $C_x$ . The output signal with frequency  $f_x$  from the digital comparator compares the frequency  $f_r$  of the second supporting quartz stabilized generator G2.

T. Trifonova, V. Markova, V. Todorov, V. Draganov are with the Faculty of Electronics, Technical University of Varna, 1 Studentska str., 9000 Varna, Bulgaria. E-mails: [t.trifonova@abv.bg](mailto:t.trifonova@abv.bg), [valliq@abv.bg](mailto:valliq@abv.bg), [todorov\\_88@mail.bg](mailto:todorov_88@mail.bg), [draganov\\_vd@abv.bg](mailto:draganov_vd@abv.bg).

Upon reaching certain, predefined ratio of the both frequencies, control signal is generated at the output of the comparator [6].

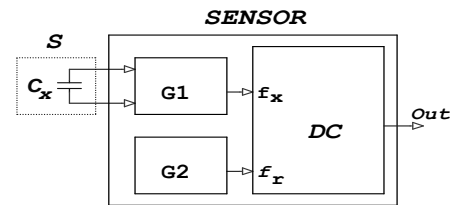


Fig.1. Scheme to monitor the capacity variation of capacitive sensor

The sensitivity of the sensor is determined by the ratio of the frequencies of both generators. It can be changed in wide ranges by setting the digital comparator [7].

To increase the sensitivity of the capacitive sensor it is necessary to reduce the instability of the two generators' frequencies, caused by alteration of the temperature and supply voltage changes of both generators.

There exist a great variety of quartz generators in the literature. Only several schemas of stabilized quartz generators - with an active element transistor, TTL or CMOS integrated circuit fulfill the requirements of our development.

A great number of investigations on selected schemes have been made. It was found different variation of the frequencies of each generator for changes in the ambient temperature and alteration of the supply voltage.

In this publication, due to the limited number of pages only the best results obtained for the two of analyzed schemes are presented. Studies have shown that two basic patterns of quartz stabilized generator - with TTL integrated circuit (Fig. 2) and with CMOS integrated circuit (Fig. 3) are the most appropriate for our sensor. Both schemes are similar. Primary capacitive converter with capacity  $C_x$  is connected to the stabilized quartz generator serially linked with quartz resonator. The frequency of the generator can be adjusted in small ranges.

## III. EXPERIMENTAL RESULTS

For greater accuracy and precision of the sensor it is necessary to use two identical generators. The experiments were performed with two pairs of generators. It was used different types of quartz resonators with serial resonance frequency  $F_s$  varying from 1 MHz to 10 MHz in dependence on the lowest temperature fluctuation.

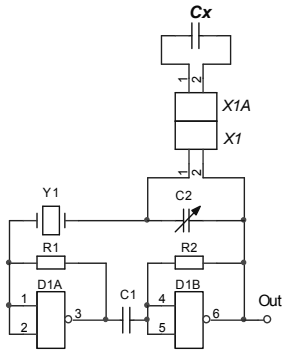


Fig.2. Capacitive converter with TTL integrated circuit

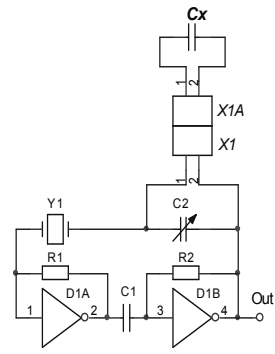


Fig.3. Capacitive converter with CMOS integrated circuit

IV. EXPERIMENTAL RESULTS

For greater accuracy and precision of the sensor it is necessary to use two identical generators. The experiments were performed with two pairs of generators. It was used different types of quartz resonators with serial resonance frequency  $F_S$  varying from 1 MHz to 10 MHz in dependence on the lowest temperature fluctuation.

A. Investigations of capacitive converter with TTL integrated circuit

Temperature dependences of the generators' frequency  $F_X$  for frequencies of the quartz resonators, respectively  $F_{S1.1}=F_{S1.2} = 1$  MHz and  $F_{S2.1}=F_{S2.2} = 5$  MHz are display on Fig.4 and Fig.5.

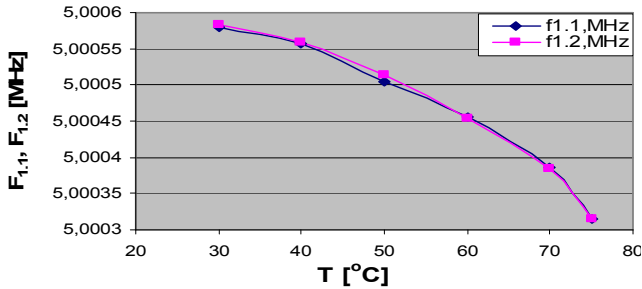


Fig.4. Dependence  $F_x = \varphi(T)$ , where  $F_S = 5$  MHz

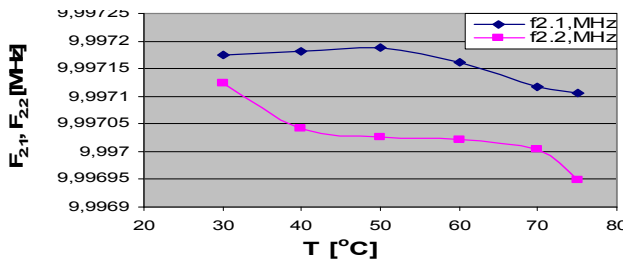


Fig.5. Dependence  $F_x = \varphi(T)$ , where  $F_S = 10$  MHz

It is clearly seen from results that the deviation of the frequencies is extremely small – below  $10^{-4}$  for  $F_S = 5$  MHz and below  $2.10^{-5}$  for  $F_S = 10$  MHz with changes in ambient temperature with  $45^\circ C$ . This corresponds to instability, less than  $10^{-5} / ^\circ C$  ( $F_S = 5$  MHz) and less than  $10^{-6} / ^\circ C$  ( $F_S = 10$  MHz).

As the output of the sensor (fig.1) is given control signal upon reaching a predefined ratio of the frequencies of the two identical generators, it is more important to compare the temperature dependencies of the frequencies ratio of both generators. The results for two quartz generators with frequencies  $F_{S1.1} = F_{S1.2} = 5$  MHz and  $F_{S2.1} = F_{S2.2} = 10$  MHz are presented, respectively on fig.6 and fig.7.

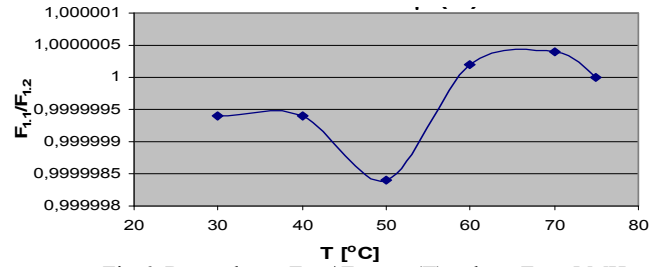


Fig.6. Dependence  $F_{1.1} / F_{1.2} = \varphi(T)$ , where  $F_S = 5$  MHz

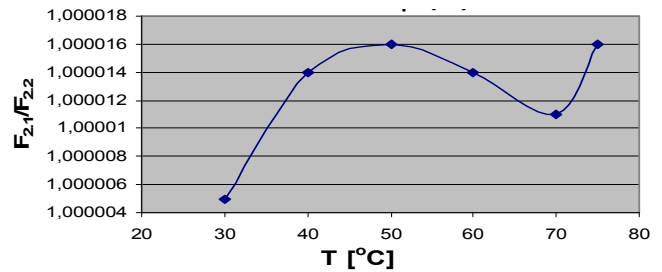


Fig.7. Dependence  $F_{2.1} / F_{2.2} = \varphi(T)$ , where  $F_S = 10$  MHz

Obviously, the deviation of ratio of both generators' frequencies for variation of environmental temperature with  $45^\circ C$  is much less – below  $2.10^{-6}/^\circ C$  for first generators and below  $10^{-5}/^\circ C$  for others. This corresponds to instability, less than  $10^{-7}/^\circ C$ .

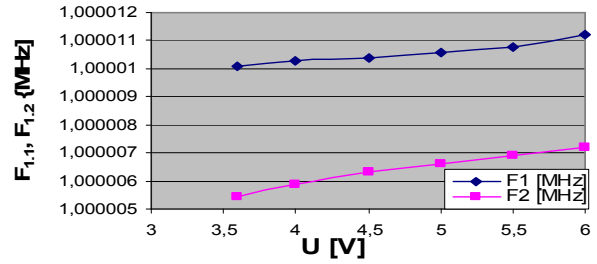


Fig.8. Dependence  $F_{1.1} = \varphi(U)$ ,  $F_{1.2} = \varphi(U)$ , where  $F_S = 1$  MHz

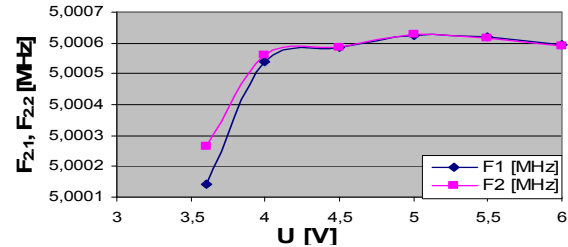


Fig.9. Dependence  $F_{2.1} = \varphi(U)$ ,  $F_{2.2} = \varphi(U)$ , where  $F_S = 5$  MHz

Fig.8 and fig.9 illustrate the dependences of the generator's frequency  $F_X$  for change of the voltage  $U_X$  if the frequencies of quartz resonators are  $F_{S1.1} = F_{S1.2} = 1$  MHz and  $F_{S2.1} = F_{S2.2} = 5$  MHz. The figures show that the frequency is changed in very small range – of the order of  $5.10^{-6}/V$  for the first resonator and below  $10^{-4}/V$  for the second in case of voltage alteration  $U$  from 4 to 6V.



Next two figures (fig.10 and fig.11) are connected with investigation the impact of the variation of the voltage on the ratio of the generators' frequencies, respectively, for frequencies of both quartz generators  $F_{S1,1} = F_{S1,2} = 1$  MHz and  $F_{S2,1} = F_{S2,2} = 5$  MHz.

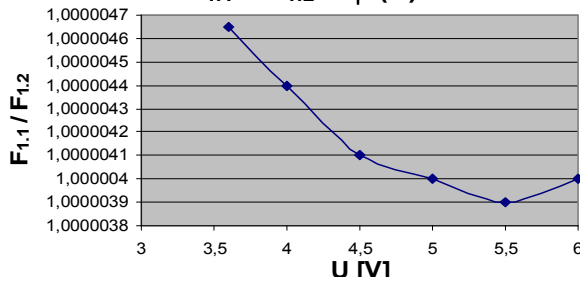


Fig.10. Dependence  $F_{1,1} / F_{1,2} = \varphi(U)$ , where  $F_S = 1$  MHz

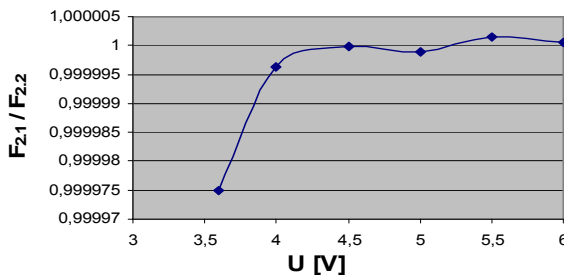


Fig.11. Dependence  $F_{2,1} / F_{2,2} = \varphi(U)$ , where  $F_S = 5$  MHz

As it was expected variation of voltage  $U$  led to alteration in ratio of the frequencies of both generators bellow  $5 \cdot 10^{-7}/V$  for  $F_S = 1$  MHz and less than  $10^{-5}/V$  for  $F_S = 5$  MHz.

*B. Investigations of capacitive converter with CMOS integrated circuit*

The results of experimental studies of the scheme on fig.3 are presented on fig.12 ÷ fig.19.

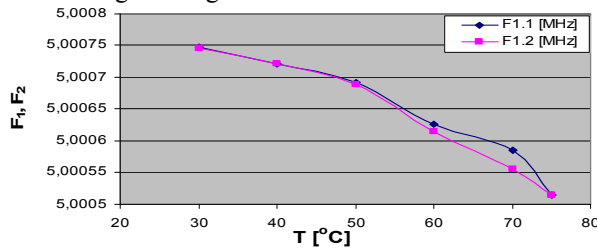


Fig.12. Dependence  $F_{1,1} = \varphi(T)$ ,  $F_{1,2} = \varphi(T)$ , where  $F_S = 5$  MHz

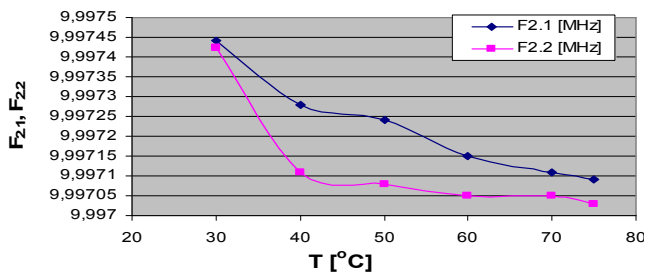


Fig.13. Dependence  $F_{2,1} = \varphi(T)$ ,  $F_{2,2} = \varphi(T)$ , where  $F_S = 10$  MHz

Dependencies of the frequencies of the generators  $F_{XY}$  by the variation of the ambient temperature  $T$ , given frequencies of the quartz resonators, respectively  $F_S = 5$  MHz and  $F_S = 10$  MHz are presented on fig.12 and fig.13. Temperature instability under these conditions is less than  $10^{-5}/^{\circ}C$  for  $F_S = 5$  MHz and bellow  $10^{-6}/^{\circ}C$  for  $F_S = 10$  MHz.

As mentioned in the previous section for proposed schemes (fig.2, fig.3) the temperature dependence of the ratio of both generators' frequencies is more importantly. It can be seen from fig.14 and fig.15 that the temperature instability of the ratio  $F_{1,1} / F_{1,2}$  is less than  $10^{-6}/^{\circ}C$  if  $F_{S1,1} = F_{S1,2} = 5$  MHz, respectively for  $F_{2,1} / F_{2,2}$  the instability is bellow  $3 \cdot 10^{-7}/^{\circ}C$  for  $F_{S2,1} = F_{S2,2} = 10$  MHz.

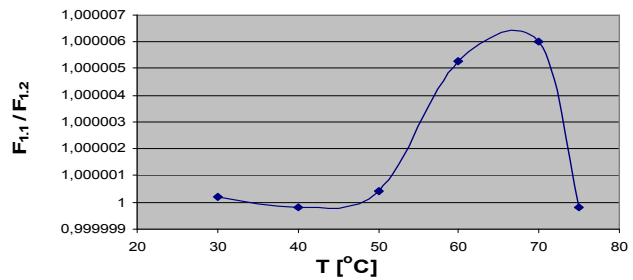


Fig.14. Dependence  $F_{1,1} / F_{1,2} = \varphi(T)$ , where  $F_S = 5$  MHz

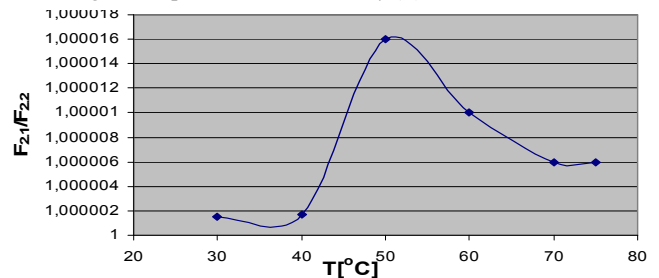


Fig.15. Dependence  $F_{2,1} / F_{2,2} = \varphi(T)$ , where  $F_S = 10$  MHz

The last group of studies are related to determining influence of changes in voltage on the frequency of generator, given frequencies of quartz resonators, respectively  $F_{S1,1} = F_{S1,2} = 1$  MHz and  $F_{S2,1} = F_{S2,2} = 5$  MHz.

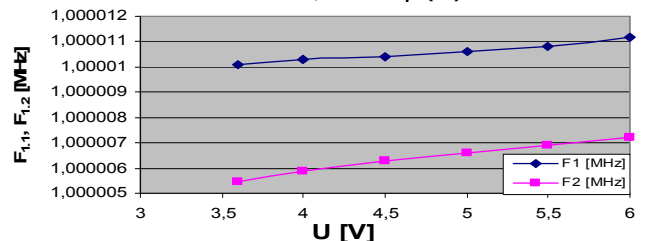


Fig. 16. Dependences  $F_{1,1} = \varphi(U)$ ,  $F_{1,2} = \varphi(U)$ , for  $F_S = 1$  MHz

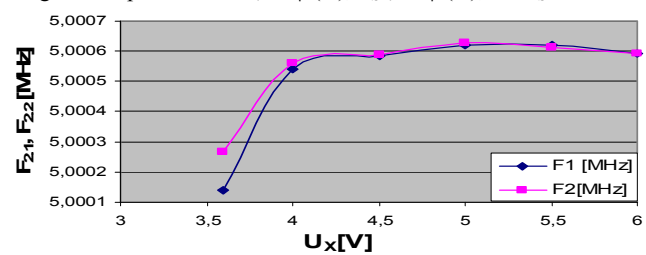


Fig.17. Dependences  $F_{2,1} = \varphi(U)$ ,  $F_{2,2} = \varphi(U)$ , for  $F_S = 5$  MHz

From the resulting dependences, shown on fig.16 and fig.17 follows that the instability of the generators' frequencies under the relevant conditions ( $F_{1,1}$ ,  $F_{1,2}$ ,  $F_{2,1}$ ,  $F_{2,2}$ ) to alteration the voltage is less than  $10^{-5}/V$  for  $F_S = 1$  MHz and bellow  $10^{-4}/V$  for  $F_S = 5$  MHz.

Finally, the ratio of frequencies of both generators dependences to variation of voltage, respectively, for frequencies of both quartz generators  $F_{S1,1} = F_{S1,2} = 1$  MHz and  $F_{S2,1} = F_{S2,2} = 5$  MHz are displayed on fig.18 and fig.19. The frequency instability of both generators is less than  $10^{-6}/V$ .

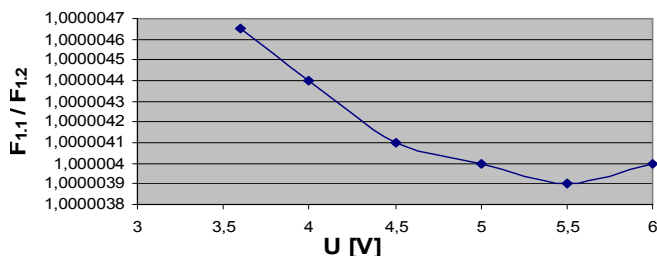


Fig.18. Dependence  $F_{1,1} / F_{1,2} = \varphi(U)$ , for  $F_S = 1$  MHz

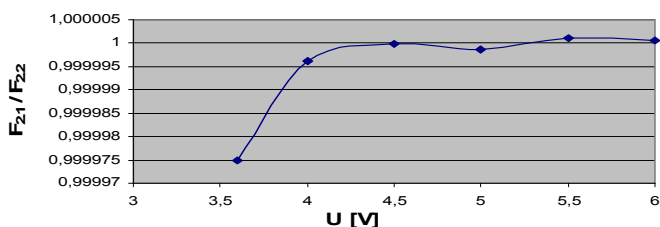


Fig.19. Dependence  $F_{2,1} / F_{2,2} = \varphi(U)$ , for  $F_S = 5$  MHz

### V. ANALYSIS

As it is well known the value of frequency instability varies depending on particular scheme and used quartz resonator. It can be concluded from made experiments that for proposed schemes of capacity sensor the influence of the variation of the ambient temperature and voltage to generated frequencies is extremely small.

In addition for further increasing the sensitivity of the designed capacitive sensor it is necessary to be ensured the equal working conditions for both (the measuring  $G1$  and the supporting  $G2$ ) generators of the scheme shown in Fig.1.

This was achieved in following ways:

1. To ensure a very small difference in operating temperatures of both generators it was proposed constructive decision using common integrated circuit for both generators. The quartz resonators were mounted much close to each other. Their temperature was aligning by an additional thermal connection with heatsink with heat-conveying paste.

Thus, it was assured the temperature difference substantially below  $1^\circ C$ , which leads to increase of the capacitive sensor's sensitivity more than 10 times, as seen from experimental results.

2. Using a common stabilized power source of both generators (fig. 20), with instability of the output voltage below  $0,125 V$ , also enhanced the sensitivity of the capacitive sensor over 10 times.

The proposed integrated stabilizer gives one additional advantage of the sensor – the option of its power to be turned on and off for a predefined period of time. This allows the insertion of the sensor in wireless sensor network.

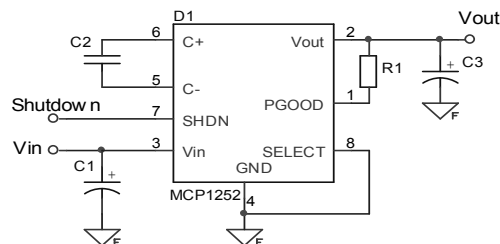


Fig.20. Source of stabilized voltage with control options

### VI. CONCLUSION

It has been made a number of investigations related to reduction of the destabilizing influence of ambient temperature and supply voltage to the frequencies of two quartz generators used in capacitive sensor. Based on results was developed high-sensitive intelligent capacitive sensor for monitoring extremes of bulk materials. The proposed sensor can be used as a node of a wireless sensor network.

### ACKNOWLEDGEMENT

This work was supported by the National Science Funds, under Grant No НПЗ/2012 "Изследване на комуникационни системи с използване на цифрови сигнални процесори".

### REFERENCES

- [1] Dargie W., Ch. Poellabauer. *Fundamentals of Wireless Sensor Networ:Theory and Practice*, Wiley, 2010. Proc. of The VII Int. Conf. ELECTRONICS'98, Book 3, pp. 108-113, 1998
- [2] J. Yick, B. Mukherjee, D. Ghosal, *Wireless sensor network survey*, Computer Networks, Vol. 52, Issue 12, pp.2292-2330, Aug. 2008
- [3] Драганов В., Д. Драганов, Н. Ненков, П. Балабански *Устройство за измерване на ниво*" Авт. свидетелство № 45168, МПК – C 01 G 23/26, 08.01.1988
- [4] Draganov V., Il. Tanchev. *Possibilities for Registering the Changes of Capacity of a Capacity Converter*. Proc. of the Int. Conf Electronics'2004, Sozopol, 22-24 Sep 2004, Book 3, pp.195-199.
- [5] Draganov V., Il. Tanchev. *Device for Measuring the Level of Bulk Materials in Bunkers*, Proc. of the Int. Conf. ICEST'2005, Nish, Serbia, Vol. 1, pp 379-381, June 2005.
- [6] Markova V., R. Dimova, V. Draganov „*An architecture design of a monitoring level sensor system*”, Annual Journal of Electronics, Vol. 5, N 2, 2011, pp. 37-39, ISSN 1313-1842
- [7] Draganov V., T. Trifonova, V. Markova, R. Dimova, D.Kaneva „*Precise measurement opportunities of very small capacity c*

# Different Implementations of Serial Pseudorandom/Natural Code Converters

Dragan Denić<sup>1</sup>, Goran Miljković<sup>2</sup>, Jelena Lukić<sup>3</sup>, Miodrag Arsić<sup>4</sup> and Dragan Živanović<sup>5</sup>

*Abstract* – The duration of the pseudorandom/natural code conversion is critical for the absolute position measurement cycle when pseudorandom position encoders are applied. Because of their simplicity, serial code converters have advantages in implementation. This paper presents serial pseudorandom/natural code converters and proposes one new faster converter. Concrete examples for the proposed solution are also presented.

*Keywords* – position measurement, pseudorandom position encoder, serial pseudorandom/natural code converter

## I. INTRODUCTION

The pseudorandom position encoders, their main advantage, have only one code track regardless to the resolution. This solution for the absolute position measurement is based on property of  $n$ -bit pseudorandom binary sequence (PRBS) that each sliding window of length  $n$ , which passed along a sequence, will extract unique code word in every moment [1]. Also, the last  $(n-1)$  bits of the current code word are equivalent to the first  $(n-1)$  bits of the subsequent code word. The PRBS is useful type of periodic signal and also has the following properties: the signal is periodic and bipolar, signal exhibits a uniform power spectral density over a wide frequency band, signal is deterministic repeatable, etc. Pseudorandom binary sequences are also used in cryptography, bit-error-rate measurements, wireless communication systems, audio applications, etc. The PRBS generator can be implemented using discrete electronics (shift register with D flip-flop cells), using a microprocessor (flexible implementation), using a FPGA-based implementation (flexible and very fast), using virtual instrumentation concept [2], etc.

The main functional parts of pseudorandom position encoder are the code reading system [3, 4], where different solutions are developed (with one, two or more heads), code scanning methods in the sense of reliable code reading

moment defining [4, 5], and error detection methods [3], which increase reliability of encoder. One more functional part of encoder, but no less important than previous ones, is pseudorandom/natural code conversion. Pseudorandom binary code is not suitable for direct application in digital electronics. There are different methods for pseudorandom/natural code conversion, and they can be separated on three distinct groups: parallel [6], serial [4] and serial-parallel code conversion [4]. Parallel solution for code conversion is fast, but expensive and impractical for long PRBS. Serial code conversion is developed as one simple and cheap way for conversion of long PRBS. However, conversion time is critical for one absolute position measurement cycle. Through development of different solutions of serial code converters the main goal is reducing of conversion time. Serial-parallel code conversion is one compromise solution, which combines serial and parallel conversion techniques. During mounting on the shaft pseudorandom encoder provides possibility of direct zero position adjustment without a significant change of hardware and software, but only when serial code conversion is used [7].

In the first part of the paper existing serial pseudorandom/natural code converters are explained, and then on new faster serial converter is proposed. This new solution employed different feedback configuration of logic gates. The presented solutions are detailed explained using appropriate concrete example.

## II. THE SERIAL PSEUDORANDOM/NATURAL CODE CONVERTERS

The simple solution for pseudorandom/natural code conversion is the serial or sequential pseudorandom/natural code conversion method [4], but in the case of high resolution, the conversion time becomes a limiting factor. This method finds the actual value of the position ' $p$ ' simply by counting the steps that the shift register with inverse feedback needs until it reaches the initial state by successive shifting from the read pseudorandom  $n$ -bit word. Serial pseudorandom/natural code conversion process for  $n = 7$  is shown in Fig. 1. Pseudorandom code on the code track is read using of only one code reading head  $x(7)$  [4]. In the code conversion process one Fibonacci generator with inverse feedback configuration is applied (Fig. 1). The Fibonacci implementation consists of a shift register in which an exclusive-OR (XOR) gates for modulo-2 sum of the binary-weighted taps are used for feedback configuration. The states of the shift register are actually sequential code words of pseudorandom binary sequence until it came to the state that corresponds to the initial code word. The forbidden state is usually referred to be 0000000, because when all the flip-flop

<sup>1</sup>Dragan Denić is with the Faculty of Electronic Engineering at University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: dragan.denic@elfak.ni.ac.rs

<sup>2</sup>Goran Miljković is with the Faculty of Electronic Engineering at University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: goran.miljkovic@elfak.ni.ac.rs

<sup>3</sup>Jelena Lukić is with the Faculty of Electronic Engineering at University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: jelenalukicpk@gmail.com

<sup>4</sup>Miodrag Arsić is with the Faculty of Electronic Engineering at University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: miodrag.arsic@elfak.ni.ac.rs

<sup>5</sup>Dragan Živanović is with the Faculty of Electronic Engineering at University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: dragan.zivanovic@elfak.ni.ac.rs

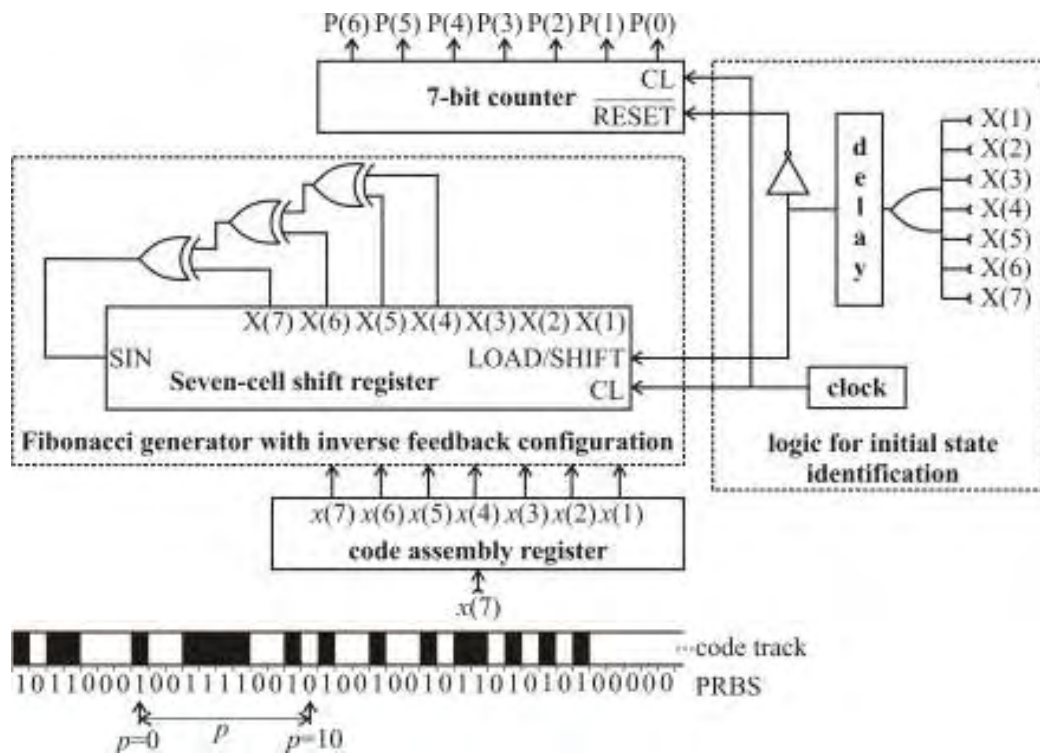


Fig. 1. Serial pseudorandom/natural code converter based on Fibonacci implementation

values are 0, the XOR will reveal a 0 regardless of the location of the taps. Therefore, the feedback value is always 0, and the shift register stays in the 0000000 state. If one of the feedback values are inverted, or XNOR instead XORs are used, the forbidden state may be altered.

Also, in the serial pseudorandom/natural code converter one 7-bit counter is added that counts steps and also the logic for the initial state identification. The basic advantage of this code converter is simplicity, and the disadvantage is the serial connection of logical elements (XOR gates) in the feedback configuration, which increases the total propagation delay and thus conversion time is limited. On the other hand, the conversion time limits the maximum rotation speed of the encoder. The table which contains for maximal length pseudorandom sequences feedback sets for different shift register sizes is given in [2, 8].

One way of reducing the code conversion time of the previous method approximately two times is based on the idea that, thanks to PRBS cycling property, the initial state could be reached using feedback sets that are used for either “direct” or “inverse” PRBS generating [9]. Depending on the previous position of the movable system it is decided which PRBS generating low (“direct” or “inverse”) would be used for current code conversion.

Another solution for code conversion process is using the Galois implementation of PRBS generator, which consists of a shift register, the content of which is modified at every step by a binary-weighted value of the output stage, using XOR gates. The pseudorandom binary sequence generator with a parallel feedback logic configuration (Galois generator of pseudorandom binary sequence) is known as a faster pseudorandom binary sequence generator [10]. The Galois

generator is generally faster than the Fibonacci in hardware due to the reduced number of logic gates in the feedback loop. Now, the total propagation delay in the feedback configuration is equal to the propagation delay of only one logical gate. The order of the Galois weights is opposite that of the Fibonacci weights, for given identical feedback set. The pseudorandom/natural code converter based on the Galois generator is shown in Fig. 2. It is added a logic that the read code word converts to the appropriate content of the shift register. When the code word is read, this logic provides the equivalent state of the shift register that is loaded in that shift register. This logic does not participate further in the code conversion process and thus negligibly influences to conversion time. Furthermore, the steps counted that are needed for the shift register with the determined and the written state come from the known initial state of the shift register. The obtained number is the result of the conversion, the same as in the case of the known serial code converter (Fig. 1). In the second part of the paper the proposed logic for translation of the read code word in the appropriate content of the shift register is explained in detail, without which it would not be possible to realise the new code converter.

### III. LOGIC FOR INITIAL ADJUSTMENT OF READ PSEUDORANDOM CODE WORD

The read  $n$ -bit pseudorandom code word (assigned as  $x = x_n x_{n-1} x_{n-2} \dots x_2 x_1$ ) in real time is not identical to the  $n$ -bit current content of the shift register (assigned as  $X = X_n X_{n-1} X_{n-2} \dots X_2 X_1$ ), which corresponds to the position of this code word in the generated pseudorandom binary sequence. For each  $n$ -bit code word of the pseudorandom binary

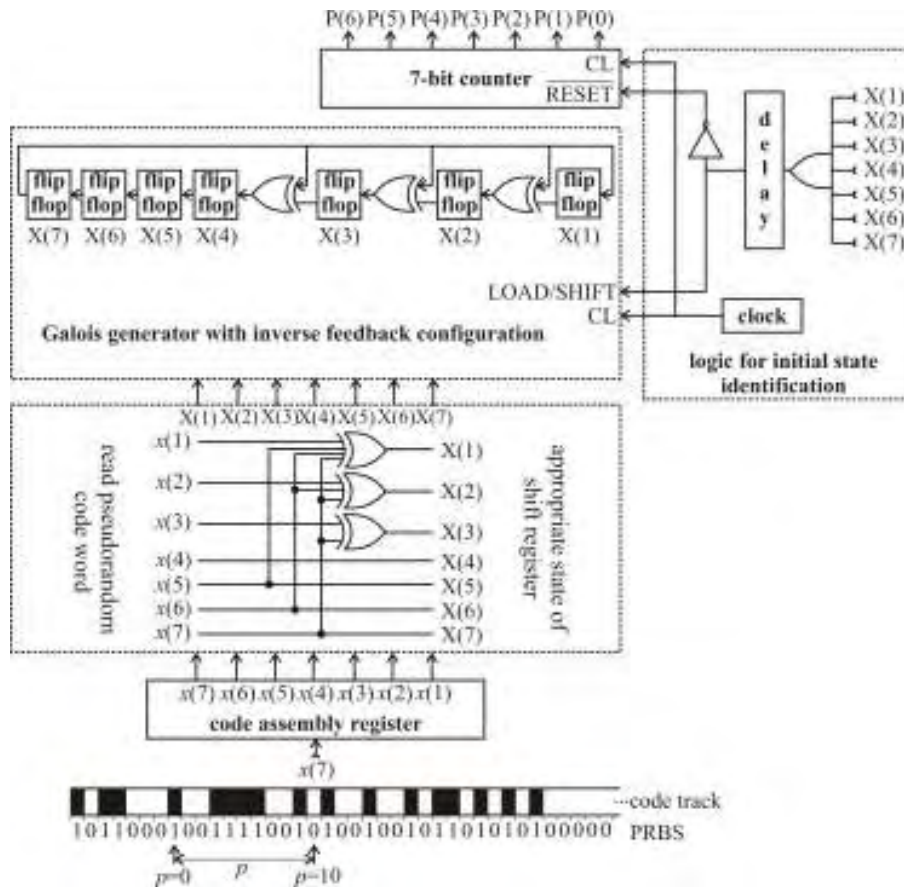


Fig. 2. Faster serial pseudorandom/natural code converter based on Galois implementation

sequence there corresponds exactly one state of the shift register with a parallel feedback logic configuration and it is possible to design a simple logic composed from XOR gates (Fig. 2). For  $n = 7$  will be shown process of logic design, and such a procedure is applicable to any other pseudorandom binary code word of arbitrary length. On the Fig. 3 is shown passing through the states of the 7-bit shift register with a parallel feedback logic, which is known as a ‘Galois’ shift register [8, 10]. On the start, suppose that the initial content of the shift register is  $\{X_1, X_2, X_3, X_4, X_5, X_6, X_7\}$  and the pseudorandom code word  $\{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$  corresponds to that content. The direct generation law of PRBS and the moving direction from  $X_1$  to  $X_7$  is applied. The pseudorandom bit output is always identical to the state  $X_7$ , and

$$X_7 = x_7 \tag{1}$$

After the first clock pulse, the content of the shift register becomes  $\{X_7', X_6', X_5', X_4', X_3', X_2', X_1'\}$ , where, in accordance with the direct generation law of pseudorandom binary sequences for  $n = 7$ :

$$\begin{aligned} X_7' &= X_6 \\ X_6' &= X_5 \\ X_5' &= X_4 \\ X_4' &= X_7 \oplus X_3 \\ X_3' &= X_7 \oplus X_2 \\ X_2' &= X_7 \oplus X_1 \\ X_1' &= X_7 \end{aligned}$$

The relations for the parallel feedback logic configuration are well known, based on the known serial feedback logic configuration of the  $n$ -bit shift register [8]. According to basic mathematical relationships  $X_7' = X_6$ , and  $X_7' = x_6$ , can be concluded

$$X_6 = x_6 \tag{2}$$

Then, after the second clock pulse the content of the shift register becomes  $\{X_7'', X_6'', X_5'', X_4'', X_3'', X_2'', X_1''\}$ . According to previous principle the following relations are obtained:

$$\begin{aligned} X_7'' &= X_6' \\ X_6'' &= X_5' \\ X_5'' &= X_4' \\ X_4'' &= X_7' \oplus X_3' \\ X_3'' &= X_7' \oplus X_2' \\ X_2'' &= X_7' \oplus X_1' \\ X_1'' &= X_7' \end{aligned}$$

Since  $X_7'' = x_5$  and  $X_7'' = X_6'$ , plus from the previous clock pulse the valid relation is  $X_6' = X_5$ , there is obtained

$$X_5 = x_5 \tag{3}$$

With the identical procedure for the next clock pulse or by writing the relations on the same principle and using the relations from the previous clock pulse, and also properties of modulo-2 sum, the following dependences are obtained:

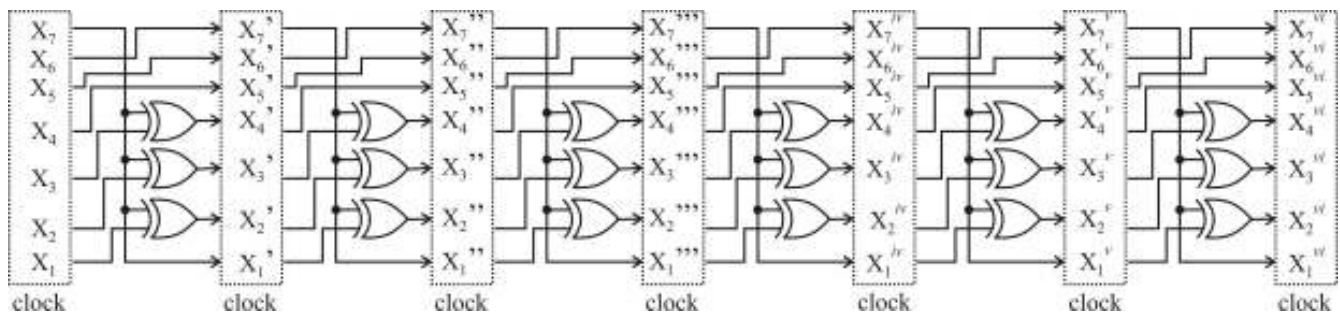


Fig. 3. Contents of the ‘Galois’ shift register through 7 clock periods

$$\begin{aligned}
 X_4 &= x_4 & (4) \\
 X_3 &= x_3 \oplus x_7 & (5) \\
 X_2 &= x_2 \oplus x_6 \oplus x_7 & (6) \\
 X_1 &= x_1 \oplus x_5 \oplus x_6 \oplus x_7 & (7)
 \end{aligned}$$

These seven relations define relations between the content of the register and the appropriate pseudorandom binary code word, and also define the logic for initial adjustment of the read pseudorandom binary code word shown in Fig. 2.

Now compare the conversion of the read pseudorandom code word {0, 1, 1, 0, 0, 0, 1} for the case of applying the first explained serial code converter and the proposed faster serial code converter. The initial code word is {1, 1, 1, 0, 0, 1, 0}. According to the algorithms described in [4], the pseudorandom/ natural code conversion is accomplished sequentially after loading it into a shift register having a reverse feedback equation  $X(1) = X(4) \oplus X(5) \oplus X(6) \oplus X(7)$  (Fig. 1). In the given example there are 10 shifts of register, and the counter state is  $p = 10$  at the end, which is actually the value of the current position of the movable system. Let us now look at the new code converter, which is shown in Fig. 2. The code track is the same as shown in Fig. 1. Now, the read code word is not written directly to the shift register, but feeds the input of logic for the initial adjustment of the read code word (Fig. 2). By the application of relations (1), (2), (3), (4), (5), (6), (7) the code word {0, 1, 1, 0, 0, 1, 1} is obtained as output. It is now directly saved in the shift registry. Now, the shift register sequentially passes through the following states: {1, 1, 0, 0, 1, 1, 0}, {1, 0, 0, 0, 0, 1, 1}, {0, 0, 0, 1, 0, 0, 1}, {0, 0, 1, 0, 0, 1, 0}, {0, 1, 0, 0, 1, 0, 0}, {1, 0, 0, 1, 0, 0, 0}, {0, 0, 1, 1, 1, 1, 1}, {0, 1, 1, 1, 1, 1, 0}, {1, 1, 1, 1, 1, 0, 0}, and {1, 1, 1, 0, 1, 1, 1} when the stop will be. The state {1, 1, 1, 0, 1, 1, 1} is the content of the register, which corresponds to the initial pseudorandom code word {1, 1, 1, 0, 0, 1, 0}. So, at the end of conversion the counter state is  $p = 10$ , which is exactly the same value as in the case of the first serial code converter.

#### IV. CONCLUSION

During development of this faster serial pseudorandom/nature code converter the goal was to reduce the conversion time. It is achieved by the reduction in the number of serial connected gates in the feedback logic, which provides less propagation delay. For implementation of code converter, the parallel feedback logic configuration is applied and there has also been designed a simple logic of initial

adjustment of the read code word into a appropriate state of the shift register, without which it would not be possible to realise the code converter proposed here.

#### ACKNOWLEDGEMENT

Research activities presented in this paper, are supported by funds of the Ministry of Education and Science of the Republic of Serbia, having the reference project number TR32045.

#### REFERENCES

- [1] S. Engelberg and H. Benjamin, “Pseudorandom sequences and the measurement of the frequency response”, *IEEE Instrum. Meas. Mag.*, vol. 8, no. 1, pp. 54–59, 2005.
- [2] G. Miljković, I. Stojković and D. Denić, “Generation and application of pseudorandom binary sequences using virtual instrumentation”, *Facta Universitatis, Series: Automatic Control and Robotics*, vol. 10, no. 1, pp. 51–58, 2011.
- [3] M. Arsić and D. Denić, “New pseudorandom code reading method applied to position encoders”, *Electron. Lett.*, vol. 29, no. 10, pp. 893–894, 1993.
- [4] E.M. Petriu, “Absolute position measurement using pseudorandom binary encoding”, *IEEE Instrum. Meas. Mag.*, vol. 1, no. 3, pp. 19–23, 1998.
- [5] D. Denić and G. Miljković, “Code reading synchronization method for pseudorandom position encoders”, *Sens. Actuators A*, vol. 150, pp. 188–191, 2009.
- [6] E.M. Petriu, J.S. Basran and F.C.A. Groen, “Automated guided vehicle position recovery”, *IEEE Trans. Instrum. Meas.*, vol. 39, no. 1, pp. 254–258, 1990.
- [7] D. Denić, I. Randelović and G. Miljković, “Recent trends of linear and angular pseudorandom encoder development”. *Int. Symp. on Power Electronics, Electrical Drives, Automation and Motion (SPEEDAM)*, Taormina, Sicily, Italy, pp. 746–750, 2006.
- [8] “Linear feedback shift register: Implementation, M-sequence properties, feedback tables”, *New Wave Instruments*, January 2004 [Online] Available: [http://www.newwaveinstruments.com/resources/articles/m\\_sequence\\_linear\\_feedback\\_shift\\_register\\_lfsr/](http://www.newwaveinstruments.com/resources/articles/m_sequence_linear_feedback_shift_register_lfsr/)
- [9] D. Denić, I. Randelović and M. Rančić, “Pseudorandom position encoder and code conversion problems”, *ICEST*, Ohrid, Macedonia, pp. 437-440, 16-19 June 2004.
- [10] S. Bourdel, E. Campo, P. Melet and L. Andrieux, “From modelling of a CDMA transceiver in indoor environment to an ASIC circuit synthesis”, *J. Telecommun. Inf. Technol.*, vol. 3, pp. 33–45, 2001.

# Cloud systems for environmental telemetry - A case study for ecological monitoring in agriculture

George Suciu<sup>1</sup>, Octavian Fratu<sup>1</sup>, Cristian Cernat<sup>1</sup>, Traian Militaru<sup>1</sup>, Gyorgy Todoran<sup>1</sup> and Vlad Poenaru<sup>1</sup>

**Abstract** – Large telemetry systems have several hundreds of RTUs that are sending data to be processed by intelligence algorithms and stored in a database that is accessible via Web interface on the Internet. In this paper we present the way in which SlapOS, an open source provisioning and billing system for distributed cloud computing, is used to gather centrally environmental information from different sensors at remote observation points.

**Keywords** – Cloud, Telemetry, Sensors, Remote monitoring, RTU.

## I. INTRODUCTION

In this paper we develop a test platform for environmental telemetry and use it as a case study for monitoring ecological parameters in agriculture. We use different types of RTUs (Radio Transmission Units) and Sensors that monitor and transmit important information such as temperature, precipitation, wind speed and leaf wetness from selected locations.

The RTUs will transmit sensor data over GSM/GPRS to our cloud platform where we can conveniently process the site-specific weather and soil data in near real-time, display it in our web-based visualization application and get detailed recommendations when and where to spray and how much to irrigate - resulting in optimized yield, quality and income.

Our system can also help keeping track of pathogen development, optimize treatments to hit a disease dead on, warn of frost, and to produce crops as environmentally friendly as possible and to improve agricultural risk management. Falling producer prices and rising costs of production are increasingly forcing agricultural businesses to optimize production costs [1]. Therefore "precision farming", the selective use of inputs such as water, fertilizers or chemicals, is now indispensable in modern agriculture. The growing environmental awareness of consumers further accelerates this process and promotes the usage of remote automatic monitoring system for field information such as the one we developed [2].

We will introduce in this article SlapOS, the first open source operating system for Distributed Cloud Computing. SlapOS is based on a grid computing daemon called slapgrid which is capable of installing any software on a PC and instantiate any number of processes of potentially infinite

duration of any installed software. Slapgrid daemon receives requests from a central scheduler the SlapOS Master which collects back accounting information from each process. SlapOS Master follows an Enterprise Resource Planning (ERP) model to handle at the same time process allocation optimization and billing. SLAP stands for "Simple Language for Accounting and Provisioning".

This structure has been implemented for cloud-based automation of ERP and CRM software for small businesses and aspects are under development under the framework of the European research project "Cloud Consulting" [3]. We will use our platform hosted on several servers running Ubuntu Linux – Apache – MySQL template with current software release. On our cloud testing environment we provide the platform for processing information from hundreds different sensors, enabling the analysis of environmental data through a large sample of RTUs.

In previous approaches RTUs were implemented in most cases on a local server and no company could aggregate enough sensor data to consider automating the treatment process.

## II. CLOUD ARCHITECTURE FOR TELEMETRY

### A. Cloud Architecture

SlapOS is an open source Cloud Operating system which was inspired by recent research in Grid Computing and in particular by BonjourGrid [4]–[5] a meta Desktop Grid middleware for the coordination of multiple instances of Desktop Grid middleware. It is based on the motto that "everything is a process".

SlapOS is based on a Master and Slave design. In this chapter we are going to provide an overview of SlapOS architecture and are going in particular to explain the role of Master node and Slave nodes, as well as the software components which they rely on to operate a distributed cloud for telemetry applications.

Slave nodes request to Master nodes which software they should install, which software they show run and report to Master node how much resources each running software has been using for a certain period of time. Master nodes keep track of available slave node capacity and available software. Master node also acts as a Web portal and Web service so that end users and software bots can request software instances which are instantiated and run on Slave nodes. Master nodes are stateful. Slave nodes are stateless. More precisely, all information required to rebuild a Slave node is stored in the Master node. This may include the URL of a backup service which keeps an online copy of data so that in case of failure of

<sup>1</sup>The authors are with the Faculty of Electronics, Telecommunications and Information Technology at Politehnica University of Bucharest, Bd. Iuliu Maniu, nr. 1-3, Bucharest 060042, Romania, E-mails: george@beia.ro, ofratu@elcom.pub.ro, cernatcristi@gmail.com, gelmosro@yahoo.com, todran.gyorgy@gmail.com, vlad.wing@gmail.com.

a Slave node, a replacement Slave node can be rebuilt with the same data.

It is thus very important to make sure that the state data present in Master node is well protected. This could be implemented by hosting Master node on a trusted IaaS infrastructure with redundant resource. Or - better - by hosting multiple Master nodes on many Slave nodes located in different regions of the world thanks to appropriate data redundancy heuristic. We are touching here the first reflexive nature of SlapOS. A SlapOS master is normally a running instance of SlapOS Master software instantiated on a collection of Slave nodes which, together, form a trusted hosting infrastructure. In other terms, SlapOS is self-hosted, as seen in Fig. 1.

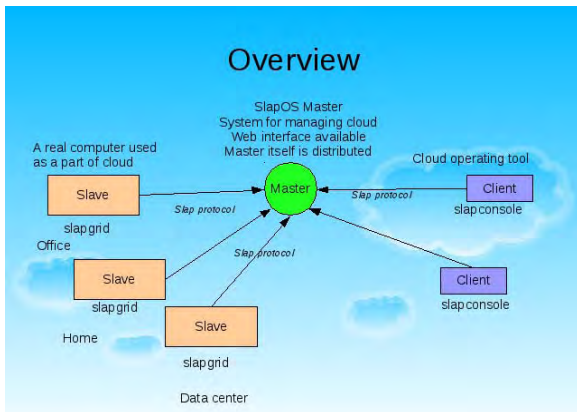


Fig. 1. SlapOS Master – Slave Architecture

**B. SlapOS Master**

SlapOS master nodes keep track of the identity of all parties which are involved in the process of requesting Cloud resources, accounting Cloud resources and billing Cloud resources. This includes end users (Person) and their company (Organisation). It includes suppliers of cloud resources as well as consumers of cloud resources. It also includes so-called computer partitions which may run a software robot to request Cloud resources without human intervention. It also includes Slave nodes which need to request to SlapOS master which resources should be allocated. SlapOS generated X509 certificates for each type of identity: X509 certificates for people like you and me who login, an X509 certificate for each server which contributes to the resources of SlapOS and an X509 for each running software instance which may need to request or notify SlapOS master. A SlapOS Master node with a single Slave node, a single user and 10 computer partitions will thus generate up to 12 X509 certificates: one for the slave, one for the user and 10 for computer partitions.

Any user, software or slave node with an X509 certificate may request resources to SlapOS Master node. SlapOS Master node plays here the same role as the backoffice of a marketplace. Each allocation request is recorded in SlapOS Master node as if it were a resource trading contract in which a resource consumer requests a given resource under certain conditions. The resource can be a NoSQL storage, a virtual machine, an ERP with web-portal interface for displaying

sensor data and Google Maps integration for RTUs localization, a Wiki, etc. The conditions can include price, region (ex. China) or specific hardware (ex. 64 bit CPU). Conditions are somehow called Service Level Agreements (SLA) in other architectures but they are considered here rather as trading specifications than guarantees. It is even possible to specify a given computer rather than relying on the automated marketplace logic of SlapOS Master.

By default, SlapOS Master acts as an automatic marketplace. Requests are processed by trying to find a Slave node which meets all conditions which were specified. SlapOS thus needs to know which resources are available at a given time, at which price and under which characteristics. Last, SlapOS Master also needs to know which software can be installed on which Slave node and under which conditions.

**C. SlapOS Slave**

SlapOS Slave nodes are relatively simple compared to the Master node. Every slave node needs to run software requested by the Master node. It is thus on the Slave nodes that software is installed. To save disk space, Slave nodes only install the software which they really need.

Each slave node is divided into a certain number of so-called computer partitions. One may view a computer partition as a lightweight secure container, based on Unix users and directories rather than on virtualization. A typical barebone PC can easily provide 100 computer partitions and can thus run 100 RTU web portals or 100 sensors monitoring sites, each of which with its own independent database. A larger server can contain 200 to 500 computer partitions.

SlapOS approach of computer partitions was designed to reduce costs drastically compared to approaches based on a disk images and virtualization. As presented in Fig. 2, it does not prevent from running virtualization software inside a computer partition, which makes SlapOS at the same time cost efficient and compatible with legacy software.

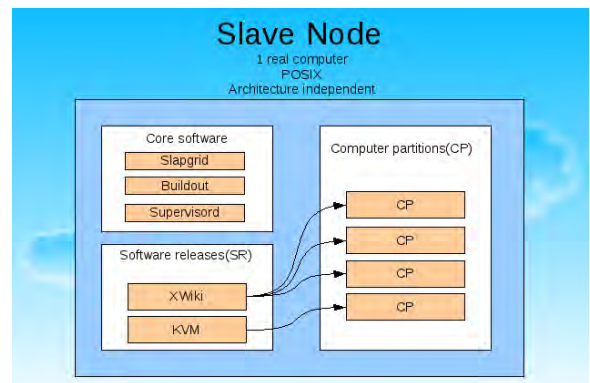


Fig. 2. SlapOS Slave Node

SlapOS Slave software consists of a POSIX operating system, SlapGRID, supervisor and buildout [3]. SlapOS is designed to run on any operating system which supports GNU's glibc and supervisor. Such operating systems include for example GNU/Linux, FreeBSD, MacOS/X, Solaris, AIX, etc



D. SlapOS kernel

SlapOS relies on mature software: buildout and supervisor. Both software are controlled by SLAPGrid, the only original software of SlapOS. SLAPGrid acts as a glue between SlapOS Master node (ERP5) and both buildout and supervisor, as shown in Fig. 3. SLAPGrid requests to SlapOS Master Node which software should be installed and executed. SLAPGrid uses buildout to install software and supervisor to start and stop software processes. SLAPGrid also collects accounting data produced by each running software and sends it back to SlapOS Master.

Supervisord is a process control daemon. It can be used to programmatically start and stop processes with different users, handle their output, their log files, their errors, etc. It is a kind of much improved init.d which can be remotely controlled. Supervisord is lightweight and old enough to be really mature (ie. no memory leaks).

Buildout is a Python-based build system for creating, assembling and deploying applications from multiple parts, some of which may be non-Python-based. Buildout can be used to build C, C++, ruby, java, perl, etc. software on Linux, MacOS, Windows, etc. Buildout can either build applications by downloading their source code from source repositories (subversion, git, mercurial, etc.) or by downloading binaries from package repositories (rpm, deb, eggs, gems, war, etc.). Buildout excels in particular at building applications in a way which is operating system agnostic and to automate application configuration process in a reproducible way.

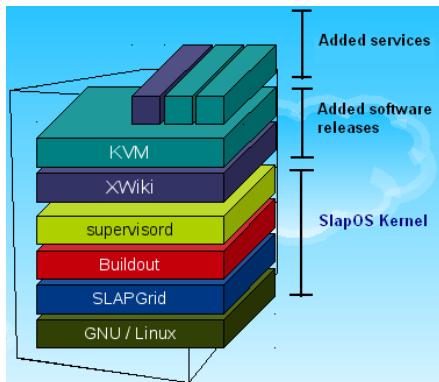


Fig. 3. SlapOS Kernel and User Software

Every computer partition consists of a dedicated IPv6 address, a dedicated local IPv4 address, a dedicated tap interface (slaptapN), a dedicated user (slapuserN) and a dedicated directory (/srv/slapgrid/slappartN). Optionally, a dedicated block device and routable IPv4 address can be defined.

SlapOS is usually configured to use IPv6 addresses. Although use of IPv6 is not a requirement (an IPv4 only SlapOS deployment is possible) it is a strong recommendation. IPv6 simplifies greatly the deployment of SlapOS either for public Cloud applications or for private Cloud applications. In the case of public Clouds, use of IPv6 helps interconnecting SlapOS Slave Nodes hosted at home without having to setup tunnels or complex port redirections.

In the case of private Cloud, IPv6 replaces existing corporate tunnels with a more resilient protocol which provides also a wider and flat corporate addressing space. IPv6 addressing helps allocating hundreds of IPv6 addresses on a single server. Each running process can thus be attached to a different IPv6 address, without having to change its default port settings. Accounting network traffic per computer partition is simplified. All this would of course be possible with IPv4 or through VPNs but it would be much more difficult or less resilient. The exhaustion of IPv4 addresses prevents in practice allocation of some many public IPv4 addresses to a single computer. After one year of experimentation with IPv6 in Romania, using IPv6 native Internet access (more than 50% of worldwide IPv6 traffic), we found that IPv6 is simple to use and creates the condition for many innovations which would else be impossible.

III. RESEARCH RESULTS AND FUTURE DESIGN

In order to collect the information from the RTUs we developed the following test platform as shown in Fig. 4. The usage of GSM/GPRS data transmission can be extended in areas where there is no coverage by using a UHF bridge operating in the fixed frequency range 430 – 440 MHz connected to a gateway that has access to the Internet.

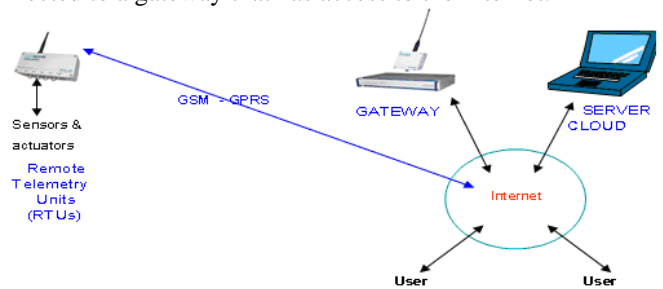


Fig. 4. General Architecture of Telemetry System

The case study was done on 2 grape yards in Romania (Bucharest and Blaj) with the following sensors, as seen in Fig. 5

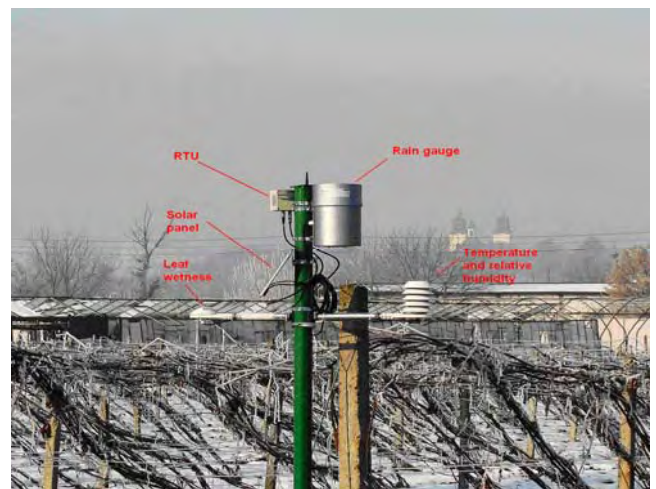


Fig. 5. General Structure of RTU and Sensors

The total quantity of rain reported by the system during the months of May – September 2011 was of 222 l/sqm with the following monthly distribution: May – 33 l/sqm, June – 116 l/sqm, July – 49 l/sqm, August - 5 l/sqm, September – 8 l/sqm. Other climatic parameters such as Precipitation, Leaf Wetness, Temperature and Relative Humidity can be seen on Fig. 6.

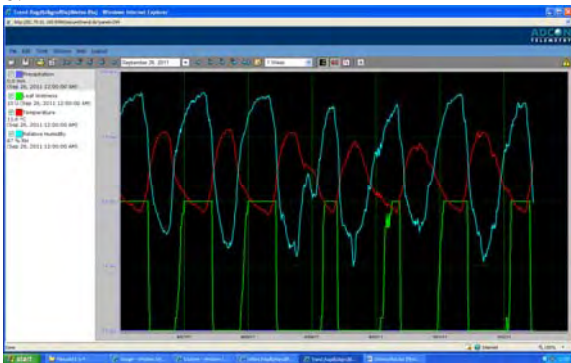


Fig. 6. Results climatic parameters during a week (26.09 – 02.10.2011)

Another important parameter we studied is the accumulation of thermal energy over time, known as degree-days or heat units. The growth and development of plants, insects, and many other invertebrate organisms is largely dependent on temperature. In other words, a constant amount of thermal energy is required for the growth and development of many organisms, but the time period over which that thermal energy is accumulated can vary. Many organisms slow or stop their growth and development when temperatures are above or below threshold levels. Degree-days and other heat unit measurements have been used for determination of planting dates, prediction of harvest dates, and selection of appropriate crop varieties.

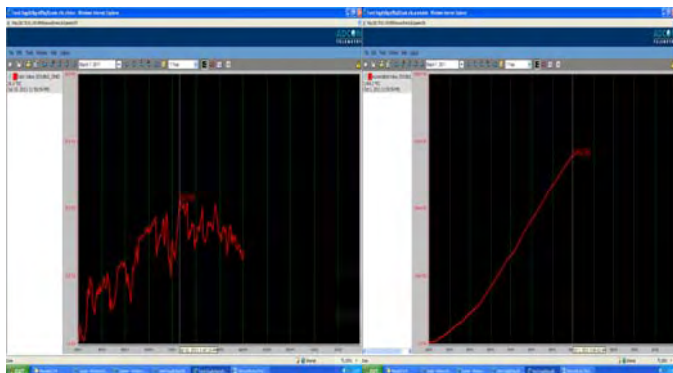


Fig. 7. Heat units graphs (daily degree-days and total)

The calculation methods available for heat unit include: Averaging, Standard, GDD (Growing Degree-Days), Single Triangle, Double Triangle, Single Sine, Double Sine and Near Real-Time. As shown in Fig. 7 we used the Averaging Method and the maximum heat unit (26,3 degree days) was calculated on the date of 10.07.2011 and the total accumulated thermal energy by the crop on 01.10.2011 was 3.484,2 degree-days.

#### IV. CONCLUSION

Our system for environmental telemetry can be adapted also to other applications besides agriculture and meteorology. Knowing how the weather will be is important but knowing how the environmental parameters are right now is just as much important for power plants, airports, wind and solar parks, incinerators and landfills - they all need wind, temperature, radiation data, etc. reliably and up to date.

Even though IPv6 is used to interconnect processes globally on a SlapOS public or private Cloud, we found that some existing software on RTUs are incompatible with IPv6. Reasons vary. Sometimes, IP addresses are stored in a structure of 3 integers, which is incompatible with IPv6. Sometimes, IPv6 URLs are not recognized since only dot is recognized as a separator in IP addresses. For this reason, we decided to provide to each computer partition a dedicated, local, non routable IPv4 address.

We hope in the future that Microsoft Windows will also be supported as a host (Microsoft Windows is already supported as a guest) through glibc implementation on Windows and a port of supervisor to Windows.

#### ACKNOWLEDGEMENT

This paper is presented as part of the project “*Valorificarea capitalului uman din cercetare prin burse doctorale (ValueDoc)*” Project co-financed from the European Social Fund through POSDRU, financing contract POSDRU/107/1.5/S/76909 and part of the project “*Cloud Consulting*” and “*TELE GREEN*”.

#### REFERENCES

- [1] Tokihiro Fukatsu, Tomonari Watanabe, Haoming Hu, Hideo Yoichi, Masayuki Hirafuji, “Field monitoring support system for the occurrence of *Leptocorisa chinensis* Dallas (Hemiptera: Alydidae) using synthetic attractants, Field Servers, and image analysis”, *Computers and Electronics in Agriculture*, vol. 80, January 2012, pp. 8–16, 2012
- [2] Jiang, J.A.; Tseng, C.L.; Lu, F.M.; Yang, E.C.; Wu, Z.S.; Chen, C.P.; Lin, S.H.; Lin, K.C.; Liao, C.S., “A GSM-based remote wireless automatic monitoring system for field information: A case study for ecological monitoring of the oriental fruit fly, *Bactrocera dorsalis* (Hendel)”, *Computers and Electronics in Agriculture* vol. 62, Issue 2, July 2008, pp. 243–259, 2008
- [3] George Suci, Octavian Fratu, Simona Halunga, Cristian George Cernat, Vlad Andrei Poenaru, Victor Suci, “Cloud Consulting: ERP and Communication Application Integration in Open Source Cloud Systems”, 19th Telecommunications Forum - TELFOR 2011, IEEE Communications Society, pp. 578-581, 2011
- [4] Heithem Abbes, Christophe C´erin, and Mohamed Jemni. Bonjourgrid as a decentralised job scheduler. In APSCC 08. Proceedings of the 2008 IEEE Asia-Pacific Services Computing Conference, pages 89–94, Washington, DC, USA, 2008. IEEE Computer Society.
- [5] Heithem Abbes, Christophe C´erin, Mohamed Jemni: A decentralized and fault-tolerant Desktop Grid system for distributed applications. *Concurrency and Computation: Practice and Experience* 22(3): 261-277 (2010)

# High-quality Primary School Education in the Field of Electrotechnics and Informatics - Beginning of the Development of Successful Engineers

Sonja Cvetkovic<sup>1</sup> and Zoran Stankovic<sup>2</sup>

**Abstract** – Taking into consideration that education of future electrical and IT engineers starts in primary school through acquisition of the initial knowledge in the field of electrotechnics and informatics, the presented abstract is dedicated that segment of education. It shows that present implementation level of knowledge acquisition and education support in this field by the current educational plans and programmes for primary schools in the Republic of Serbia. It shows the outcome of the undertaken researches based on a poll conducted amongst the pupils and the questionnaire referred to the preference of subjects, teaching and methodical units and other activities related to knowledge acquisition in the field of electrotechnics and informatics. It also explains possibilities, available methods and strategies for pupil education quality improvement in these fields which are supported by the current educational plans and programmes for primary schools in the Republic of Serbia

**Keywords** – Electrical engineers education, Primary school education, Education quality

## I. INTRODUCTION

The education of future electrotechnics engineers starts in primary school through acquisition of the initial knowledge in the field of electrotechnics and informatics. Therefore the quality of teaching and knowledge acquisition of the primary school pupils in these fields is extremely important for future successful education development of quality experts – engineers who will be able to face business challenges of the modern society in which technical development happens very fast. The first part of work is dedicated to the implementation of the knowledge acquisition in the field of electrotechnics and informatics within the new curriculum and syllabus for primary schools in Serbia [1]. It was shown in which school subjects (courses) and with how many classes the education of the mentioned fields was carried out. In the second part of the paper were shown the results of the researches conducted amongst the primary school pupils with the goal to estimate how satisfied they are with the quality of the acquired knowledge in the field of electronics and informatics which are determined with the current curriculum and syllabus, and also to estimate generally how much they are interested for the further education within these fields. The last part of the paper is dedicated to the abilities, methods and strategies for the improvement of the education quality of pupils in the mentioned fields which are supported by the curriculum for primary schools in Serbia [1-14].

<sup>1</sup>Sonja Cvetkovic is with the Primary school "Cele Kula" Radnih brigada 28, 18000 Nis, Serbia, E-mail: sonjacvetkovic61@gmail.com

<sup>2</sup>Zoran Stankovic is with the Univeristy of Nis, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: zoran.stankovic@gmail.com

## II. IMPLEMENTATION OF ELECTRICAL AND INFORMATICS KNOWLEDGE ACQUISITION IN SERBIAN PRIMARY SCHOOLS

Acquisition of the first terms and basic knowledge from the field of electrotechnics and informatics starts in the first four grades of primary school within the school subjects (courses) *World around us* and *Nature and Society*. In the table I is shown which topics in the field of electro-technics and informatics are taught, within which school subjects and how many classes according to the current curriculum and syllabus for primary schools in Serbia.

TABLE I. ACQUISITION OF THE INITIAL KNOWLEDGE IN THE FIELD OF ELECTROTECHNICS AND INFORMATICS

Grade	Course name	Topics related to eletrotechnics and informatics and teaching hours
1th	World around us	Communication devices (1) Computer and computer components (1)
2th	World around us	Computer and information devices (4)
3th	Nature and society	Electrical properties of materials, Current, Circuit (4)
4th	Nature and society	Electrical and magnetic properties of materials, Electrostatics, Current, Circuit, Experiments in the filed of electrostatics, current, circuit and magnetism (6)

In the upper grades, electrotechnics and informatics topics are mostly covered by the compulsory courses: *Physics* and *Technical and informatics education* and by elective course *Information technology and computer technique* which scores a high percentage in most schools in the Republic of Serbia.

General aim of the *Physics* is for the pupils to familiarise themselves with natural occurrences and principal laws of nature, to get basic scientific literacy, to recognize physical phenomena through research and to apply physics laws in their everyday life and work. Table II shows the curriculum for this course per classes in accordance to the current teaching plans and programmes for primary schools regulated by the Ministry for Education and Science of the Republic of Serbia [1].

TABLE II. PHYSICS - CURRICULUM

Course name: <b>Physics</b>	
<b>6th grade</b>	Course duration in teaching hours: <b>72 per school years</b>
Introduction (2), Kinematics (14), Forces (14), Measurement (15), Mass and density (15), Pressure (12)	
<b>7th grade</b>	Course duration in teaching hours: <b>72 per school years</b>
Forces and motion (25), Friction forces (12), Body balance (11), Mechanical work, energy and power (15) Thermal phenomena (9)	
<b>8th grade</b>	Course duration in teaching hours: <b>68 per school years</b>
Oscillatory motion and waves (8), Liht waves and phenomena (15), Electrical field (10), Current Electricity (19), Magnetic field (6), Basic of atomic and nuclear physics (8), Physics and modern word (2)	

You may notice that topics which in the narrower sense belong to the core field of electro-technics are more present in the 7<sup>th</sup> class. As this is a more comprehensive subject it requires a solid background in mathematics and other fundamental areas of physics. Almost the entire 8<sup>th</sup> class curriculum is dedicated to the electrotechnics (Fig. 1). The teaching plans and programmes include laboratory practical drills and experiments with interactive work with pupils as a support in the process of knowledge acquisition of the relevant fields of physics. Learning quality critically depends on the level of the school teaching aids equipment. With regards to the electro-technics teaching, the physics laboratory at primary school “Cele Kula”, Nis is an example of a well-equipped teaching aids school which include: complete electrostatics experiments, influent machine, mathematical and electrical pendulum, Faraday’s cage, electrical car models, alternator, voltmeter, various resistors and conductors, magnetic field, various magnets, magnetic needle, compass, Ersted’s experiment. This is similar in most Serbian schools. Teachers of physics believe that teaching of physics will be more qualitative, interesting and attractive to pupils if the schools had larger numbers of teaching aids, but teaching aids expansion is limited by the budget which is regulated by the authorised state institutions.

Apart from the regulated number of regular class hours, knowledge in this subject is also gained from supplement class hours aimed at less successful pupils and from complementary class hours in the format of young physicist groups. These groups are aimed at pupils with special interest in physics who would like to promote their high level of acquired knowledge by the means of taking part in competitions which are supported by the Ministry of Education and Science of the Republic of Serbia [2].

TABLE III. TECHNICAL AND INFORMATICS EDUCATION - CURRICULUM

Course name: <b>Technical and informatics education</b>	
<b>5th grade</b>	Course duration in teaching hours: <b>62 per school years</b>
Introduction (2), Traffic (8), Graphic communications (16), From idea to realization (8), Materials and technologies (12), Energy systems (4), Constructive modeling (12)	
<b>6th grade</b>	Course duration in teaching hours: <b>72 per school years</b>
Introduction to the architecture and construction (4), Technical drawing (8), Information technology (16), Building materials (4), Energy systems (4), Technical resources in construction (4), Traffic systems (2), Housing culture (4), Constructive modeling (22), Technical resources in agriculture (4)	
<b>7th grade</b>	Course duration in teaching hours: <b>72 per school years</b>
Introduction to mechanical engineering (2), Technical drawing in mechanical engineering (8), Information technology (14), Materials (2), Measurement and control (2), Materials processing technology (4), Machines and Mechanisms (16), Robotics (2), Energy systems (6), Constructors modeling - Modules (16)	
<b>8th grade</b>	Course duration in teaching hours: <b>68 per school years</b>
Information technology (16), Electrical materials and installation (10), Electrical machines and devices (14), Digital electronics (12) From idea to realization - Modules (16)	

The general goal of the school subject (course) *Technical and informatics education* is to get the pupils acquainted with the technical and technologically developed surrounding through the acquisition of new technical and informatics knowledge and skills, as well as to enable them to apply the

acquired knowledge and skills in the everyday life. In table III is given the curriculum for this school subject signed by the Ministry of Education and Science of Serbia [1]. The school subject *Technical and informatics education* has grown out from the previous school subject *Technical education* which was studied according to the old curriculum. With the new curriculum this school subject leaves more room for studying informatics which is specifically done by introducing some creative elements in the teaching process, which will be more talked about in the section IV [1]. Besides the regular classes, there are also complementary classes held for the pupils who have dispositions for this subject and for those who want to enter the competitions in this field [2].

The goal of the school subject *Information technology (IT) and computer technique* is to enable pupils to use computers, to gain informatics literacy as well as to enable them to apply the use of computers and IT in everyday life and work (Table IV). According to the current curriculum for primary schools, this is an optional subject. Besides the classes which are held as regular classes, there are also complementary classes held for the pupils who have dispositions for this subject and for those who want to enter the competitions in this field [2].

TABLE IV. INFORMATION TECHNOLOGY AND COMPUTER TECHNIQUE – CURRICULUM

Course name: <b>Information Technology and Computer Technique</b>	
<b>5th grade</b>	Course duration in teaching hours: <b>36 per school years</b>
Computer operating system (14), Text editing (14), Introduction to multimedia (8)	
<b>6th grade</b>	Course duration in teaching hours: <b>36 per school years</b>
Text editing (10), Internet (4), Computer graphics (10), Computer animation (3) Elective modules (9)	
<b>7th grade</b>	Course duration in teaching hours: <b>36 per school years</b>
Internet (6), Audio processing (4), Video processing (6), WWW presentation design (10) Elective modules (10)	
<b>8th grade</b>	Course duration in teaching hours: <b>68 per school years</b>
Computer spreadsheets (10), Elective modules (16), Software programming (14), Computer networks and WWW presentations (14), Development of individual projects in the filed of information technology and computer technique (14)	

Figure 1 shows the percentage of hours per subject *Physics* and *Information technology and computer technique*, which is directly related to the study in the field of electrotechnics and informatics. It can be seen that the presence of of these teaching hours increases from lower to higher grades.

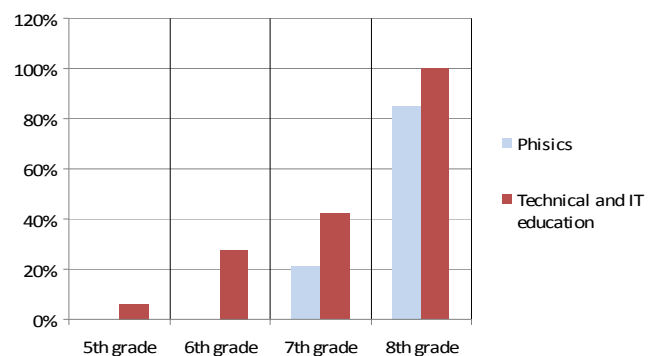


Fig. 1. The percentage of teaching hours in which pupils learn lessons that directly relate to electrotechnics and informatics

Performing quality teaching in the field of IT and Computer Technique requires that each primary school is equipped with specialized IT rooms. This issue will be discussed in more details later in section IV.

### III. SOME SURVEY RESULTS AS AN SUPPORT TO THE EVALUATION OF THE PRIMARY CURRICULUM IN THE FIELD OF ELECTROTECHNICS AND INFORMATICS

There was a survey conducted amongst the primary school pupils in the region of Nish in order to give an approximate evaluation how satisfied they are with the way and quality of the acquired knowledge in the field of electro-technics and informatics which are determined by the current curriculum for primary schools in Serbia, and also to find out how much interested they are in the further education in these fields. There were two schools from the closer urban region of Nish chosen for this survey, "Cele kula" and "Dusan Radovic", and one school from the wider rural region of Nish, "Ivan Goran Kovacic" in Nish Spa. In the survey participated 407 pupils in total from 5th to 8th grade, 218 of them from "Cele kula", 102 from "Dusan Radovic" and 87 pupils from "Ivan Goran Kovacic" (observed according to the grades: V-75 pupils, VI-41 pupils, VII-52 pupils and VII-239 pupils). The pupils answered seven questions which referred to education in the field of electrotechnics and informatics.

In figure 2 are shown the pupils' answers to the question how interesting for them are the teaching units in the field of electrotechnics and informatics. Only a small number of the pupils answered that these units weren't interested which speaks a lot about the pupils' interests for these fields. In figure 3 the pupils answers show that the number of pupils who want to continue to attend the lectures on electrotechnics and informatics is much larger than the number of those who don't want that. There is a prevailing wish for attending IT and informatics. Picture 4 shows how big is the interest of pupils for choosing professions such as engineer of electronics, informatics and computers. At least 10% of the pupils would like that their future profession is engineer of electronics, informatics or computers, and the majority of pupils would like to use the knowledge acquired from the fields mentioned above in their future jobs. Answers in figure 5 show that the majority of the pupils think that there should be more practical work in the field of electrotechnics and informatics. That indicates that current equipment and teaching aids aren't in accordance with the novelties in the curriculum and syllabus. So, the teaching aids should be modernized to enable the good quality practical work of pupils. Figure 6 shows that a significant majority of pupils chooses elective course *IT and computer technique* which again speaks about the popularity of IT and computer science. In figure 7 can be seen that almost one half of the 6th grade pupils attends complementary classes or are members of science (physics) section, while this number significantly declines in the 8th grade. This is explained with the increasing number of pupils' obligations in higher grades. Figure 8 shows that there is an evident interest of pupils for competitions in the field of electrotechnics and informatics, but it should be worked even more on making these competitions more participated by pupils.

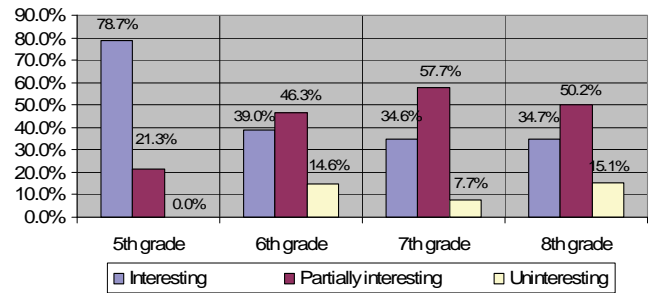


Fig. 2. Pupils' answers to the question of how interesting for them are the teaching units in the field of electrotechnics and informatics

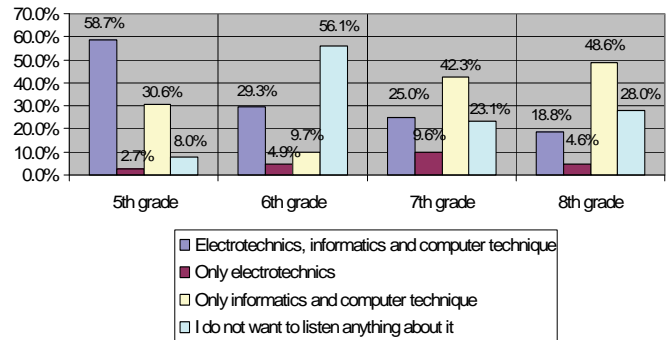


Fig. 3. Pupils' answers to the question of what they want to listen in the courses of future studies

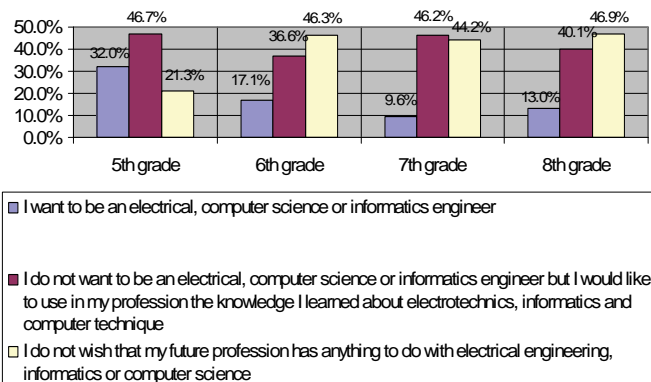


Fig. 4. Pupils' answers to the question of which profession they want to choose after completing their education

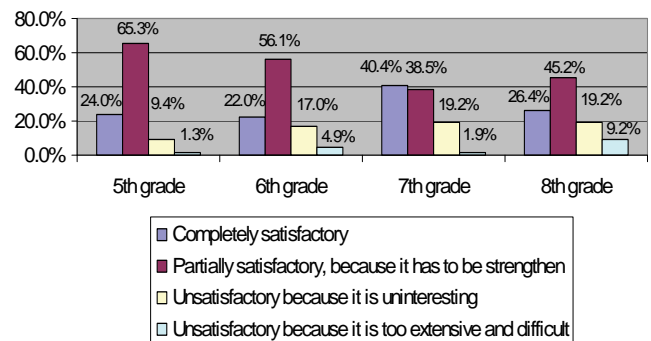


Fig. 5. Pupils' answers to the question how the practical work (work in laboratories and workshops, work with computers, conducting experiments) in the field of electrotechnics, informatics and computer technique is satisfactory

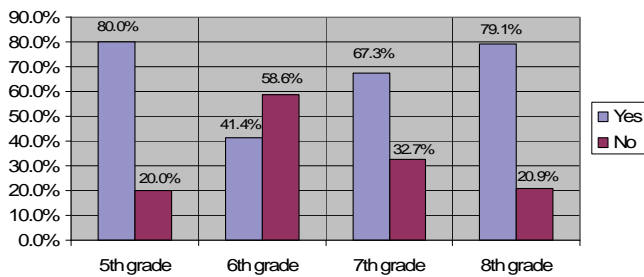


Fig. 6. Percentage of pupils who chose elective course *Information technology and computer technique*

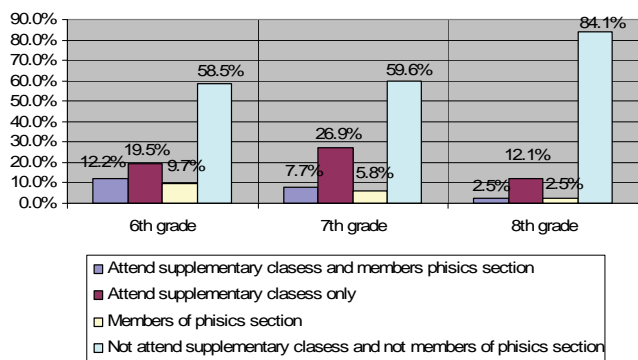


Fig. 7. Percentage of pupils who attend supplementary classes in physics or who are members of physics section

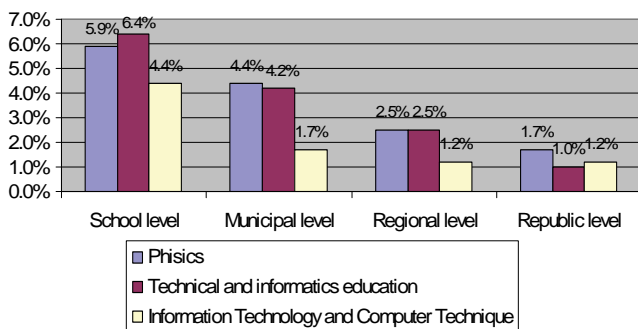


Fig. 8. Percentage of pupils who participated at different levels of competition (overview is given by the subjects and levels of competition)

#### IV. OPPORTUNITIES FOR IMPROVING THE PRIMARY EDUCATION QUALITY IN THE FIELD OF ELECTROTECHNICS AND INFORMATICS

Beside the Ministry of Education and Science of the Republic of Serbia which is the main and responsible institution for implementing of all levels of education, in Republic of Serbia today there are a few more important state institutions whose work have an influence on pupils' primary education. They are [3-5]:

- National education council of the Republic of Serbia
- Center for improving education and schooling
- Center for assessment of education quality and schooling

The work of these institutions is coordinated and they work very closely with each other. These institutions have the

relevant working mechanism which open doors for different possibilities for pupils' primary education improvement in the field of electrotechnics and informatics. Amongst the activities with significant importance we would like to mention the following:

- Harmonizing of educational plans and programmes with the EU standards and modern tendencies in the field of electrotechnics and informatics
- Improving of the teaching aids in the field of electrotechnics and informatics
- Introducing modern information-communication technologies (ICT) to support new methodologies in education
- Professional development of teachers in the field of electrotechnics and informatics through attending regular lessons for continuous professional development and enabling them to participate in professional conferences and seminars
- Organising seminars, lectures and other activities for pupils with an aim to compliment the education in the fields of electrotechnics and informatics and increase their interest in further education in these particular fields
- Improving the organisation and increasing the general technical and educational level of competition in the area of electronics and informatics in order to attract more pupils to participate in these competitions

Today's rapid development of electrotechnics and informatics and their increasing presence in all areas of social living requires continuous updating of primary school curriculum with the latest news in these fields. Modern plans and programmes in high schools across the EU have the fastest dynamics of adapting to those changes. One of the popular solutions to adapt the plans and programmes to changes in primary schools is their coordination with relevant programmes in high schools [6]. National Education Council in the Republic of Serbia has a vital role in continuous monitoring and analyzing of the state of education on all levels and its coordination with the European principles and values. In accordance with this the council makes decisions on changes and coordination of the primary schools plans and programmes in educational areas of interest [3]. Primary schools participation in TEMPUS projects, financed and approved by the EU, is a great opportunity to get help and support for the process of modernizing primary schools plans and programmes through their coordination with high schools programmes [6].

The improvement of quality of education in primary schools is unthinkable without introduction of certain educational standards which are in coordination with the EU standards. During the last few years the Republic of Serbia has been making huge efforts to incorporate and implement such standards in primary education. As a direct result of those efforts, in 2009, National Education Council has passed an act "the educational standards for the end of compulsory education". Those standards originated from working on project by the Ministry of Education and Sport of the Republic of Serbia named: "Development of schooling in the

Republic of Serbia” and its project component “Standard development and assessment” brought to life by the Center for assessment of education quality and schooling [5]. The basic problem that occurred after adopting these standards was the fact that they only covered 10 subjects in primary education. As far as the subjects containing electrotechnics and informatics were concerned these standards included only the physics while excluding other subjects such as technical and informatics education and information technology and computer technique. In the light of this obvious shortage, today there have been efforts to develop adequate standards of achievements for these subjects so they can give a significant contribution in increasing the quality of education in these subjects [7].

Due to its specific nature implementation of high quality education in the area of electrotechnics and informatics is difficult to achieve if primary schools do not have modern teaching aids required for this particular area. Significant number of schools in the Republic of Serbia is facing a problem of using dated and inadequate teaching aids needed to achieve educational targets in the area of electrotechnics. Constant repairing and upgrading of the old equipment and buying of the new one which will be in accordance with new modern curricula, requires continuous funding. This is particularly relevant for the equipment used in laboratory experiments and training. Finding significant financial means in the time of financial crisis and limited and restricted primary schools’ budgets may be an impossible task. One of the ways to resolve this problem, which has been given more attention lately, is to introduce the concept of digital school, where teaching is delivered in the space of digital cabinets (classrooms equipped with computers and other ICT equipment) [9]. According to this concept, some adequate additional interactive multimedia books will be used to support the increase of quality of schooling in physics, technics and informatics [10]. In accordance with that, instead of buying a complex and expensive equipment for practical training in electrotechnics, an adequate and cheaper interactive programme will be used to simulate the same experiment on the computer screen. Using this method will make it possible for larger number of experiments to be done without the need to expand the infrastructure of the laboratories as practical training will be delivered in the virtual lab. The situation is quite similar when it comes to training in the area of informatics, however we will talk about this a bit later in the part relating to the introduction of ICT in primary schools teaching.

At the beginning of 2011 huge support was given to the concept of digital school, introduction of ICT in teaching and increasing the quality of education in primary schools in general by stating the project named “Digital School” which is managed and funded by the Ministry of Telecommunications and Information Society of the Republic of Serbia. The aim of this project is to fully equip 2910 schools (83%) with digital cabinets in the Republic of Serbia [9]. Including March 2012 digital cabinets were fully established in 2808 schools (80%) in the Republic of Serbia. One of the schools with fully equipped digital cabinet is primary school “Cele kula“ in Nis (Fig. 9).



Fig 9. Modern digital cabinet in primary school “Cele kula“ in Nis realised through a project “Digital School” [9]

Implementation of digital cabinets in primary schools is widely opening doors for introduction of information/communication technologies in teaching which is essential for increasing education to a higher level particularly in the area of computing and informatics. Working within computer networks, access to modern internet services, using multimedia and other information/ communication technologies are opening possibilities for implementation of methodical innovations in teaching such as learning based on working on project and problem orientated teaching, also known as “problem solving teaching” [11]. These innovations are very suitable for improving teaching in subjects such as technical and information education and information technology and computer technique because the knowledge required is based on active participation of pupils in solving the real life problems through team work on chosen subjects. On the other side, the implementation of ICT in teaching has reduced the problem of shortage of adequate literature on problem solving teaching as the information is available on the Internet [11]. With the objective to implement internet services to the majority of schools the Ministry of Education and Science of the Republic of Serbia has signed a contract with the one of the biggest internet providers in Serbia, Telecom Serbia, to install ADSL in primary and secondary schools. The contract aims at installing ADSL in all primary schools in Serbia until 2013 with the speed of 16Mbps which is sufficient for the quality internet access.

Primary teachers’ training to deliver schooling based on the new plans and programmes and application of new methods as well as capability of constant improvement of their knowledge with new scientific resources results in their chosen subject is one more important reason that influences the quality of education in electrotechnics and informatics. According to the legal regulation about continuous professional development, teachers’ training qualification, tutor and expert advisor, accepted by the Ministry of Education and Science of the Republic of Serbia, a teacher is legally obliged to attend 100 hours programme over 5 years while at least 60 hours are dedicated to the compulsory and up to 40 hours to elective curriculum [12]. Every year Center for improvement of education and teaching approves a catalogue of programmes for professional development for teachers [4]. Taking into

consideration that the Minister of Education and Science of the Republic of Serbia has given priority to information – communication technologies in the last 3 years, programmes for professional development in the field of technology and informatics are significantly more present in the catalogues and are available to teachers. Yet another important way of teachers' professional development is their participation in professional conferences and seminars. One example of professional conference which is important for primary school teachers' development who are teaching electrotechnics and informatics is International Conference on Technics and Informatics – TIO organised by the Faculty of Technical Science in Cacak [13].

With the aim to make electrotechnics and informatics more popular among primary school pupils, as well as enable them to widen their knowledge in this area, they are given an opportunity to attend professional seminars, summer schools, lectures and workshops. "Summer school of science" hugely popular among primary school pupils in Serbia, supported by UNESCO, takes place in Petnica and it can offer rich and interesting programmes in the field of electronics and computing [14]. Primary school "Cele Kula" in Nis as well as significant number of other schools in Serbia organise a very popular "Science Fair" where numerous experiments and examples of practical teaching are demonstrated to pupils. Also, the numbers of joint workshops for primary and secondary pupils have been increasing in order to motivate the primary school pupils choose some of the electrotechnics modules in their secondary education.

Improvement of organisation and increase of the general technical and educational level of competition in the field of electrotechnics and informatics is yet another way of widening pupils' knowledge which has the impact on the increase of the quality of education in general. Ministry of Education and Science has been making efforts to find additional finances for improving the organisation of competitions and adequately rewarding the achievements in order to increase the number of pupils – competitors. Engaging financial sponsors outside of education can help resolving this problem.

## V. CONCLUSION

New educational plans and programmes for primary schools which are regulated by the Ministry of Education and Science of the Republic of Serbia have been coordinated with modern tendencies in the field of electrotechnics and informatics. They offer acquisition of the latest knowledge in these fields which are developing very fast today and a special attention is paid to the units referring to computers and informatics. A huge interest of pupils for studying and acquiring knowledge from the field of electro-technics and informatics, which is shown in the conducted survey, justifies these changes in the primary schools curriculum. These curriculum and syllabus have established a foundation for acquiring high-quality knowledge in the mentioned fields which represents a beginning of education for successful engineers. The survey shows that a huge number of pupils demand the increase of practical work in the fields of electro-technics and computers, which inevitably leads to the fact that

school aids should be modernized in accordance with the curriculum and syllabus. Serbian Government is making great effort with its ministries to enable such modernization of school aids in primary schools. An example for this is the project "Digital School". As the quality of education depends on how much the teachers are trained to transfer modern knowledge to pupils, special attention is paid to teachers' training in the field of electrotechnics and informatics so that they are constantly involved in special trainings, attending obligatory classes and being able to participate conferences and seminars. Today there are evident efforts being made in order to adopt appropriate educational standards for all school subjects (courses) which are in accordance with EU standards which will lead to the further improvement in quality of primary school education. Special attention is paid to coordination of syllabus and curriculum of primary schools with those in secondary schools and faculties. In the end, it is very important to point out that there are great efforts made to organize competitions, seminars, lectures and other activities for the pupils who want to complement the education in the field of electrotechnics and informatics, which surely leads to an increase in pupils' interest for further education in these fields.

## REFERENCES

- [1] Education Programme for Primary Schools (School year 2011/2012), Serbian Ministry of Education and Science, 2011.
- [2] Competitions and Festivals Calendar for Primary School Pupils (School year 2011/2012), Serbian Ministry of Education and Science, 2011, <http://www.mpn.gov.rs/prosveta/page.php?page=156>.
- [3] National Education Council of the Republic of Serbia <http://www.nps.gov.rs>
- [4] National Institute for Education and Upbringing Improvement, <http://www.zuov.gov.rs/>
- [5] National Institute for Education and Upbringing Quality Evaluation, <http://www.ceo.edu.rs>
- [6] TEMPUS Project JP 516762-2011: "Harmonization and Modernization of the Curriculum for Primary Teacher Education", <http://www.tempus.ac.rs/projects-tempus/view/797/49/>
- [7] Education Standards for End of Compulsory Education, Serbian Ministry of Education and Science, National Institute for Education and Upbringing Quality Evaluation, Belgrade, 2009 <http://www.ceo.edu.rs/images/stories/publikacije/Obrazovni%20standardi%202009.pdf>
- [8] M. Sanader, "Educational Standards of Achievements in the Course on Technology and Informatics in Primary Schools", 3<sup>rd</sup> International Conference on Technics and Informatics in Education - TIO 2010, Conference Proceedings, pp.176-192, Cacak, Serbia, 2010.
- [9] Project "Digital School", Serbian Ministry of Culture, Media and Information Society, Serbian Digital Agenda, <http://www.digitalnaagenda.gov.rs>
- [10] Kvar Media–Multimedia workshop, <http://www.kvarmedia.co.rs>
- [11] M. Paroskaj, M. Mitrovic, " Use of ICT in Modernize and Bringing Up Quality of Technical and Informatical Education in Primary School", 3<sup>rd</sup> International Conference on Technics and Informatics in Education - TIO 2010, Conference Proceedings, pp.468-474, Cacak, Serbia, 2010.
- [12] Rulebook of continuous professional training and acquiring of teacher, educator and professional associate titles, Sl. glasnik RS (br. 14/2004, 56/2005)
- [13] International Conference on Technics and Informatics in Education – TIO, <http://www.tfc.kg.ac.rs/tio2012/>
- [14] Petnica Science Center, <http://www.petnica.rs>



## AUTHOR INDEX

### A

Acevska V. ....	387, 447
Acevski I. ....	387, 447
Acevski N. ....	235
Aćimović S. ....	407
Agatonovic M. ....	311
Aleksandrova M. ....	495, 509
Aleksieva V. ....	439
Alexiev V. ....	89
Altimirski E. ....	37
Andonova A. ....	541
Angelov K. ....	37, 51
Angelov K. ....	85
Angelov P. ....	521, 525
Antolović I. ....	212
Antonov S. ....	375
Apostolov P. ....	117
Aprahamian B. ....	564
Arnaudov R. ....	255
Arsić M. ....	259, 269
Asenov O. ....	359, 428
Atanasov I. ....	327, 418
Atanasovski M. ....	243
Atlagic B. ....	463, 467
Avramova N. ....	403

### B

Babic D. ....	135
Bakardjieva T. ....	167
Balabanov G. ....	89
Balzhiev P. ....	255
Bankov K. ....	71
Banković B. ....	552, 556
Békefi Á. ....	39
Bekjarski A. ....	367, 603
Blagojević D. ....	303
Bock W. ....	255
Bodurov G. ....	509
Bogdanović N. ....	303
Bojchev D. ....	471
Bonev B. ....	25, 37
Brodić D. ....	151, 573, 577
Brusev T. ....	513
Bucsa I. ....	5

### C

Cernat C. ....	175, 273
Cherneva G. ....	371, 493, 571
Cholakova I. ....	499, 509
Čičević S. ....	407
Ćirić D. ....	113, 131
Craciunescu R. ....	1, 5
Cvetkovic S. ....	277
Cvetković T. ....	299

### D

Damjanovic M. ....	163
Demirev V. ....	21
Denić D. ....	259, 269
Denishev K. ....	495
Despotović V. ....	573, 577
Devic S. ....	463, 467
Dimitrijevic T. ....	29
Dimitrov B. ....	544, 548
Dimitrov D. ....	243
Dimitrov K. ....	55, 307, 599, 603
Dimitrov V. ....	283, 567
Dimitrova E. ....	571
Dimkina E. ....	371
Dimov A. ....	485
Dimova R. ....	283
Dimovski T. ....	43
Djosic S. ....	163
Djugova A. ....	59
Djukovic M. ....	463
Dobrev D. ....	307
Dobrikov G. ....	509
Dochev I. ....	595
Docheva L. ....	367, 595
Dochkova-Todorova I. ....	206
Dokoski G. ....	471
Dončov N. ....	299, 311
Đorđević G. ....	105
Đorđević M. ....	212
Drača D. ....	77
Draganov I. ....	143, 155
Draganov V. ....	265
Dzhakov R. ....	121

### E

Eftimov T. ....	255
Eremieva M. ....	159, 263

### F

Fahlberg-Stojanovska L. ....	447
Fehér A. ....	39
Fratu O. ....	1, 5, 93, 273
Furkov G. ....	232

### G

Gadjeva E. ....	228, 505
Gajić D. ....	216, 190
Gajic M. ....	463
Galabov M. ....	435
Gechev M. ....	51
Georgiev M. ....	71
Georgieva T. ....	343
Georgieva V. ....	247
Gerasimov K. ....	239
Gesheva K. ....	509
Gjorgjievska S. ....	194

Goleva R. ....	89
Gorecan Z. ....	463, 467
Gradinarova B. ....	167
Guliashki V. ....	208

## **H**

Hadjidimitrov A. ....	247
Halunga S. ....	1, 5
Hristov G. ....	351, 355

## **I**

Ilarionov R. ....	451
Iliev G. ....	323
Iliev I. ....	47, 319, 321
Iliev Ivo ....	502
Iliev M. ....	351, 355
Ivanov H. ....	55
Ivanov H. ....	435
Ivanova E. ....	97

## **J**

Jakimovska D. ....	194
Jakimovski G. ....	194
Jakšić B. ....	315
Janković D. ....	178
Janković M. ....	113
Janković S. ....	407
Jankulovska M. ....	387
Jelenković M. ....	131
Jevtic M. ....	163
Jokovic J. ....	29, 299
Jordanova L. ....	599, 307
Jovanovic Z. ....	331

## **K**

Kanev J. ....	74
Karailiev H. ....	451
Karailiev V. ....	451
Karova M. ....	403, 475
Katsov R. ....	493
Kirilov L. ....	208
Kitov C. ....	347
Kolev G. ....	509, 495
Kolev N. ....	33
Kolev S. ....	319, 321
Koleva E. ....	159, 263
Koleva P. ....	359
Kopta A. ....	295
Kostić V. ....	552, 556
Kostov N. ....	395, 479, 482
Kotevski A. ....	411
Kovačević M. ....	198, 220
Kovacheva M. ....	224
Krastev G. ....	422
Krasteva I. ....	485

## **L**

Lehtinen V. ....	135
------------------	-----

Lozanovska A. ....	171
Lubich L. ....	287
Lukić J. ....	259, 269

## **M**

Malecic A. ....	585
Malenko M. ....	194
Manev S. ....	319, 321
Manojlović P. ....	295
Marinov A. ....	548
Marinova G. ....	479, 482
Markova G. ....	399
Markova V. ....	265
Markovic D. ....	331
Marković I. ....	178
Martinovic L. ....	467
Matijasevic J. ....	431
Mićić Z. ....	295
Mihajlović V. ....	212, 220
Mihic D. ....	212
Mihov G. ....	383
Mihov Y. ....	101
Mijoski K. ....	235
Mijoski T. ....	235
Mikarovski G. ....	411
Mikhov M. ....	251
Milenković A. ....	178
Milić D. ....	315, 335
Milijic M. ....	291
Milinković S. ....	407
Militaru T. ....	175, 273
Milivojević D. ....	151, 573, 577
Milivojević Z. ....	151
Miljković G. ....	259, 269
Miljković V. ....	577
Milosavljević A. ....	220
Milosavljevic S. ....	335
Milovanović B. ....	29, 299, 311
Milovanović D. ....	303
Milovanovic I. ....	291
Milutinović V. ....	299
Mironov R. ....	139, 603
Mirtchev S. ....	89
Mitrevski P. ....	43
Mitrović N. ....	552, 556
Mitrović S. ....	407
Mitsev T. ....	25, 33, 55, 599
Mladenović S. ....	407

## **N**

Nagy L. ....	59
Nagy S. ....	39
Necov B. ....	71
Nedelchev M. ....	47
Nedelkovski I. ....	387
Nenov I. ....	493

Nenova M.....	339
Nikolić B. ....	105
Nikolov B. ....	395, 482
Nikolov G.....	529, 544,581
Nikolov N.....	425
Nikolova B. ....	513, 529, 581
Nikolova K.....	125
Nikolova M. ....	159, 263
Novakov P.....	283

## O

Obradović D.....	295
------------------	-----

## P

Pacheco C.....	63
Pacheco de Carvalho J.....	63
Panagiev O.....	9, 13
Panajotović A.....	77
Pandiev I.....	224, 517
Pargovski J.....	171
Pavlov G. ....	371, 493
Pavlov M.....	573, 577
Pavlović N.....	407
Penchev P.....	109
Pencheva E. ....	418
Penev I.....	403, 475
Perić D.....	81
Perić M. ....	81
Pesovic U.....	331
Petkov E.....	459
Petkov P. ....	239
Petkova Y.....	403
Petković M.....	105
Petronijević M. ....	552, 556
Petrov A.....	525
Petrova-Antonova D.....	485
Petrović M.....	315
Pleshkova-Bekiarska S.....	375
Poenaru V.....	175, 273
Popova A. ....	147
Popović M.....	81
Poulkov V.....	359
Predić B. ....	198

## R

Radev D.....	97
Radic J. ....	59
Radmanović M.....	216
Radonov R. ....	593
Raev R. ....	97
Rančić D.....	212, 220, 431
Randjic S.....	331
Rankovska V. ....	489
Reis A. ....	63
Ristić A.....	560

Roganović M.....	198
Ruzin I.....	171

## S

Sabeva V. ....	159, 263
Sadinov S. ....	74, 85
Sechkova T.....	143
Sekulović N. ....	77
Serafimov N.....	513
Shtarbakov V.....	564
Shupak M.....	25
Simeonov I.....	121
Simić M.....	259
Sirakov E. ....	129
Sit L.....	311
Slavov M. ....	109
Smiljakovic V.....	533
Spalević P.....	315
Spalevic Z.....	431
Spasova V. ....	502
Stančić I. ....	295
Stanimirovic A.....	202
Stankovic A. ....	186
Stanković D. ....	182, 186
Stanković R. ....	190
Stankovic Z.....	277, 291, 311
Stavru S.....	485
Stefanov T.....	455
Stefanova M.....	428
Stefanović Č. ....	182
Stefanovic D.....	335
Stefanovic H.....	335
Stoimenov L.....	202
Stojanović D.....	198
Stojanović M.....	560
Stoyanov G.....	125
Stratev A.....	232
Streblau M.....	564
Suciu G.....	175, 273
Šunjevarić M.....	81

## T

Tabakov S.....	502
Tahrilov H. ....	544, 548, 564
Tasić D. ....	560
Tasić V. ....	573, 577
Tatić D. ....	182, 186
Temelkovski I.....	315
Tentov A. ....	194, 471
Todoran G.....	175, 273
Todorov M.....	529
Todorov V.....	265
Todorova M.....	425
Todorović B.....	81
Tomic D.....	463, 467

Tomić S.....	131
Tomov Y. ....	323
Trifonov T. ....	121
Trifonov V.....	418
Trifonova T. ....	265
Trpezanovski L. ....	243
Tsankov B. ....	101
Tsenov A.....	67
Tsvetkova I.....	351, 355
Tarpov I. ....	493

## **V**

Valchanov H.....	443
Valkov G. ....	228
Varbanova N.....	85
Vasić B.....	105
Veiga H. ....	63

Videkov V.....	232, 593
Videnovic-Misic M.....	59
Vulović D.....	202
Vulpe A. ....	93

## **Y**

Yordanov H. ....	17
Yordanova S.....	395, 479, 482

## **Z**

Zahariev P. ....	351, 355
Zdravković J. ....	131
Zhilevski M.....	251
Zhivomirov H.....	129, 379
Živanović D.....	269
Zivanovic Z. ....	533
Zwick T.....	311