

---

**51st INTERNATIONAL SCIENTIFIC CONFERENCE ON  
INFORMATION, COMMUNICATION  
AND ENERGY SYSTEMS AND TECHNOLOGIES**

---



## **Proceedings of Papers**

---

Organized by



Faculty of Technical Sciences, Bitola, Macedonia



Faculty of Electronic Engineering, Niš, Serbia



Faculty of Communications and Communication Technologies, Sofia,  
Bulgaria

Under auspices of

Macedonian Ministry of Education and Science

In cooperation with:

Faculty of Tourism and Organizational Sciences – Ohrid

**28 - 30 June 2016, Ohrid, Macedonia**

ICEST 2016 - 51st INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION, COMMUNICATION AND ENERGY SYSTEMS AND TECHNOLOGIES - 28 - 30 June 2016, Ohrid, Macedonia

Proceedings of Papers

Editor: Prof. Dr. Cvetko Mitrovski

Technical Editor: Jove Pargovski

Published by: Faculty of Technical Sciences, Bitola, Macedonia

Number of copies printed: 30

ISBN-10 9989-786-78-X

ISBN-13 978-9989-786-78-5

EAN 9789989786785

CIP - Каталогизација во публикација

Национална и универзитетска библиотека "Св. Климент Охридски", Скопје

004(062)

620.9(062)

621.3(062)

INTERNATIONAL scientific conference on information, communication and energy systems and technologies (51 ; 2016 ; Ohrid)

ICEST 2016 : proceedings of papers / 51st International Scientific Conference on information, communication and energy systems and technologies, 28-30 June 2016, Ohrid, Macedonia ; [editor Cvetko Mitrovski]. - Bitola : Faculty of technical sciences, 2016. - 488 стр. : илустр. ; 30 см

Фусноти кон текстот. - Библиографија кон трудовите. - Регистри

ISBN 978-9989-786-78-5

1. Mitrovski, Cvetko [уредник]

а) Информатика - Собири б) Енергетика - Собири в) Електротехника - Собири

COBISS.MK-ID 101838346

**Printing of this edition has been financially supported by EVN Macedonia**





Dear Colleagues,

The 51st International Scientific Conference on Information, Communication and Energy Systems and Technology ICEST 2016, was held from 28 to 30 June at the Faculty of Tourism and Organizational Sciences - Ohrid, Macedonia.

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

As to the previous ICEST Conferences, we have papers and authors from institutions all over the world. This year we have 237 authors (and coauthors) from 10 countries with 113 scientific and application oriented papers for oral and poster presentations.

I hope that all participants had taken opportunities not only to exchange their knowledge, experiences and ideas but also to make contacts and establish further collaboration.

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

On the behalf of the Technical Program Committee, I hope that we will meet again at the next ICEST Conference.

On the behalf of the Technical Program Committee,

**Prof. D-r Cvetko Mitrovski**  
Conference Chairman



## TECHNICAL PROGRAM COMMITTEE

### GENERAL CHAIRMAN:

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

### VICE CHAIRMEN:

B. Milovanović University of Niš, Serbia

K. Dimitrov Technical University of Sofia, Bulgaria

### MEMBERS OF 2016 PROGRAM COMMITTEE

Acevski N.,	University "St. Kliment Ohridski", Bitola, Macedonia
Andreevski I.,	University "St. Kliment Ohridski", Bitola, Macedonia
Arapinovski B.,	University "St. Kliment Ohridski", Bitola, Macedonia
Atanasov I.,	Technical University of Sofia, Bulgaria
Atanasovski M.,	University "St. Kliment Ohridski", Bitola, Macedonia
Bekiarski A.,	Technical University of Sofia, Bulgaria
Bogdanov M.,	University "Ss Cyril and Methodius", Skopje, Macedonia
Boumbarov O.,	Technical University of Sofia, Bulgaria
Češelkoska V.,	University "St. Kliment Ohridski", Bitola, Macedonia
Denić D.,	University of Niš, Serbia
Dimcev V.,	University "Ss Cyril and Methodius", Skopje, Macedonia
Dobrev D.,	Technical University of Sofia, Bulgaria
Dochev I.,	Technical University of Sofia, Bulgaria
Dokić B.	University of Banja Luka, Bosnia and Herzegovina
Dončov N.,	University of Niš, Serbia
Eftimov T.	Plovdiv University "Paisii Hilendarski", Bulgaria
Georgieva V.,	Technical University of Sofia, Bulgaria
Iliev G.,	Technical University of Sofia, Bulgaria
Iliev I.,	Technical University of Sofia, Bulgaria
Jakšić Z.	IHTM Institute, Belgrade, Serbia
Janković D.,	University of Niš, Serbia
Jeftić M.,	University of Niš, Serbia
Jevremović A.,	Singidunum University, Belgrade, Serbia
Jordanova L.,	Technical University of Sofia, Bulgaria
Jovanović Z.,	University of Niš, Serbia
Kostov M.,	University "St. Kliment Ohridski" Bitola, Macedonia
Kountchev R.,	Technical University of Sofia, Bulgaria
Lutovac M.	Singidunum University, Belgrade, Serbia
Makal J.	Tech. University of Byalistok, Poland
Mančić D.,	University of Niš, Serbia
Marinova G.,	Technical University of Sofia, Bulgaria
Marković V.,	University of Niš, Serbia
Miletiev R.,	Technical University of Sofia, Bulgaria
Mirchev S.,	Technical University of Sofia, Bulgaria
Mitsev T.,	Technical University of Sofia, Bulgaria
Mojsovska N.,	University "St. Kliment Ohridski", Bitola, Macedonia

Nakamatsu K.	University of Hyogo, Japan
Nikolov T.,	Technical University of Sofia, Bulgaria
Nikolova B.,	Technical University of Sofia, Bulgaria
Nikolova Z.,	Technical University of Sofia, Bulgaria
Nusev S.,	University "St. Kliment Ohridski", Bitola, Macedonia
Pantić D.,	University of Niš, Serbia
Pencheva E.,	Technical University of Sofia, Bulgaria
Petković P.,	University of Niš, Serbia
Petrović P.	IRITEL Institute, Belgrade, Serbia
Pleshkova S.,	Technical University of Sofia, Bulgaria
Poulkov V.,	Technical University of Sofia, Bulgaria
Radevska P. M.,	University "St. Kliment Ohridski", Bitola, Macedonia
Radoičić V.	University of Belgrade, Serbia
Spalević P.	Faculty of Technical Sciences, K. Mitrovica, Serbia
Spirovski M.,	University "St. Kliment Ohridski", Bitola, Macedonia
Stanković R.,	University of Niš, Serbia
Stanković Z.,	University of Niš, Serbia
Stefanovski B.,	University "St. Kliment Ohridski", Bitola, Macedonia
Stoimenov L.,	University of Niš, Serbia
Stojanović D.,	University of Niš, Serbia
Stojčev M.,	University of Niš, Serbia
Stoyanov G.,	Technical University of Sofia, Bulgaria
Todorov G.,	St. Cyril and St. Methodius University of Veliko Tarnovo, Bulgaria
Todorova M.	St. Cyril and St. Methodius University of Veliko Tarnovo, Bulgaria
Trpezanovski Lj. ,	University "St. Kliment Ohridski", Bitola, Macedonia
Tsankov B.,	Technical University of Sofia, Bulgaria
Tsenov A.,	Technical University of Sofia, Bulgaria
Veinović M.,	Singidunum University, Belgrade, Serbia
Veselić B.,	University of Niš, Serbia
Valkova-Jarvis Z.	Technical University of Sofia, Bulgaria
Zieleznik L.	Brookes University of Oxford, UK

## CONFERENCE ORGANIZING COMMITTEE

### CHAIRMAN:

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

### INTERNATIONAL COORDINATORS:

Milovanović B., University of Niš, Serbia

Dimitrov K., Technical University of Sofia, Bulgaria

### LOCAL COORDINATOR

Pargovski J., University "St. Kliment Ohridski" Bitola, Macedonia

### MEMBERS:

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

Dimitrijević T., University of Niš, Serbia

Dimitrov K., Technical University of Sofia, Bulgaria

Dochev I., Technical University of Sofia, Bulgaria

Goleva R., Technical University of Sofia, Bulgaria

Iliev G., Technical University of Sofia, Bulgaria

Koleva P., Technical University of Sofia, Bulgaria

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

Milijić M., University of Niš, Serbia

Stanković Z., University of Niš, Serbia

Tsankova J., Technical University of Sofia, Bulgaria

## CONFERENCE SECRETARIAT

### ICEST 2016

Faculty of Technical Sciences

Ivo Lola Ribar bb, 7000 Bitola, Macedonia

phone: +389 47 207 711

e-mail: [info@icestconf.org](mailto:info@icestconf.org)

## UPDATED INFORMATION

Updated information can be obtained from the official

Conference web site:

<http://www.icestconf.org>

## **ICEST HISTORY**

The ICEST Conference appears to succeed a series of conferences started from 1963 under the name “Day of the Radio”. In 1977 the name of the Conference was changed into “Communication, Electronic and Computer Systems”.

Since 2000 it has become an international conference under the name EIST (Energy and Information Systems and Technologies). The first two EIST Conferences were successfully organized by the Faculty of Communications and Communication Technologies Sofia and the Faculty of Technical Sciences Bitola.

In 2002, for the first time, the Faculty of Electronic Engineering Niš has accompanied the Conference organizers. Also, the Conference changed its name becoming ICEST (International Scientific Conference on Information, Communication and Energy Systems and Technologies).

After two Conferences organized in Niš 2014, and Sofia 2015, the Faculty of Technical Sciences Bitola was host of ICEST 2016 Conference.

## LIST OF ICEST 2016 REVIEWERS

Aleksandar Atanasković  
University of Niš, Serbia

Aleksandra Panajotović  
University of Niš, Serbia

Anelia Doseva  
Technical University of Sofia, Bulgaria

Angel Colov  
Technical University of Sofia, Bulgaria

Boban Veselić  
University of Niš, Serbia

Cvetko Mitrovski  
University “St Kliment Ohridski” - Bitola,  
Macedonia

Dimitar Dimitrov  
University “St. Kiril I Metodij” – Skopje,  
Macedonia

Dimiter Dimitrov  
Technical University of Sofia, Bulgaria

Dušan Milošević  
University of Niš, Serbia

Emil Manolov  
Technical University of Sofia, Bulgaria

Evelina Pencheva  
Technical University of Sofia, Bulgaria

Georgi Tsenov  
Technical University of Sofia, Bulgaria

Hristomir Yordanov  
Technical University of Sofia, Bulgaria

Ivo Dochev  
Technical University of Sofia, Bulgaria

Ivo Draganov  
Technical University of Sofia, Bulgaria

Jovan Stefanovski  
JP Strezevo-Bitola, Macedonia

Julian Velchev  
Technical University of Sofia, Bulgaria

Kalin Dimitrov  
Technical University of Sofia, Bulgaria

Kamelia Nikolova  
Technical University of Sofia, Bulgaria

Liljana Docheva  
Technical University of Sofia, Bulgaria

Lupco Trpezanovski  
University “St Kliment Ohridski” - Bitola,  
Macedonia

Marin Nedelchev  
Technical University of Sofia, Bulgaria

Milan Plazinić  
University of Kragujevac - FTN Čačak, Serbia

Miroslav Lutovac  
Singidunum University, Serbia

Mitko Kostov  
University “St Kliment Ohridski” – Bitola,  
Macedonia

Nebojsa Doncov  
University of Niš, Serbia

Nebojša Mitrović  
University of Kragujevac - FTN Čačak, Serbia

Oleg Panagiev  
Technical University of Sofia, Bulgaria

Predrag Eferica  
University of Niš, Serbia

Rumen Mironov  
Technical University of Sofia, Bulgaria

Seferin Mirtchev  
Technical University of Sofia, Bulgaria

Simeon Tsvetanov  
Technical University of Sofia, Bulgaria

Siniša Randzić  
University of Kragujevac-FTN Čačak, Serbia

Slobodan Dzukić  
University of Kragujevac- FTN Čačak, Serbia

Stefan Panic  
University of Pristina, Serbia

Stojce Deskovski  
University “St. Kliment Ohridski” - Bitola,  
Macedonia

Uroš Pešović  
University of Kragujevac- FTN Čačak, Serbia

Vera Markovic  
University of Niš, Serbia

Veska Georgieva  
Technical University of Sofia, Bulgaria

Zlatka Valkova  
Technical University of Sofia, Bulgaria

Zoran Jovanović  
University of Niš, Serbia

Zoran Stankovic  
University of Niš, Serbia



# TABLE OF CONTENTS

## PLENARY SESSION

<b>Initial Steps in Development of Computer Science in Macedonia and Pioneering Contribution to Human-Robot Interface using Signals Generated by a Human Head .....</b>	<b>3</b>
Stevo Bozinovski	

## RADIO COMMUNICATIONS, MICROWAVES, ANTENNAS

<b>Coupled Halfwave Microstrip Resonators with Dumbbell Type Defected Ground Structure .....</b>	<b>13</b>
Marin Nedelchev, Alexander Kolev	
<b>Effect of RF Receiver Nonlinearity on Mobile Telecommunication Systems.....</b>	<b>17</b>
Youssef Abd Alrahman and Ilia Iliev	
<b>Omnidirectional Double Biquad Omni-Antenna for 2.4 GHz Wireless Link Application.....</b>	<b>21</b>
Kalina Kalinovska and Peter Z. Petkov	
<b>Synthesis of Wideband Edge Coupled Microstrip Filters with Defected Ground Structures.....</b>	<b>25</b>
Marin Nedelchev, Alexander Kolev, Ilia Iliev	
<b>Modified Fractal Super J-pole Antenna .....</b>	<b>29</b>
Boncho Bonev, Peter Petkov and Nina Spassova	
<b>Experimental Determination of 2biQUAD Antenna Radiation Pattern Using CRM Method.....</b>	<b>33</b>
Kalina Kalinovska and Kliment Angelov	
<b>Infrared investigation on the thermal field in the case of influence of low frequency magnetic signals on live tissue.....</b>	<b>37</b>
Kalin Dimitrov	
<b>Statistical Characteristics of Optimal Divergence of Laser Beam at FSO Systems ...</b>	<b>41</b>
Tsvetan Mitsev	
<b>ANN Approach for Return Loss Analysis of Ohmic Series RF MEMS Switches.....</b>	<b>45</b>
Tomislav Ćirić, Zlatica Marinković, Marija Milijić, Olivera Pronić-Rančić, Vera Marković, Larissa Vietzorreck	
<b>Calculation of Capacitance of the Rectangular Coaxial Lines with Offset Inner Conductor by Strong FEM.....</b>	<b>49</b>
Vladimir V. Petrovic and Žaklina J. Mančić	
<b>Design and Analysis of Realistic Vehicle Traces Model based on the Evolutionary Algorithms .....</b>	<b>53</b>
Danijel Čabarkapa and Petar Pavlović	
<b>Transmitting coherent carrier frequencies in modern CATV/HFC systems.....</b>	<b>57</b>
Oleg Borisov Panagiev	
<b>Control of Photon Energy Sharing for Photovoltaic Power Charge in Power Grid of Optic Fiber With Enhanced Effectiveness .....</b>	<b>61</b>
Jovan Shikoski, Juli Zlatev, Tinko Eftimov, Georgi Stanchev	
<b>Polynomial-Based Extraction Procedure for Determination of HEMT Noise Wave Temperatures .....</b>	<b>65</b>
Vladica Đorđević, Zlatica Marinković, Olivera Pronić-Rančić and Vera Marković	
<b>Study the impact of the instability of oscillators in the Head End modules.....</b>	<b>69</b>
Oleg Borisov Panagiev	
<b>Mixer Linearization in Direct Conversion Receiver.....</b>	<b>73</b>
Aleksandar Atanasković, Aleksandra Đorić and Nataša Maleš-Ilić	

## TELECOMMUNICATION SYSTEMS AND TECHNOLOGY

<b>Novel Transport protocol for 5G Networks .....</b>	<b>79</b>
Ivan Petrov and Toni Janevski	
<b>OP of SC Diversity for OFDM-based WLAN System Operating in Microcell and Picocell Environment.....</b>	<b>83</b>
Aleksandra Panajotović, Dragan Drača and Nikola Sekulović	
<b>Chromatic Polynomial Heuristics for Connectivity Prediction in Wireless Sensor Networks .....</b>	<b>87</b>
Vladislav Vasilev	
<b>Hybrid Environment Service Orchestration for Fog Computing .....</b>	<b>91</b>
Stojan Kitanov and Toni Janevski	
<b>Transformation of the existing operator in the service provider for business customers .....</b>	<b>95</b>
Zlata Radak and Nataša Gospić	
<b>2D Localization of Source of Stochastic EM Radiation by using Neural Networks...</b>	<b>99</b>
Zoran Stankovic, Ivan Milovanovic, Nebojsa Doncov, Bratislav Milovanovic	
<b>Analysis and evaluation of the human factor in the communication and information systems .....</b>	<b>103</b>
Evgeniy Hubenov, Antonio Andonov, Zoya Chiflidzhanova	
<b>Software development of Content Management System in Online-CADCOM platform.....</b>	<b>107</b>
Ognyan Chikov	
<b>Investigation of algorithms for VoIP signaling .....</b>	<b>111</b>
Todorka Georgieva, Ekaterina Dimitrova and Slava Yordanova	
<b>Contemporary Trends for Increasing the Reliability of the SCADA systems Communication Level.....</b>	<b>115</b>
Emiliya Dimitrova and Vasil Dimitrov	
<b>Transmission of multimedia data in Wireless Sensor Networks .....</b>	<b>119</b>
Miloš Kosanovic and Mirko Kosanovic	
<b>Performance analysis of FSO transmission over M atmospheric turbulence channels.....</b>	<b>123</b>
Petar Spalević, Bojan Prlinčević, Stefan Panić, Mladen Veinović, Aleksandra Kokić Arsić	
<b>Performance Analysis of Wireless Macrodiversity SSC Receiver in Gamma-Shadowed nLoS Fading Channel .....</b>	<b>127</b>
Caslav Stefanovic	
<b>Outage Performance of Satellite-Terrestrial Multiuser Networks with Fixed Gain AF Relay.....</b>	<b>131</b>
Aleksandra M. Cvetkovic, Jelena A. Anastasov and Nemanja M. Zdravkovic	
<b>Responsive Web Design in Online-CADCOM platform .....</b>	<b>135</b>
Ognyan Chikov	

## SIGNAL PROCESSING

<b>Statistical Analysis of Round-off Noise in First-Order IIR DSP Systems .....</b>	<b>141</b>
Zlatka Valkova-Jarvis, Kamelia Nikolova and Venera Dimitrova	
<b>Novel CIC FIR Filter Functions: Design and Influence of Added Filter Blocks .....</b>	<b>145</b>
Biljana P. Stošić, Vlastimir D. Pavlović	
<b>Model of the Operational Conveyor based on Currentsteering output stage .....</b>	<b>149</b>
Slobodan Djukic and Milan Veskovc	

<b>The Evaluation of Speech Intelligibility in the Orthodox Church on the Basis of MOS Test Intelligibility Logatom Type CCV .....</b>	<b>153</b>
Dijana Kostić, Zoran Milivojević and Violeta Stojanović	
<b>Research and signal analysis with different classes of wavelets .....</b>	<b>157</b>
Mariana Todorova and Reneta Ivanova	
<b>Optimization of the parameter of 2P Keys interpolation kernel according to the criterion of similarity to sinc function in the spectral domain.....</b>	<b>161</b>
Nataša Savić, Zoran Milivojević, Zoran Veličković, Darko Brodić	

## DIGITAL IMAGE PROCESSING

<b>Kernel based Extreme Learning Machines for Image Classification .....</b>	<b>167</b>
Stevica S. Cvetković, Miloš B. Stojanović, Saša V. Nikolić, Goran Z. Stančić	
<b>New interactive interface for visualization of 3D meshes.....</b>	<b>171</b>
Nicole Christoff, Hadrien Greef, Sébastien Gormond-Mazet, Jean-Luc Mari, Laurent Jorda and Agata Manolova	
<b>Comparison of Complex Hadamard Transform with other Unitary Transforms .....</b>	<b>175</b>
Rumen P. Mironov	
<b>Some Finger Vein Databases Evaluation .....</b>	<b>179</b>
Ivo Draganov, Darko Brodić	
<b>Tracking Dynamics Objects on Binocular Vision System.....</b>	<b>183</b>
Rosen Spirov, Neli Grancharova and Svilen Soyanov	
<b>An Approach for Renal Vessels Evaluation in US Images.....</b>	<b>187</b>
Veska Georgieva, Stephan Vassilev	
<b>The Insertion of the Encrypted Low-Resolution Watermark in the Uncompressed Video.....</b>	<b>191</b>
Zoran Veličković, Zoran Milivojević and Marko Veličković	
<b>Face Recognition using LBP and PCA under Different Lighting Condition .....</b>	<b>195</b>
Dipak P. Patil, Pallavi. S. Sonar and Svetlin I. Antonov	

## COMPUTER SYSTEMS AND INTERNET TECHNOLOGIES

<b>Formal Verification of Models for M2M Device Registration.....</b>	<b>201</b>
Evelina Pencheva, Ivaylo Atanasov and Anastas Nikolov	
<b>Autonomous Agent Model for Connectivity Management in M2M Communications.....</b>	<b>205</b>
Ivaylo Atanasov, Anastas Nikolov and Evelina Pencheva	
<b>Integrated HPC Data-log Monitoring System.....</b>	<b>209</b>
Kostadin Mishev, Sonja Filiposka, and Anastas Mishev	
<b>E-content Development and Task Solution Using the Content Management System of Online-CADCOM.....</b>	<b>213</b>
Galia Marinova and Ognyan Chikov	
<b>SCADA system for monitoring and control of small industrial producing unit.....</b>	<b>217</b>
Hristo Nenov	
<b>Project of Crochet Software Development .....</b>	<b>221</b>
Elena Zaharieva-Stoyanova	
<b>Graph Construction Algorithm for finding the Shortest Path in a Maze.....</b>	<b>225</b>
Milena Karova, Ivaylo Penev, Mariana Todorova, Hristiyan Bobev, Neli Kalcheva	

## INFORMATICS AND COMPUTER SCIENCE

<b>A Hybrid Metaheuristic Algorithm for Flexible Job Shop Scheduling “PSO&amp;TS” ..</b>	231
Asen Tochev and Vassil Guliashki	
<b>NoTree.....</b>	235
Iliya Tronkov	
<b>Random Generation of Bent Functions on Multicore CPU Platform.....</b>	239
Miloš Radmanović and Radomir S. Stanković	
<b>Education of Patients by using Mobile Learning System.....</b>	243
Dusan Tatić and Dragan Janković	
<b>Versatile Program Model of Avatar Face.....</b>	247
Milan Adamović, Dragan Ivetić, Aleksandar Kupusinac	
<b>Hadoop – Next Generation of Business Intelligence Systems .....</b>	251
Borivoje Milosevic, Srdjan Jovkovic and Marko Jankovic	
<b>Formal Methods for Data Analysis .....</b>	255
Blagovest Kirilov	
<b>An Approach in Spatial Databases.....</b>	259
Mariana Stoeva and Violeta Bozhikova	
<b>Studying the process of software development through intensive use of software tools.....</b>	263
Violeta Bozhikova, Mariana Stoeva and Dimitrichka Nikolaeva	
<b>Statistical Analysis of Text Encryption Software.....</b>	267
Darko Brodić, Ivo R. Draganov and Zoran N. Milivojević	

## ELECTRONICS

<b>Analytical Modeling of a ZVS Bidirectional Series Resonant DC-DC Converter... .</b>	273
Aleksandar Vuchev, Nikolay Bankov, Yassen Madankov and Angel Lichev	
<b>Load and Control Characteristics of a ZVS Bidirectional Series Resonant DC-DC Converter .....</b>	277
Aleksandar Vuchev, Nikolay Bankov, Yassen Madankov and Angel Lichev	
<b>Two-Phase Switching-Mode Converter with Zero Voltage Switching Designed on CMOS 0.35 <math>\mu</math>m Technology .....</b>	281
Tihomir Brusev, Georgi Kunov and Elissaveta Gadjeva	
<b>Converters with Energy Dosing for Charging of EV’s Li-ion Batteries.....</b>	285
Nikolay Madzharov and Goran Goranov	
<b>Mathematical model of thermoelectric Peltier module .....</b>	289
Liliya Staneva, Ivaylo Belovski, Anatoliy Alexandrov and Pavlik Rahnev	
<b>Mathematical model of control system based on programmable environment.....</b>	293
Anatolii Aleksandrov, Goran Goranov, Pavel Hubenov	
<b>Electronic Module for 2D Positioner Manuel Control .....</b>	297
Nikola Draganov, Totka Draganova and Dragan Draganov	
<b>Study of Nonlinear Effects in Parallel Gyrator Resonance Circuits.....</b>	301
Atanas Tanev, Ivan Uzunov, Marin Hristov, Sotir Ouzounov	
<b>Problems and Properties of a Current Amplifier When Realized in Ultra Deep Sub-micron Technology.....</b>	305
Simeon Kostadinov, Ivan Uzunov, Dobromir Gaydazhiev	
<b>Optimization and analysis ALL Digital PLL For resonant inverter.....</b>	309
Goran Goranov, Nikolay Madzharov	
<b>Design and Realization of a 12W Quasi-Resonant Flyback Converter .....</b>	313
Zoran Zivanovic and Vladimir Smiljakovic	

<b>Problems in assessing the reliability of electronic components and systems by reliability testing.....</b>	317
Toncho Papanchev, Anton Georgiev and Nikolay Nikolov	
<b>Indices for Reliability Assessment of a Star Structured Complex Electronic System .....</b>	321
Nikolay Nikolov, Anton Georgiev and Toncho Papanchev	
<b>Using Wireless Interfaces in a Smart Home Model.....</b>	325
Valentina Rankovska	

## ENERGY SYSTEMS AND EFFICIENCY

<b>Determination of Optimal Modules Number in Photovoltaic Strings for Inverter Power Maximization.....</b>	331
Ljupco Trpezanovski, Dimitar Dimitrov	
<b>Transient Stability of Asynchronous Generator on Distribution Network.....</b>	335
Metodija Atanasovski, Mitko Kostov, Nikola Acevski, Blagoj Arapinoski, Elena Kotevska	
<b>Economic Energy Scheduling of an Islanded Microgrid.....</b>	339
Galia Marinova and Vassil Guliashki	
<b>Assessing the Profitability of New Investment Projects on Construction of Small HPP on HPS Crn Drim.....</b>	343
Goce Bozinovski and Atanas Iliev	
<b>Determination of Harmonics Level in Local Electrical Distribution System.....</b>	347
Ginko Georgiev, Silviya Letskovska, Kamen Seymenliyski and Pavlik Rahnev	
<b>Model study of Lighting Protection of 110/20 kV Substation.....</b>	351
Margreta Vasileva, Neli Velikova and Danail Stanchev	
<b>Calculation of parameters of Three-Phase Induction Motor with Double Squirrel Cage with FEM 3D.....</b>	355
Blagoja Arapinoski, Mirka Popnikolova Radevska, Vesna Ceselkoska, Mile Spirovski and Metodija Atanasovski	
<b>Electromagnetic analysis on Three – Phase Induction Motor whit Cage with FEM 2D.....</b>	359
Nikolce Talevski, Mirka Popnikolova Radevska, Blagoja Arapinoski, Mile Spirovski	
<b>Single-Line Ground Fault's Impedance Impact to Electrical Parameters in Ungrounded Electrical System .....</b>	363
Milica Rašić, Milan Radić and Zoran Stajić	
<b>Influence of Induction Generator's Windings Resistance Variation on Optimal CVT Regulation Curves .....</b>	367
Milan Radić, Milica Rasić and Zoran Stajić	
<b>Some properties of electrochromic low vacuum evaporated copper(I) oxide thin films .....</b>	371
Ratka Neshkovska, Mimoza Ristova and Julijana Velevska	
<b>FTIR spectra of electrochromic copper(I) oxide thin films prepared by different methods .....</b>	375
Ratka Neshkovska, Mimoza Ristova and Julijana Velevska	

## CONTROL SYSTEMS

<b>Multicomponent Analysis of Basic Physico-chemical Parameters of Bulgarian Yoghurt .....</b>	381
Mariyana Sestrimska, Tanya Titova, Veselin Nachev and Chavdar Damyanov	

<b>Comparative analysis of more commonly used recursive methods for parameter estimation in adaptive systems.....</b>	385
Ivan V. Grigorov, Nasko R. Atanasov	
<b>Mobile Robot Guidance in Presence of Obstacles, using the Potential Fields Method .....</b>	389
Vesna Antoska-Knights, Zoran Gacovski, Stojce Deskovski	
<b>Method and device for registration and analysis of car crashes at smart traffic lights. Hardware model. ....</b>	393
Rosen Vitanov and Georgi Stanchev	
<b>Method and device for registration and analysis of car crashes at smart traffic lights. Software model.....</b>	397
Rosen Vitanov and Rumen Arnaudov	
<b>Control-Command and Signalling Subsystems – Testing Framework within the APIS Process.....</b>	401
Denitsa Kireva-Mihova	
<b>Application of Automatic Landing using GPS-GBAS.....</b>	405
Olivera Petrovska, Vesna Antoska Knights and Stojche Deskovski	

## MEASUREMENT SCIENCE AND TECHNOLOGY

<b>Evaluation of Wines Based on Mathematical and Statistical Analysis.....</b>	411
Krastena Nikolova, Ivan Maslinkov, Irena Ivanova	
<b>Tablet Computers and Extremely Low-Frequency Magnetic Field Emission.....</b>	415
Darko Brodić, Milena Jevtić, Jordan Radosavljević and Ivo R. Draganov	
<b>Functional and Timing Analysis of Improved Serial Pseudorandom/Natural Code Converter .....</b>	419
Goran Miljković, Dragan Denić, Milan Simić and Aleksandar Jocić	
<b>Review and Contribution to the Best Measurement and Calibration Capabilities in the Field of Electrical Energy and Power.....</b>	423
Marija Cundeva-Blajer and Bojan Iliev	
<b>Analysis of the noise parameters of quad structure of MEMS linear accelerometers .....</b>	427
Emil Iontchev, Rosen Miletiev, Vladimir Bashev and Rumen Yordanov	
<b>Analysis of the noise parameters of quad structure of MEMS angular rate sensors .</b>	431
Rosen Miletiev, Emil Iontchev, Vladimir Bashev and Rumen Yordanov	
<b>Laboratory Performance Studies of WPA IEEE 802.11a PTP Links.....</b>	435
José A. R. Pacheco de Carvalho, Cláudia F. F. P. Ribeiro Pacheco, Hugo Veiga, António D. Reis	

## REMOTE ECOLOGICAL MONITORING

<b>Frequency characteristics and impedance as a criterion for purity by the monitoring of soil .....</b>	441
Svetlana Gerganova-Savova, Daniela Toneva-Zheynova and Yanko Yanev	
<b>Client-Server System for Improving Ecological Conditions of the Environment .....</b>	445
Dijana Jagodić, Dejan Vujičić, Siniša Randić	

## ENGINEERING EDUCATION

<b>An Overview of the Possibilities of the Variational Analysis for Exploration of Electric Circuits.....</b>	451
Emil Panov and Miroslava Doneva	

<b>Laboratory Equipment for Energy Accumulation from Renewable Sources.....</b>	<b>455</b>
Silviya Letskovska, Kamen Seymenliyski, Ginko Georgiev	
<b>Project Based Learning .....</b>	<b>459</b>
Danica Milosevic	
<b>Collaborative Learning in Higher and Secondary Education .....</b>	<b>463</b>
Tania Vasileva, Stela Stefanova and Vassiliy Tchoumatchenko	
<b>Laboratory stand for experiments in real time with applications in training on control non-linear systems.....</b>	<b>467</b>
Veselin Nachev, Tanya Titova and Plamen Nikovski	
<b>Using PLC for Control on Asynchronous Drives – Laboratory Simulator .....</b>	<b>471</b>
Vasil Dimitrov and Emiliya Dimitrova	
<b>Importance of Learning to Code, Coding to Learn in Higher Education .....</b>	<b>475</b>
Anton J. Knierzinger and Boyka Gradinarova	
<b>Infrared investigation on the thermal field of the human face during the EEG session.....</b>	<b>479</b>
Kalin Dimitrov, Stanio Kolev, Iglika Andonova, Tsvetan Mitsev	
<b>Contemporary Approach to Teaching GSM with USRP and OpenBTS.....</b>	<b>483</b>
Slavica Marinković, Amela Zeković, Ivan Pavlović and Milutin Nešić	





# Plenary Session



# Initial Steps in Development of Computer Science in Macedonia and Pioneering Contribution to Human-Robot Interface using Signals Generated by a Human Head

Stevo Bozinovski <sup>2</sup>

**Abstract** – This plenary keynote paper gives overview of the pioneering steps in the development of computer science in Macedonia between 1959 and 1988, period of the first 30 years. As part of it, the paper describes the worldwide pioneering results achieved in Macedonia. Emphasize is on Human-Robot Interface using signals generated by a human head.

**Keywords** – pioneering results, Computer Science in Macedonia

## I. INTRODUCTION

This paper represents a research work on the history of development of Computer Science in Macedonia from 1959 till 1988, the first thirty years period. The history can be presented in many ways and in this paper the measure of historical events are *pioneering steps*. The paper lists pioneering steps in development of Computer Science in Macedonia, and points out some of them which are worldwide pioneering results.

Many colleagues were directly contacted during the writing of this report; however the responsibility for this report is solely on the author. So it can be viewed as a personal view, although the author made every effort this report to be an objective presentation of the pioneering events of the development of Computer Sciences in Macedonia.

## II. 1959 -1967: PIONEERING STEPS BY PEOPLE FROM MACEDONIA, DONE OUTSIDE MACEDONIA

### A. 1959- People employed in nuclear energy research in institutes outside Macedonia

In the period 1959-1988 which is covered by this report, Macedonia was a state of the Yugoslavia federation. There were 6 states: Bosnia and Herzegovina, Croatia, Macedonia, Montenegro, Serbia, and Slovenia.

The development of science during 1950's placed significant importance on nuclear energy. At that time the principal institutes in Yugoslavia such as Vinca in Belgrade, Rudjer Boskovic in Zagreb, and Josef Stefan in Ljubljana

were engaged in support of nuclear energy research. People from Macedonia interested in advanced research and study were part of this effort. Most significant work was done by Aleksandar Hrisoho, who working together with Branko Soucek in Zagreb, wrote a paper about magnetic core memories in 1959 [1]. He received formal recognition for this work, the award "Nikola Tesla" [2]. He also worked on analog-to-digital conversion on which he wrote a paper [3], received a PhD [4], and was cited [5]. So it can be said that authors from Macedonia in early 1960's already achieved PhD and citation for work related to computer sciences. Their work was in languages other than Macedonian because they worked outside Macedonia.

### B. 1961-1967: People from Macedonia visiting programming courses outside Macedonia

During 1950's and 1960's there was advancement in computer architecture worldwide and European industry was part of it. First computers appeared which were programmed in machine language and in initial form of assembler language. People from Macedonia took part in participating in seminars and courses in computer programming. In 1961 Dimitar Bitrakov followed the course on machine language programming in Ljubljana. In 1964 he followed another course in Belgrade, on machine language named Autocode III. Both used European computers as models of computer architecture. Other Macedonian scientist following the courses offered on programming at that time was Dimitra Karcicka.

## III. 1968: START OF COMPUTER SCIENCE IN MACEDONIA

### A. 1964- 1968: Start of Computer Science in Macedonia

The event which influenced development of Computer Science in Macedonia happened in Soviet Union, where in 1964 Glushkov published a computer science book named Introduction to Cybernetics [6]. It was translated into English in 1966 [7] and in Serbian in 1967 [8]. Here is a list of the book chapters:

Ch1. Abstract theory of algorithms; Ch2. Boolean functions and propositional calculus; Ch3. Automata theory; Ch4. Self-organizing systems; Ch5. Electronic digital computers and programming; Ch6. The predicate calculus and the automation of the processes of scientific discovery.

As can be seen that was actually an advanced book in Computer Science and Artificial Intelligence. The most

<sup>1</sup>Stevo Bozinovski is with the Department of Mathematics and Computer Science, South Carolina State University, USA and with the Faculty of Computer Sciences and Engineering Technologies, Sts Cyril and Methodius University, Macedonia, sbozinovski@scsu.edu

advanced chapter was Chapter 4, on self-organizing systems where Glushkov presented theory of neural networks, pattern recognition, and learning in perceptrons.

In 1967 that book appeared at a Book Fair in Skopje, at two people relevant for this story bought it. One was Dr Georgi Cupona, who was one of the initiators for establishment of Mathematical Institute with Numeric Center (MINC) in 1966, the first institution in Macedonia which in its title has computational aspects of mathematics. The second person, independently, was the author of this paper, at that time a high school student. Professor Cupona immediately recognized the importance of the book. First he suggested to his spouse Gorica Ilieva, to propose to the author of this text to work on the third chapter of the book (Automata Theory), as a high school maturation research project. Being a regular participant in mathematics competition in Macedonia and Yugoslavia, the author of this text was known to Dr. Cupona. In the beginning of 1968, Cupona decided that this book will be basis of a seminar, named Seminar on Cybernetics organized by MINC. Nine lectures were assigned to university professors from Mathematics, Biology, and technical departments. The tenth lecture, on abstract automata, Cupona assigned to the author of this text. At that time no university professor was working on abstract automata. So in May 1968, the author of this text delivered the first lecture on Abstract Automata, in front of university professors and other interested audience, and also wrote, with help of Dr Cupona, the first paper on Computer Science written in Macedonian language [9][10]. This event can be recognized as a starting point in development of Computer Science in Macedonia.

#### *B. 1965 – 1967: IBM and the first computer in Macedonia*

In 1965 IBM started production of its “scientific minicomputer” IBM 1130. With a price tag of \$30,000-40,000 IBM started a marketing offensive to offer that computer to United Nations and governments of many countries. Through the United Nations aid for Skopje 1963 earthquake, the first IBM 1130 was installed in Macedonia in 1967. The configuration contained magnetic core memory of 8K (16 bit words) with 512K disk space. Operating system was Disk Monitor System which was on disk, along with compilers and loaders. Programs were input from punched cards. The computer came as part of the UNESCO support, for development of the Institute for Earthquake Engineering and Engineering Seismology (IZIIS). The Computing Center was established, with the first Director being Dimitar Bitrakov.

The education on single-task Fortran programming in Macedonia started a year before, in 1966. It was delivered by Dr. Masanori Izumi, from Japan, in English language as part of the graduate study carried out by the Institute of Earthquake Engineering and Engineering Seismology (IZIIS). One of the early students and user of the computer was Dimce Petrovski.

It should be noted that in both Belgrade and Zagreb such computer came a year later, in 1968. It should also be noted that IBM 1130 was a revolution in computer science in Europe, not just an evolution. It brought a new philosophy: High level programming languages such as Fortran, made

some previous effort in Europe in developing computers irrelevant. In 1967 Macedonia was in the mainstream computer programming, a step ahead of other Yugoslavia states.

It can be said that Fortran programming and use of IBM1130 computer was initiated by IBM, both in Macedonia and worldwide. As opposite to that, the Computer Science itself was self-initiated in Macedonia, by a vision of Dr Cupona, and the author of this text as executor of his idea.

## **IV. PIONEERING STEPS 1969-1988**

### *A. 1968-1973: First generation of Computer Science majors in Yugoslavia and influence on development of computer science in Macedonia*

The trend of following Nuclear Sciences in 1950's and early 1960's has changed during 1960's toward interest in electronics. There were students enrolling in such programs in Belgrade, Zagreb, and Ljubljana. However, the University of Zagreb was the first to formally start a computer sciences program in 1970.

In 1968 students from Macedonia enrolled the Electrical Engineering Department in Zagreb (ETF Zagreb). ETF Zagreb was organized in units named “zavods” as opposite to “cathedras” which at that time were usual university units. A “zavod” gives a sense of orientation toward industry oriented work, in addition to study and teaching work (usually assumed for a cathedra). One of the zavods, on initiative of professor Stanko Turk, in 1970 was established as “Zavod za Racunarsku Tehniku i Informatiku”), which was about development of computer sciences. ETF Zagreb was the first university department in Yugoslavia which formally established such a computer science program (original “Racunaska tehnika i informatika”). So the mentioned three students from Macedonia were in the first generation of students in Yugoslavia that obtained a formal degree in computer sciences. Among few students of that first generation who obtained highest grade “excellent”, one was from Macedonia. Both the computer science major and concept of zavod, influenced the later development of computer sciences in Macedonia in at least two ways.

First was the name which was given later to the unit of the Electrical Engineering Department in Skopje, Macedonia, to be “Zavod za kompjuterska tehnika i informatika” after the same name from the Electrical Engineering department in Zagreb. The name for the major was “Kompjuterska tehnika I informatika” The names were originally proposed by the author of this paper on 1976..

Second was the influence of the book by professor Branko Soucek published in USA in 1972 named “Minicomputers in data processing and simulation” [11], which gave very advanced knowledge on computer science especially on computer interfacing and process control. From today's perspective we can say that it was most advanced book in computer science by an author from Yugoslavia at that time, and the undergraduate students in ETF Zagreb were able to

listen to lectures by the author of that book. The educational model of a computer used in the book was Digital Computer Corporation PDP 8-minicomputer. The book is about process computers, dealing with interfacing to a process and processing signals. In 1976 Soucek published his second book in USA, named "Microprocessors and minicomputers" [12]. That book also starts with PDP-8 as a model computer, but then goes toward microprocessors rather than toward process control. Students following first generations of master's and doctoral program at ETF Zagreb, among them the author of this text, were able to learn from that book.

*B. 1969- : First courses and book on Fortran programming at undergraduate level*

Although Fortran programming was taught in Macedonia since 1966 by foreign instructors, the undergraduate courses taught by Macedonian instructors started in 1969 on both Mathematics department (taught by Dimitra Karcicka) and Electro-Mechanics department (taught by professor Pane Vidincev). During early 1970's the first book on Fortran programming was published by Dimitar Bitrakov [13] in Macedonian language.

*C. 1970-: First employment in a computer science company*

During 1970's students from Yugoslavia were able to work in foreign countries. Two students from ETF Zagreb, Branko Doljak and Stevo Bozinovski enrolled the educational program of IBM Deutschland. Doljak worked in the plant in Mainz and Bozinovski in the sales unit in Bayreuth. The geschäftsstelle (GS) Bayreuth was selling computers and had system engineering (SE) support. Bozinovski was "Werkstident -IBM Mitarbeiter" and part of the SE support of GS Bayreuth. The education was in IBM educational centers in Dusseldorf, Hannover, Stuttgart, and Munich. The knowledge gained was on operating systems for IBM 1130, IBM System/3 and IBM 380/135, as well programming languages Fortran, Cobol, RPG, and PL/I.

Many other specializations abroad in computer science, especially in the USA, took place during 1970's and after that.

*D. 1970-: The first computer programs on Artificial Intelligence as part of development of Computer Sciences in Macedonia*

The author of this paper wrote his first Artificial Intelligence (AI) program in 1970 and it was a chess endgame two kings and a rock. Given any position with those three pieces on the chess table, computer program would recognize the position correctness and will play until it wins.

*E. 1971-: Recognition of a student work: First- of-May prize*

In 1971 the author proposed a work for First-of-May award of the University of Zagreb, about neural networks named perceptrons and teaching for pattern recognition [14]. A program was written simulating complete perceptron, with

connections from retina to associative units as well as connections from associative to decision units. The university prize was obtained delivered by the University Rector Dr. Ivan Supek

*F. 1971- : Innovation in digital electronics*

During early 1970's students from Macedonia were studying electronics in centers outside Macedonia, because there was not such a study in Macedonia. Although many students just followed the known solutions, there were original solutions proposed. One of them was proposed by the author of this text. While the previous circuits from Macedonian authors were in area of A/D conversion, this is the first original pure digital circuit, proposed by a Macedonian author. It is a simplest *controllable flip-flop, which is built only of 4 inverters*. It was known that two inverters can keep a binary variable, but to control it, until 1971 only the 2-NAND and 2-NOR solutions were known. The inverters-only flip-flop idea was gradually developed considering possibilities of Wired-AND connections of the 1971 Resistor-Transistor Logic (RTL) integrated technology. This 4-inverters flip-flop was used for the first time in 1972 in a work on a digital integrated circuit for simulation of conditioned reflexes [15]. The simulated neural network was also published in a journal [16] as well as in the book of professor Santic [17] who was mentor of this project. More details of obtaining this flip-flop are given in [18].

*G. 1973-: Stipend for talented students*

Following the first generation of its computer science students, who were graduating during 1972 and 1973, in March 1973 Electrical Engineering department in Zagreb opened a master's degree study in Computer Science.

In 1973 the author of this text obtained the newly established stipend for young talents from the Macedonian state government, for master's degree study at the University of Zagreb. The talent had to be shown in various areas and the evidence included mathematics competitions, best high school student recognition, first prize for poetry obtained in Bitola, and prize for drawing competition obtained in Skopje.

*H. 1976- : First educational materials in computer sciences in addition to Fortran programming materials*

From 1966 till 1976, the first 10 years of learning in Fortran programming, there were various written teaching materials in Fortran programming, including the books mentioned above.

During 1970's reorganization activities were carried out also at the Electro-Mechanical Department. In 1970 the major in Electronics was established. Part of that program was a course of Logic Circuits which was initially taught by professor Borivoje Lazic, from University of Belgrade. Among various department cathedras, the Cathedra of Cybernetics was established in 1972 under the leadership of Pane Vidincev. In 1977 the Department became part of the University Center of Mathematics and Technical Sciences



(UCMTN) with slightly different name, “Electrical and Mechanical Department”. The Institute of Electronics was established with Milco Kocare as director. In 1976 the author of this paper joined the Cathedra of Cybernetics at the time of university centers reorganization, and become assistant for teaching Logic Circuits and Organization of Digital Computers. The instructor of Logic Circuits that year was Borivoje Lazic from Belgrade.

The first booklet on computer science not related to Fortran programming was written by the author of this work in 1976 [19] and was about Operating Systems and Systems Programming. Basically it contained knowledge the author gained during his work in IBM in Germany of IBM 370/135 systems. The booklet was made available to the students enrolled in the course on Organization of Digital Computers which in that year was taught by Tome Bosevski, a nuclear engineer returning from Vinca, as proposed by Aksenti Grnarov, lecturer of the Cathedra for Cybernetics, who originally introduced this course.

In 1977 a booklet on logic circuits and computer programming was prepared [20] for the Mathematical School, organized by MINC. The material of the booklet was basically following the book [11].

In 1977 the booklet on digital integrated circuits [21] was prepared showing digital integrated circuits available from Electronic Industry (Ei) from Nis, Yugoslavia. It contained gates, combinational circuits, sequential circuits, memories, and interface modules.

#### *I. 1977- : The name “Informatika” appears in courses and institutions*

In 1977 the Medical Department introduced the course named “Medical Informatics” in its graduate program. The previously existing course of named “Cybernetics” was renamed “Biomedical Cybernetics”. In such way in 1977 the word “Informatics” formally became part of the higher education in Macedonia. Let us mention that three people were involved, Vanco Kovacev from Medical Department, Branko Trpenovski from Mathematics Department, and the author of this paper, who proposed the name of the new course and was first instructor of an Informatics course.

In the same year the name “informatika” appeared as a name of an institution. The Institute of Applied Mathematics and Informatics was established at the Department of Mathematics. As mentioned before the first formal institution dealing with computer sciences was Mathematical Institute with Numeric Center, established in 1966. The author of this paper joined MINC in 1974, after submitting his Master’s Thesis at the Electrical Engineering department in Zagreb. MINC was truly research and application institution, and was not connected to teaching, although some of the employees were engaged in teaching various university courses. In 1976 university went into transformation into University centers, and one of the centers was University center for mathematical and technical sciences (UCMTN). It that reorganization MINC was dissolved and become part of the Mathematical department, where an institution was formed in 1977 named Institute for Applied Mathematics and Informatics.

#### *J. 1977- : Realization of hardware devices for external customers*

In 1976 an idea for development of hardware devices for customers was presented by the author of this paper to professor Pane Vidincev, head of the Cathedra for Cybernetics. The idea was accepted and the laboratory part for digital circuits development was built. At the beginning only this lab had only a professional lab table and a movable arm lamp. A space was assigned in the room where the Cathedra of Cybernetics had its analog computer. The Cathedra had two lab rooms, one with digital computer IBM1130 and the other with analog computer MEDA.

The first customer was Institute of Physiology in 1977. The device delivered was named “Programmer for Reaction-time meter” and was the first digital controller built in Macedonia and the first hardware product ever delivered to a customer by the Electrical Engineering Department. It was built with small scale integration elements using abstract automata principles. The student engaged in building the first controller was Mite Gjorgjeski. Several controllers were built for external customers, and the new technology used was firmware. The firmware based controllers built in 1979 used chips 74188 PROM and 2708 EPROM.

#### *K. 1978- : First lab space dedicated to digital circuits for external customers*

In 1978 the first lab space was dedicated to development of digital devices for external customers, by placing a dedicated lab desk. Before that, the students interested in work with digital circuits had to be guests in labs of other Cathedras. This lab space was built inside the lab for Computing Machines and Automatics which was previously developed by prof. Vidincev, as part of Cathedra for Cybernetics. The lab space was later expanded to two lab desks and it was active till 1986 when it was moved in a new building, where a new lab was formed named Laboratory of Intelligent Machines, Bioinformation Systems, and Systems Software (LIMBISS).

#### *L. 1978- : First symposium related to computer sciences*

In 1978 the first symposium was organized related to computer science. It was named Biomedical Cybernetics and covered applications of computers in medicine among other topics. Participants were from Yugoslavia. The organizers were Stevo Bozinovski and Metodija Kamilovski from Electrical Engineering Department, and Dragan Mihajlov and Margita Kon-Popovska from Mathematics Institute. A symposium proceedings was published and was available at the time of the symposium.

#### *M. 1978- : First professional society in Macedonia related to computer sciences*

The first society related to computer sciences was Biocybernetics Society of Macedonia, formed in 1978, during the mentioned symposium. The president was professor

Dimitar Hrisoho from Medical department, brother of the aforementioned Aleksandar Hrisoho. The Society organized seminars and symposia, among other events.

*N. 1979- : First patent filed, as computer science development*

In 1979 first patent was filed with the Yugoslav patent institution by the author of this paper. It was about programmable controller. The patent was later approved. It was the first patent from Macedonia related to computer science.

*O. 1979- : First microprocessor in Macedonia*

The first microprocessor based application was developed by the author of this work in 1979 during the stay with the Institute for Information and Communication (Nachrichtentechnik und Informatik) of the Technical University in Munich. It was a system for generating square wave for a signal generator and was based on Texas Instruments TMS 9900 microprocessor [22].

The same year the first microprocessor Intel 8008 purchase in Munich and was brought to the Electrical Engineering Department and was introduced in the education process by the author of this text. In 1979 a booklet was prepared on microprocessors and microcomputers [23] containing single clock step micro-programming as well as interfacing.

*P. 1979- : Developing curriculum for a major in computer science*

In 1979 Electrical Engineering Department had a discussion for some time on future reorganization of the Department. The Cathedra of Cybernetics had a discussion on having own major with name Computer Technique, Informatics, and Automatics (orig "nasoka Kopjuterka Tehnika, Informatika i Avtomatika"). As propose by the author of this paper the organization unit for development of computer science would be named "Zavod za Kompjuterska Tehnika i Informatika". The basic curriculum for computer science was developed by Grnarov and Bozinovski. In principle Grnarov developed all the hardware oriented courses, and Bozinovski software oriented courses. Exception was the course on microprocessors and minicomputers" which was developed by Bozinovski.

In 1979 Mathematics and Natural Sciences Department developed a major in "Mathematics-Informatics". The Mathematics Department organized also specialization (one year program) entitled Informatics and Application of Computing Machines.

*Q. 1979- : Spreading computer science in Bitola, Macedonia*

In 1979 the author of this text was elected a faculty in the rank of university Lecturer, of the Technical Department in Bitola, Macedonia. The course was named "Numerical Methods and Digital Computers" and was taught at the

Mechanical Engineering program of the department. The engagement was allowed by prof. Dimitar Bitrakov from Mathematical Department in Skopje, and prof. Mihail Bombol from Technical Department in Bitola. The course was taught in Spring part of 1980 and had 20 students.

Later in 1985 the author of this work was elected as adjunct associate professor at the Technical Department in Bitola to teach the course on "Traffic Transport Cybernetics" which was part of transportation program. The course was taught during 1986. The course assistant was Cvetko Mitrovski.

*R. 1980- : The first industrial fair presentation of Institute of Electronics products*

In 1980 the author of this text presented the hardware and software products at the International Fair on Contemporary Medicine. There were 180 companies presenting their products and Institute of Electronics was one of them. The following hardware products were presented: stabilized voltage source, biostimulator, biopotential amplifier, biopotential pre-amplifier, tapping counter, reactionmeter RE79, universal 8 channel microcontroller, PROM-programmer, Programmer for reaction-time meter RE79. Software products consisted of statistical packages and two diagnostic packages for Cornel and MMPI psychiatry tests. The invitation flyer was prepared by three organizations: Institute of Electronics of the Electrical Engineering Department, Biocybernetics Society of Macedonia, and Association of health organizations of Macedonia. A catalog of the products was prepared for the booth visitors [24].

In 1980, after the fair, for continuous activities related to presentation of Electrical Engineering Department, the highest body of the Department, the Department Board ("Fakultetski Sovet") proposed Bozinovski for a state recognition "11 October" after a formal procedure, with expert committee recommendation, then Faculty Board ("Nastavnicki Sovet") recommendation, finally Department Board approval. Later other people applied for this recognition but using different procedures, not involving Faculty Board and the Department Board.

*S. 1981- : The first membership in an international professional society*

In 1981 Bozinovski become a member of the Institute of Electrical and Electronics Engineers (IEEE) an international society which covers computer sciences, and remained as member (Senior Life Member) till today. The same year he became a member of the New York Academy of Sciences, after he has been invited to join.

*T. 1982-1983: The first upgrade of an IBM computer with own hardware and software*

The first extension of a computer system capability using own hardware-and-software interface was done on the IBM Series/1 computer. It was installed at the Mechanical

Engineering Department, and it was a demand for using that computer as a process computer. It was rack-based and modular, with a unit for handling analog inputs and outputs. However it was not designed as laboratory computer, it had no interface panel toward its analog and digital ports. A project was funded by the Mechanical Engineering Department for building interface and software so that the computer can be used as laboratory computer for signal processing and process control. The people involved were Dragi Danev, Vladimir Dukovski, Ljuben Dudevski, Stevo Bozinovski, and the student Vladan Popovic. The project was completed in 1982 as hardware and in 1983 as software support for signal processing. After that the computer IBM Series/1 become the first process computer used by faculty and students from both Electrical Engineering and Mechanical Engineering departments.

*U. 1982 -: The first multitasking (multithreading) software*

In 1982 the first multi-task (parallel) programming was written and introduced as a way of programming, in the course of Systems Software. The multi-task (or multi-thread) programming was a capability of the Event Driven Language (EDL) which was part of the Event Driven Executive (EDX) operating system for the IBM Series/1 computer. It was noticed that the standard flow-charts used in single-task programming are not applicable, so an original representation was developed [25][26]/

*V. 1983 -: The first multitasking software architecture and the first Macedonian robot*

The first multi-task programming application was for the first Macedonian Robot [21]. It was built out of toy car to which three touch sensors, one voltage sensor, and one timer were added. The robot was named Adriel-1 (ADaptive Robot of the Institute of ELectronics). The demonstration was the known artificial intelligence task: search in a maze-like environment following the walls until it find a "door", actually a voltage source that the voltage sensor detects. The abstract automaton graph was used as method of designing the robot behavior. This multitasking software architecture of a robot was a worldwide pioneering achievement.

*W. 1984 -: The first government grant related to computer sciences*

In 1984 the first state grant was obtained for a computer science related project, named Macro-project Adaptive Industrial Robots. The institution providing state grants at that time was Macedonian Association for Scientific Activities (orig. "Zaednica za naucni dejnosti na Makedonija"). The policy at that time was funding macro-projects involving more than one institution and collaborating with industry.

The project built mostly vehicle robots. It had two research directions: Signal Processing Robotics, and Flexible Manufacturing Systems. After completion of the project in

1988, it had an extension with name Adaptive Intelligent Industrial Robots, till 1992.

*X. 1985 -: The first paper published in a journal of the series IEEE Transactions*

In 1985 the first paper from computer sciences in Macedonia appeared in a world-wide prominent journal, from the IEEE Transaction series, named IEEE Transactions on Systems, Man, and Cybernetics [27]. The paper was about teaching neural networks and was result of the author's work on his PhD Thesis.

At this point a macedonian author published a paper in a high ranking journal. Although it was a pioneering step and a milestone for development of computer science in Macedonia, the paper by itself was not a pioneering step in worldwide science. It did not open a new road in science. It was rather an incremental step, a contribution, to the efforts started by McCulloch and Pitts [28], who first opened a road toward artificial neural networks, and Rosenblatt [29] who first opened a road toward learning in neural networks.

*Y. 1985 -: The first Editor of a journal*

In 1984 Macedonian Society for Biocybernetics organized its thirds symposium, with title Biological and Non-biological Beings. The author of this paper was invited by the Journal Automatika to be the Guest Editor of a special issue, [30], dedicated to the mentioned symposium. The Introduction written by the Guest editor, established first connection between genetic engineering a robotics.

*Z. 1986 -: The first description of a speech control of a robot*

In 1986 the first result of signal processing robotics was achieved, the control of a robot using speech commands. The first language used for talking to a robot was Japanese, and that result was achieved in Japan [31]. However the Japanese group did not give description of the signal processing performed, because it was not the main goal of that group. Their goal was to build the first humanoid robot, named Wabot-1 and they did it. The first description of a speech processing for robot control was given by the project AIR in 1986 [32] and was published in 1989 [33]. It should be emphasized that Macedonian language is the second language which robots understood, after Japanese language.

*AA. 1987 -: Application of Artificial Intelligence in Operating Systems tuning*

Artificial intelligence has been important part of development of Computer Science in Macedonia since 1968 when the first contact with concepts of neural networks and pattern recognition has been made through the book [6]. First program was a chess endgame program and after that series of programs were written for pattern classification and learning in neural networks (e.g. [14]).



Tuning of operating systems has been of interest since 1984 [34]. Most advanced application has been made in 1987 when a IBM 4331 VM/SP operating system has been adaptively tuned for performance improvement [35].

*AB. 1988 -: The most important result of Computer Science in Macedonia: Controlling a robot using EEG signals*

In 1988 project AIR achieved to control a robot using human EEG signals [36][37][38]. Later it become evident that it was the only such result in 20<sup>th</sup> century. The second result of EEG control was achieved 11 years later, in 1999, but with an animal, a rat, using very invasive procedure, with electrodes inside the rat brain. This is arguably the greatest result of Macedonian computer science so far, this paper being written in 2016. Let us mention that the members of the team m besides the author of this work were Liljana Bozinovska, neurophysiologist, and Mihail Sestakov, a student.

*AC. 1988 -: A worldwide pioneering result: Brain-Computer Interface using CNV potentials*

In 1973 Vidal stated a challenge which became known ad Brain-Computer Interface challenge [39]. The challenge was to control external objects using biosignals. Before that challenge usually biofeedback method was used, where internal parameters such a heart rate was controlled. Usually a visual feedback was used. In the 1973 challenge Vidal gave a special attention to the Contingent Negative Variation (CNV) signal.

The first brain-computer interface using CNV potential was achieved in 1988 in Macedonia [40][41].

*AD. 1988 -: "Zavod za komjuterska tehnika in informatika" formally established*

The proposal given by the author of this text that a zavod or development of computer science should be built based on the ETF Zagreb model, become a reality in 1988. By formal decision of the Department Board ("Fakultetski Sovet") this Zavod was established.

*AE. 1989 -: A worldwide pioneering result: Controlling robots using EOG signals*

Here we will mention on achievement which took place in 1989. In 1989 the first control of a robot using electrooculogram (EOG) was achieved [37][38]. A robot was controlled to go forward, back, left, and right using just the signals from movement of the eyes.

## V. CONCLUSION

This work covers period of first signs of work on computer science by the people from Macedonia in 1959 till 1988 when

the most important result known so far was achieved by the Macedonian computer science.

The first signs of computer science work from Macedonian authors date back to 1959. At that time people from Macedonia were engaged in nuclear engineering research in other part of Yugoslavia and as part of it devices related to computer science were developed. The first computer came to Macedonia in 1967 by IBM initiative, and it was one year before coming to other part of Yugoslavia.

The year 1968 is viewed as start of development of computer sciences in Macedonia, because that year the first computer science paper was written in Macedonian language. The paper was on automata theory and was written by the author of this text, with initiative of professor Gorgi Cupona.

The most important achievement of Computer Science in Macedonia is the 1988 worldwide pioneering result of control of a robot using signals emanating from a human brain. Before 1988 that achievement was assumed to be a science fiction under the name of psychokinesis. Other worldwide pioneering results achieved before 1988 include control of a buzzer using CNV potentials, multitasking software architecture for control of a robot, and controllable flip-flop using 4 inverters only. It is also interesting that the Macedonian language was the second one used to control robots, after the Japanese language, which was first language used to control robots.

Macedonian nation has its place in the pioneering steps in development of computer science, worldwide.

## ACKNOWLEDGEMENT

The results presented here were made possible, in part, by the State of Macedonia Association for Scientific Activities, with financial support of the macro-project Adaptive Industrial Robots through the grant Nr 090110384 in the period 1984-1988.

## REFERENCES

- [1] B. Soucek, A. Hrisoho, "Testing cores of ferromagnetic memories using fast current pulses" (in Croatian), *Elektrotehnicki Vjesnik* 9-10, 313-315, 1959
- [2] B. Soucek, A. Hrisoho, Testing cores of ferromagnetic memories using fast current pulses (in Croatian), *Elektrotehnicki Vjesnik* 9-10 (1959) 313-315
- [3] Decision on recognition "Nikola Tesla", to K. Maksimilijan, A. Hrisoho, B. Soucek, and B. Turko, Foundation for recognition of scientific workers, Republic of Croatia, Zagreb, Meeting 15 April, 1964
- [4] A. Hrisoho, Analog-to-digital conversion in systems with multiparameter analysis (in Croatian) 93 pages, University of Zagreb, 1968
- [5] E. Kowalski, *Nuclear Electronics*, Chapter 4: Analog-to-Digital Converters, Springer Verlag, 1970
- [6] V. Glushkov, *Introduction to Cybernetics* (In Russian) Kiev: Ukrainian Academy of Science, 1964
- [7] V. Glushkov, *Introduction to Cybernetics* Translation from Russian: G Kranc, Academic Press, 1966

- [8] V. Glushkov, *Introduction to Cybernetics* (In Serbian) Translation from Russian: R. Tomovic and M. Uscumlic. Zavod za Izdavanje Udzbenika, 1967
- [9] S. Bozinovski. "Abstract Automata" (in Macedonian, edited by G. Cupona) Topic 10. First Seminar on Cybernetics, Mathematical Institute with Numeric Center, 1968
- [10] S. Bozinovski, "Professor Cupona and the first lecture on abstract automata in Macedonia". *Mathematica Macedonica* 8: 79-94, 2010
- [11] B. Soucek, *Minicomputers in Data Processing and Simulation*. John Wiley and Sons, 1972
- [12] B. Soucek, *Microprocessors and Microcomputers*, John Wiley and Sons, 1976
- [13] D. Bitrakov. *Introduction to Programming Language Fortran IV*. (in Macedonian) Prosvetno Delo, 1972
- [14] S. Bozinovski, "Perceptrons issue: Pattern recognition teaching" Work for First-of-May student scientific competition. University of Zagreb, 1971
- [15] S. Bozinovski, V. Mesaric "An example of digital simulation of a formal neuron: Simulation of Pavlov's conditioned reflex using integrated circuits" (In Croatian) Seminar work, Course of Biomedical Electronics, Electrical Engineering Department, Zagreb, Yugoslavia, 1972
- [16] S. Bozinovski., "Introduction to Psychocybernetics," (In Croatian) *Revija za Psihologiju* 5 :115-128, 1975
- [17] A. Santic, *Biomedical Electronics* (In Croatian), Zagreb, Skolska Knjiga, 1995
- [18] S. Bozinovski, G. Jovancevski, N. Bozinovska, "Simulation of a module-based digital networks using object oriented programming," in *Recent Advances in Java Technology: Theories, Applications, Implementations*, J. Power and J. Waldron (Eds.) Dublin: Trinity College, Computer Science Press, pp. 11-19., 2002
- [19] S. Bozinovski, Operating Systems and Systems Programming booklet for students of the Electro-Mechanical Department following the course on Switching Techniques, 1976
- [20] S. Bozinovski, Elements of the design of hardware and software of digital minicomputers.(In Macedonian, 44 pages) Mathematical School, Mathematical Institute with Numeric Center, 1977
- [21] S. Bozinovski, Digital Integrated Circuits (In Macedonian) Cathedra of Cybernetics, Electrical Engineering Department, Skopje, 1977
- [22] S. Bozinovski, Rectangular Curve for Function Generator (In German, 18 pages) Institut fur Nachrichtentechnik und Informatik, Technische Universitat, Munchen, 1979
- [23] S. Bozinovski, Microprocessors and Microcomputers (in Macedonian) Electrical Engineering Department, Cathedra of Cybernetics, 1979
- [24] S. Bozinovski, Catalog of Electronic and Software Products for Medicine and Psychology. Institute of Electronics, Electrical Engineering Department, Skopje, 1980
- [25] S. Bozinovski, M. Sestakov, "Multitasking operating systems and their application in robot control" (in Macedonian) In: Proc. Workshop on Macedonian Informatics, Skopje, pp. 195-199, 1983
- [26] S. Bozinovski, "Parallel programming for mobile robot control: Agent-based approach" Proceedings IEEE Conf. on Distributed Computing Systems, Poznan, pp. 202-208, 1994
- [27] S. Bozinovski, "A representation theorem of linear pattern classifier training", IEEE Transactions on Systems, Man, and Cybernetics 15 pp. 159-161, 1985
- [28] W. McCulloch, A. Pitts, "A logical calculus of the ideas immanent in nervous activity", *Bulletin of Mathematical Biophysics* 5: pp. 115-133, 1943
- [29] F. Rosenblatt, "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain", *Psychological Review* 65, pp. 386-408, 1958
- [30] S. Bozinovski, "Guest Editor's Introduction", Special Issue on Biological and Non-Biological Beings, *Automatika* 25: 128, 1985
- [31] I. Kato, S. Ohteru, H. Kobayashi, K. Shirai, A., Uchiyama, "Information-power machine with senses and limbs", In Proc. CISM-IFTOMM Symposium on Theory and Practice of Robots and Manipulators, Udine, Italy, pp. 12-24, 1973
- [32] G. Grujovski, "Realization of a system for recognition of isolated words and control of a mobile robot with speech commands" (in Macedonian) Technical Report, Project Adaptive Industrial Robots, based on Grujovski's Diploma Thesis, mentor S. Bozinovski, Electrical Engineering Faculty, University Cyril and Methodius, Skopje, Yugoslavia, 1986
- [33] G. Grujovski, S. Bozinovski, "Realization of a system for speech control of a mobile robot" (In Macedonian) Proceedings of the 6th Yugoslav Symposium on Applied Robotics and Flexible Automation, Novi Sad, Yugoslavia, pp. 227-235, 1989
- [34] A. Bogdanovski, S. Bozinovski Adaptive tuning of an operating system (in Macedonian) Project Adaptive Industrial Robots, ETF Skopje, 1984
- [35] S. Bozinovski, G. Jovancevski, Operating system tuning: An Artificial Intelligence learning system approach" (in Macedonian) Proc. Conf. ETAN, Bled, Yugoslavia, 1987
- [36] S. Bozinovski, M. Sestakov, L. Bozinovska, "Using EEG Alpha Rhythm to Control a Mobile Robot", In G. Harris and C. Walker (eds.) Proc 10th Annual Conference of the IEEE Engineering in Medicine and Biology Society, New Orleans, LA, vol 10, track 17: Biorobotics, pp. 1515-1516, 1988
- [37] S. Bozinovski, M. Sestakov, G. Stojanov, L. Bozinovska, "Bioelectric mobile robot control" (In Macedonian) Proc 6th Yugoslav Symposium on Applied Robotics and Flexible Automation, Novi Sad, Yugoslavia, pp. 237-242, 1989
- [38] S. Bozinovski, "Mobile Robot Trajectory Control: From Fixed Rails to Direct Bioelectric Control", In O. Kaynak (ed.) Proc IEEE International Workshop on Intelligent Motion Control, Istanbul, Turkey, vol. 2, pp. 463-467, 1990
- [39] J. Vidal, "Toward direct brain-computer communication", *Annual Rev. Biophysics and Bioengineering*, pp. 157-180, 1973
- [40] L. Bozinovska, M. Sestakov, G. Stojanovski, S. Bozinovski, "Intensity variation of the CNV potential during the biofeedback training guided by a personal computer" (in Serbian). *Neurologija* 1988, suppl 2, Abstracts, 37
- [41] L. Bozinovska, S. Bozinovski, G. Stojanov, M., Sestakov, "Introduction of biofeedback in the CNV paradigm" (In Serbian) Proc Conf. ETAN, Novi Sad, Yugoslavia, p. XII. 93-98, 1989
- [42] L. Bozinovska, S., Bozinovski, G., Stojanov, "Electroexpectogram: Experimental design and algorithms", Proc. IEEE International Biomedical Engineering Days, Istanbul, Turkey, pp. 58-60, 1992

# **Radio Communications, Microwaves, Antennas**



# Coupled Halfwave Microstrip Resonators with Dumbbell Type Defected Ground Structure

Marin Nedelchev, Alexander Kolev<sup>1</sup>

**Abstract** – The paper presents design, and simulation of coupled halfwave microstrip resonators with dumbbell type defected ground structures. For the design purposes, the coupling coefficient of the coupled resonators with different sizes of the dumbbell structure is investigated. Each topology is simulated in full-wave electromagnetic simulator and the coupling coefficient is extracted. Design graphs of the coupling coefficients are presented.

**Keywords** – microstrip coupled, resonator, coupling coefficient, defected ground, dumbbell

## I. INTRODUCTION

Microstrip filters are key components in the communication systems. They have continuous development in theory and realizations. Edge coupled halfwave microstrip filters are widely used in the practice because of their straightforward design process, easy manufacturing and adjustment, light weight and acceptable size on the printed circuit board. Therefore the improvement of the edge coupled filter characteristics and design process is worth.

One of the problems facing the wideband and ultrawideband filters is the realization of very small gaps between the coupled resonators and the fabrication tolerances connected to their manufacturing. The filter response is also affected by the precision of the manufacturing process of the small gaps. Utilizing different shaped slots in the ground plane of the microstrip line- defected ground structures (DGS) [1-5], it is possible to enhance the coupling coefficient and particularly the electrical coupling. There are proposed different types of DGS topologies in the literature not only for the analysis of the effect of the defects, but also for the implementation of the defects in various structures in microwave filters and devices. DGS is realized by introducing a shape defected on a ground plane so it will disturb the shielded current distribution depending on the shape and dimension of the defect. The disturbance at the shielded current distribution will influence the input impedance and the current flow of the design. The positions and lengths of the dumbbell sections control the response and insertion loss attenuation of the coupled lines. DGS increases the effective

values of dielectric constant of substrates ( $\epsilon_{\text{eff}}$ ), thus, decreases the wavelength and the overall length of the design [6].

Fig.1 shows the different defected ground elements' shapes proposed in [1] and the most used DGS element is the dumbbell shaped.

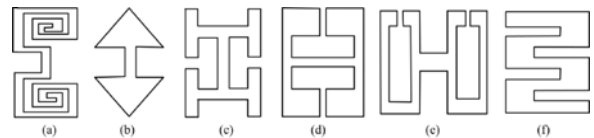


Fig.1 Types of DGS elements-(a) spiral head, (b) arrowhead-shot, (c) "H"-shape slot, (d) a square open-loop with a slot in middle section, (e) open-loop dumbbell and (f) interdigital DGS

Fig. 2 shows the edge coupled halfwave resonators with the dumbbell DGS in the ground plane.

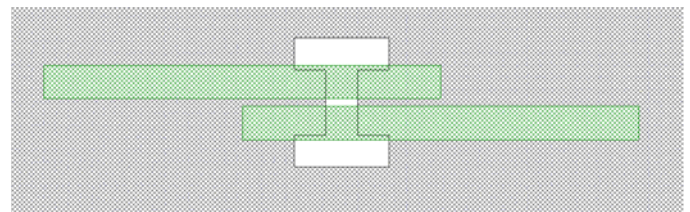


Fig.2 Parallel coupled halfwave resonators with dumbbell DGS

This paper researches the coupling coefficient between two edge coupled halfwave resonators with dumbbell DGS with respect to the variation of its dimensions. As the dumbbell DGS has a lot of degrees of freedom, the coupling coefficients for different dimensions of the dumbbell are computed using a full-wave electromagnetic simulator. The results of the research may be used in microstrip wideband filter design.

## II. DUMBELL DEFECTED GROUND STRUCTURE

All the simulations, design procedures in the paper are performed for dielectric substrate FR-4 with height 1.5mm, relative dielectric constant  $\epsilon_r = 4.4$  and loss tangent  $\text{tg } \delta = 0.02$ .

The form of the dumbbell type defected ground structure and its dimensions is shown on Fig. 3. It consists of an main slot connected on both ends to rectangular dumbbell heads.

<sup>1</sup>Marin Veselinov Nedelchev and Alexander Kolev –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)



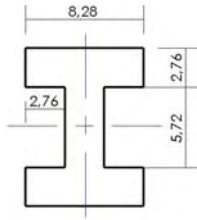


Fig.3. Form and size of the DGS (all sizes in mm)

The edge halfwave coupled resonators have width corresponding to the width of a line with a characteristics impedance of  $50\Omega$ . For the dielectric substrate FR-4, it is computed that the width of the line is  $w = 2.76 \text{ mm}$ .

Fig.4 shows two coupled resonators with length  $l=32.82\text{mm}$  for center frequency  $f_0 = 2.4\text{GHz}$ . The length is determined using a simulation in a full wave EM simulator.



Fig.4 Two coupled resonators with length 32.82mm and width 2.76mm

The responding coupling coefficient for synchronously tuned resonators can be calculated easily by the resonance frequencies of even and odd mode [2], when the coupled resonators are overcoupled:

$$k = \frac{f_e^2 - f_o^2}{f_e^2 + f_o^2} \quad (1)$$

A full wave EM simulator based on the Method of the Moments (MoM) is used to identify the resonance frequencies in the response [7,8]. Figure 5 shows the frequency response of the coupled resonators, when there is no DGS and when we place a dumbbell DGS with the form and dimensions shown in Fig.3.

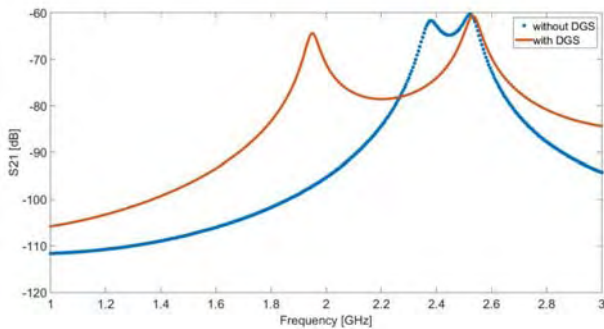


Fig. 5 Frequency response of the coupled resonators with and without dumbbell DGS

It is seen that the slot in the ground plane and the corresponding current disturbance enhances the electric coupling between the resonators. The split resonance frequency of the even mode, when DGS is used, is far away from the split resonance frequency, when no DGS is used. The odd mode resonance frequency is slightly higher than when no DGS is used. Consequently the dumbbell DGS under the coupled resonators enhances the electric part of the

coupling coefficient. There is another physical description of this effect. In the microstrip coupled resonators without DGS, in the coupling mechanism many of the electric lines start from the top layer and end on the ground plane passing through the dielectric substrate. In the presence of the slot a part of these lines are forced to end on the other resonator, enhancing this way the electric part of the coupling. In this way the effective dielectric permittivity for even mode is much higher than the effective dielectric permittivity for odd mode. The DGS topology and size define the coupling coefficient behavior and enhancement.

This fact allows additional degrees of freedom in the synthesis of coupled resonator circuits [9].

### III. PARAMETRIC ANALYSIS OF DGS SIZE IMPACT ON COUPLING COEFFICIENT

This section presents the results from the parametric study of the coupling coefficient between coupled halfwave resonators with dumbbell DGS in the ground plane. This parametric study will give the designer comprehensive understanding of the effect of the DGS under the coupled halfwave resonators. The study lead to design curves for coupled halfwave resonators with dumbbell DGS in the ground plane. It will assist the design process of microstrip filters synthesis in the ISM band of 2.4GHz.

The parametric study is performed in full wave EM simulator. Geometrically, the center point of the dumbbell DGS is placed in the center point of the coupled resonators.

Fig.6 shows the four parameter changes in the dumbbell DGS size- head height, head width, slot width and slot height.

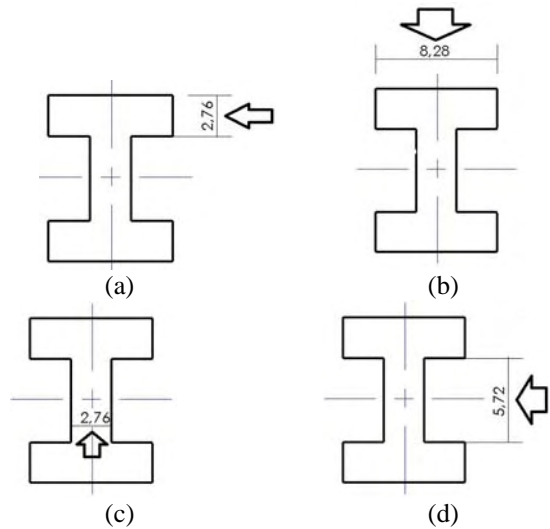


Fig.6 Dumbbell-type DGS size variation.

(a) height of the head, (b) width of the head, (c) width of the slot line (d) length of the slot line

Using full wave EM simulator, the split resonance frequencies in the magnitude response of the coupled resonators are obtained. Using Eq.1, the coupling coefficient is computed for heights of the dumbbell head from 0.2mm to 5mm. All the rest dimensions of the DGS topology remain unchanged. The corresponding family of curves for the

coupling coefficient are shown on Fig.7. Generally, the coupling coefficient is monotonically decreasing with the increasing the distance between the resonators. The gap between the coupled resonators is changed discretely from 0.2mm to 3.8mm. As the head of the dumbbell becomes bigger, the coupling coefficient is also increasing. For one and the same gap between the resonators, the value of the coupling coefficient can be at least two times bigger. The height of the dumbbell head has significant impact on the value of the coupling coefficient. It leads to technologically easy tuning of the coupling coefficient by changing the height during the filter adjustment.

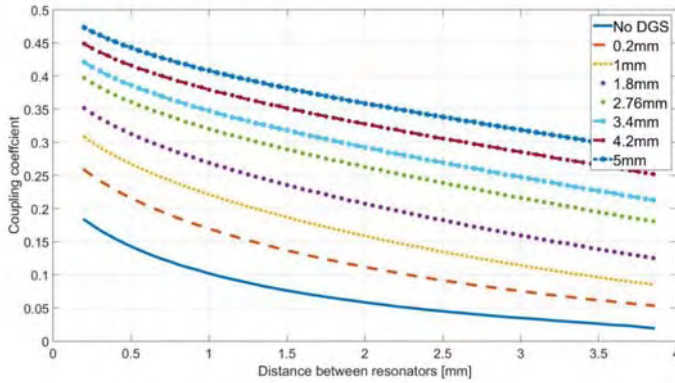


Fig.7 Coupling coefficient for height of the DGS head variation from 0.2mm to 5mm

Fig.8 shows the results from the study of the coupling coefficient, when the width of the DGS head is changed (as shown on Fig.6b). Again, the coupling coefficient is monotonic, but the big increase of the width does not lead to corresponding significant increase of the coupling coefficient. As it is seen from Fig.8, the presence of dumbbell DGS increase at least two times the value of the coupling coefficient.

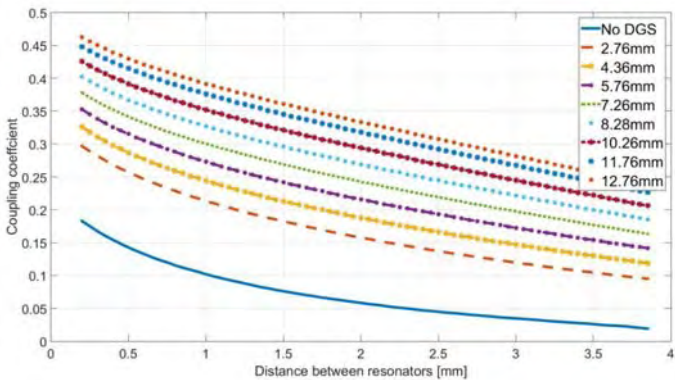


Fig.8 Coupling coefficient for width of the DGS head variation from 2.76mm to 12.76mm

Fig.9 figures out the results for the coupling coefficient when the slot of the dumbbell DGS is changed discretely from 0.2mm to 5.52. The current in the ground plane caused by the slot width is comparably small and the coupling coefficient enhancement is not so well pronounced like the cases when the other parameters are changed. The coupling coefficient increases in 30 to 50% compared to the smallest DGS slot width.

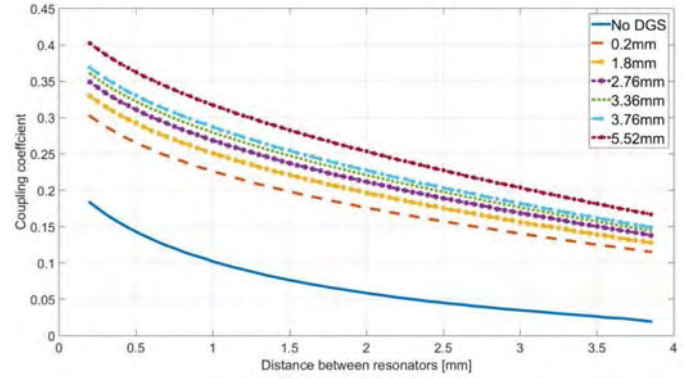


Fig.9 Coupling coefficient for width of the DGS slot variation from 0.2mm to 5.52mm

Fig.10 shows the coupling coefficient dependence on the distance between the resonators and the length of the slot of dumbbell DGS. The values of the coupling coefficient are at least 25% bigger than the values without DGS. The higher is the length, the bigger is the coupling coefficient.

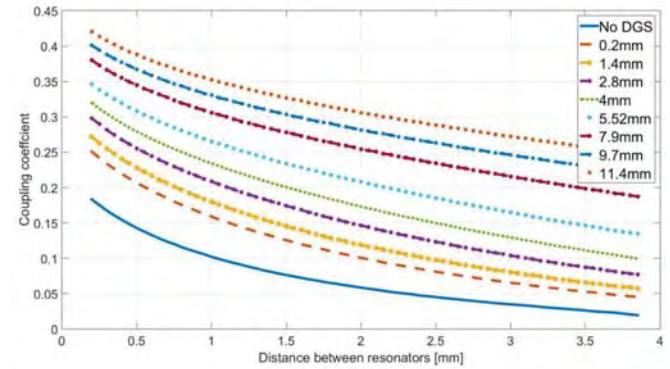


Fig.10 Coupling coefficient for length of the DGS slot variation from 0.2mm to 11.4mm

The parametric study of the coupling coefficient between coupled halfwave resonators confirm the assumption, that introduction of defects of specific topology in the ground plane of the microstrip technology will enhance it. This approach may be used in the wideband and ultrawideband microstrip filter design. From the study it was described the mechanism of the coupling coefficient increasing. The derived design graphs can be used by microstrip filter designers for the ISM band on 2.4GHz.

#### IV. CONCLUSION

This paper presents the coupling coefficient between two edge coupled halfwave resonators with dumbbell DGS with respect to the variation of its dimensions. The coupling coefficient between the resonators is derived using the split resonance method. The results from the studies are presented in several design graphs. They can be used in microstrip filter design in the ISM band on 2.4GHz on FR-4 substrate.

## ACKNOWLEDGEMENT

The authors would like to thank Prof. Ilia Iliev, PhD, Dean of the Faculty of Telecommunication, Technical University of Sofia for the encouragement during the study and the bright ideas.

## REFERENCES

- [1] L.H. Weng, Y.C.Guo, X.W. Shi and X.Q.Chen, "An Overview on Defected Ground Structure", Progress In Electromagnetics Research B., vol. 7, pp. 173-189, 2008.
- [2] J.Hong, M.J Lancaster, "Microstrip Filters for RF/Microwave Applications", John Wiley and Sons Inc, 2001.
- [3] G.Breed, "An Introduction to Defected Ground Structures in Microstrip Circuits", High Frequency Electronics, pp 50-54, 2008.
- [4] C.Carg, M.Kaur, "A Review of Defected Ground Structures (DGS) in Microwave Design", International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering, Vol.2 Issue 3, 2004.
- [5] S.K. Parui, S.Das, "A New Defected Ground Structure for Different Microstrip Circuit Applications ", Radioengineering Vol. 16 No 1, 2007.
- [6] D.M. Pozar, "Microwave Engineering", John Wiley and Sons Inc., 2012.
- [7] D.G. Swanson Jr, W.J.R Hoefer, "Microwave Circuit Modeling Using Electromagnetic Field Simulation", Artech House 2003.
- [8] D.B. Davidson, "Computational Electromagnetics for EF and Microwave Engineering", Cambridge University Press, 2005.
- [9] Z.Hou, X. Li and C. Hao, "Design of Wideband Filter Using Split-ring Resonator DGS", Progress in Electromagnetics Research Symposi,China, March 24-28, 2008.



# Effect of RF Receiver Nonlinearity on Mobile Telecommunication Systems

Youssef Abd Alrahman<sup>1</sup> and Ilia Iliev<sup>2</sup>

**Abstract** – This paper proposes a modified model of signal-to-interference-and-noise ratio (SINR) which include the interference due to intermodulation products. Using this model, we study the effect of receiver's nonlinearity on the performance of telecommunication systems. We examine the problem of interference caused by intermodulation products generated by unwanted interference signals at receivers. Specifically, third-order intermodulation distortion (IMD3) that could be at frequencies within working band and drive circuits into saturation causing blocking for wanted signal at same frequency. Using Matlab software and nonlinear amplifier model, we achieve a practical calculations for third-order intercept point (IP3). Effect of Intermodulation products on SINR is shown using the third-order input intercept point (IIP3) specifications for GSM receiver and SINR modified model and Matlab software.

**Keywords** – SINR, GSM, RF receiver, Nonlinearity, Intermodulation.

## I. INTRODUCTION

In according to manage the increasing in the number of mobile subscribers around the world and their demands, more operators and base stations and frequencies reuse should be achieved. Which lead to a diverse range of challenges in frequency planning. Most important one is intermodulation because - as an example - third-order intermodulation products generated by two GSM-1800 signals could fall in UMTS band and interfere with one UMTS signal [1]. Thus, although mobile stations receive desired signals only from their operating base station, high-power signals from the neighboring base stations for another service providers, regardless of their carrier frequencies, may cause desensitization and blocking by forcing their circuits into saturation.

The intermodulation in mobile nonlinear systems can generate two different types of interference: the one that is called in-band, that is generated inside the GSM-900, GSM-1800 or UMTS band and is derived from inside frequencies, and the out of band interference that is generated mixing terms from the GSM-900, GSM-1800 or UMTS band [1].

The most problematic of all the intermodulation distortions are third-order intermodulation products with frequencies  $2f_a - f_b$  and  $2f_b - f_a$ . This is because when the difference between  $f_a$  and  $f_b$  is small, the intermodulation products fall

in-band close to the fundamentals. If the fundamentals at frequencies  $f_a$  and  $f_b$  are two strong interfering signals, they could produce third-order intermodulation products within the low-noise amplifier of the victim receiver and corrupt the desired signal. The performance metric that characterizes nonlinear devices for such third-order intermodulation products is the “third-order intercept point” (IP3) [2].

In this paper, we use modified model of signal-to-interference-and-noise ratio (SINR) which include the interference due to intermodulation products which could be generated due to receiver nonlinearity. This modified model could be used to analyze receiver nonlinearity characteristic to minimize intermodulation products effect, on the other hand it could be used to modify mobile networks planning technique and frequency allocation algorithms to avoid generating these unwanted interference signals beside of receiver nonlinearity. We use Matlab software to achieve a practical calculations for third-order intercept point (IP3) and effect of intermodulation products on SINR.

## II. SINR MODELS

The signal-to-interference-and-noise ratio is defined as the power of a certain signal of interest divided by the sum of the interference power (from all the other interfering signals) and the power of noise all at the output of receiver. In the mobile telecommunication systems standard SINR ( $SINR^s$ ) is widely used to calculate signal-to-interference-and-noise ratio [4] [5] [6].

$$SINR^s = \frac{P_{signal}}{N + P_{interference}} \quad (1),$$

where  $P_{signal}$  and  $N$  are respectively the desired signal and thermal noise power at the receiver, while  $P_{interference}$  is interference power form all external interference resources.

Practically, the most studied interference in the cellular mobile systems are co-channels and adjacent channels interference. This lead to the enhanced SINR model ( $SINR^e$ ) which take into account these two main source of interference.

$$SINR^e = \frac{P_{signal}}{N + \sum_i P_i} \quad (2),$$

where  $P_i$  is the power of  $i^{th}$  co-channel and adjacent channel interfering signals. The received signal power  $P_{signal}$  and  $P_i$  could be calculated using one model for signal propagation and path loss. There are many models for co-channel and adjacent channel interference available to use [7] [8].

Standard and enhanced SINR models are used to estimate performance of wireless networks. But these models not

<sup>1</sup>Youssef Abd Alrahman is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: tel.eng.josef@live.com.

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

considering probability of interference due to nonlinearity of RF receiver. Intermodulation products generated in nonlinear stages in receiver such as mixers and amplifiers could cause interference with the desired signal and therefore should be included.

### III. ANALYSIS OF INTERMODULATION INTERFERENCE

#### A. Intermodulation Modeling

Intermodulation products are produced when two or more in-band signals as given in Eq. (3) are applied to the input of a nonlinear receiver has transfer function given in Eq. (4), these signals mix within the receiver and they form new components that are not harmonics of the input frequencies. These components are termed as intermodulation distortion products, as shown in Fig.1 [2] [9].

$$x = a \cos 2\pi f_a t + b \cos 2\pi f_b t \quad (3)$$

$$y(t) = \sum_{i=0}^n k_i * x(t)^i \quad (4)$$



Fig.1. Even and odd-order intermodulation products.

The  $n$ th order intermodulation product is given at frequency  $\pm n_a f_a \pm n_b f_b$  where  $|n_a| + |n_b| = n$ . The transfer function order could be limited to 3, because the higher order products are of less significance.

In this paper, we consider only the products at frequencies  $2f_1 - f_2$  and  $2f_2 - f_1$  which are commonly referred as third-order intermodulation distortion (IMD3) products. If the distance between the frequencies of two in-band interferers ( $f_1$  and  $f_2$ ) is similar to the distance between one of the interferers and the useful signal at  $f_0$ , one of the resulting intermodulation products may fall close or within the receiver channel and may corrupt the desired signal. The IMD calculation method based on intercept points is as follow [3]:

When two incoming signals with frequencies at  $f_1$  and  $f_2$  interact, only two third-order intermodulation products of type IM3(2,1) are possible. In case of three signals with frequencies at  $f_1$ ,  $f_2$  and  $f_3$ , intermodulation products of types IM3(2,1) and IM3(1,1,1) are possible. IM3(2,1) and IM3(1,1,1) products frequencies  $f_{IMD}$  are as given in Eqs. (5) and (6) respectively.

$$f_{IMD} = 2f_a \pm f_b \quad (5)$$

$$f_{IMD} = f_a \pm f_b \pm f_c \quad (6)$$

where  $a, b, c \in \{1, 2, 3\}$  and  $a \neq b \neq c$ .

Power of intermodulation product IM3(2,1) can be calculated by:

$$P_{IMD-2} = 3(P_{e-in} + G_{RX}) - 2OIP_3 \quad (7),$$

where  $P_{IMD-2}$  is output power of the third-order intermodulation product IM3(2,1) in dBm,  $OIP_3$  is 3rd order output intercept point of the receiver in dBm,  $G_{RX}$  is receiver gain in dB and  $P_{e-in}$  is input power of incoming signal into the receiver in dBm.

$$P_{e-in} = (2P_1 + P_2)/3 \quad (8),$$

where  $P_1, P_2$  are powers of incoming signals at frequencies  $f_1, f_2$  correspondingly. Power of intermodulation product IM3(1,1,1) can be calculated by :

$$P_{IMD-3} = 3(P_{e-in} + G_{RX}) - 2OIP_3 + 6 \quad (9),$$

where  $P_{IMD-3}$  is output power of the third-order intermodulation product IM3(1,1,1) in dBm.  $P_{e-in}$  in this situation is given by:

$$P_{e-in} = (P_1 + P_2 + P_3)/3 \quad (10),$$

where  $P_1, P_2, P_3$  are powers of incoming signals at frequencies  $f_1, f_2, f_3$  correspondingly. The equivalent intermodulation product level recalculated to the input of the receiver  $P_{in-IMD}$  (dBm) is equal to:

$$P_{in-IMD} = P_{IMD} - G_{RX} \quad (11).$$

Or it could be calculated by:

$$P_{in-IMD} = 3P_{e-in} - 2IIP_3 + 6 \quad (12).$$

Interferences caused by intermodulation products in the receiver occur when the following two conditions are fulfilled:

$$F_R - 0.5B_{IF} \leq f_{IMD} \leq F_R + 0.5B_{IF} \quad (13).$$

and

$$SIR < A \quad (14),$$

where  $F_R$  is tuning frequency of the receiver,  $B_{IF}$  is passband value of the IF stage or based band filter bandwidth if there is no IF stage and  $A$  is co-channel protection ratio.

#### B. Modified SINR model

As shown in Fig.2, intermodulation products create unwanted components at frequencies which may interfere with the desired signal. In a worst case, the IMD3 product will fall directly into the desired channel (similar to co-channel interference). With increasing of the frequency offset between the IMD3 product and the desired signal, their impact to SINR becomes less harmful (similar to adjacent channel interference). Modified model of SINR will include the interference due to third-order intermodulation distortion (IMD3) products and is given by Eq. (15).

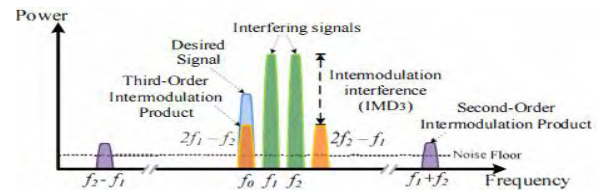


Fig.2. Third-order intermodulation distortion causing interference in the desired channel of a receiver.

$$SINR^m = \frac{P_{signal}}{N + \sum_i P_i + P_{IMD3}} \quad (15),$$

where  $P_{IMD3}$  represents interference power caused by third-order intermodulation distortion and it can be expressed by:

$$P_{IMD3} = \sum_k \sum_j P_{IMD-2} + \sum_k \sum_j \sum_l P_{IMD-3} \quad (16),$$

where  $k \neq j \neq l$  are the number of frequencies used in the telecommunication system and  $f_k, f_j, f_l$  are the signals which produce components with frequency equal to  $f_{IMD}$  as given in Eqs. (5) and (6) and this  $f_{IMD}$  is achieve the condition in Eq.(13) or it shouldn't be taken into account. In mobile telecommunication systems, there are many ways to minimize interference due to co-channel and adjacent signals during frequency planning such sectoring and cell-splitting and that will increase  $SINR^e$  [8] [10]. In this paper we consider interference generated just form intermodulation products, so modified model of SINR can be expressed as:

$$SINR^m = \frac{P_{signal}}{N + \sum_k \sum_j P_{IMD-2} + \sum_k \sum_j \sum_l P_{IMD-3}} \quad (17).$$

#### IV. THIRD-ORDER INTERCEPT POINT $IP_3$ CALCULATION USING MATLAB SOFTWARE

We calculate third-order intercept point for nonlinear system such as nonlinear amplifier by using simple model as in Fig.3 with two signals with frequencies (990,1010 MHz), considering amplifier with gain 0 dB and noise figure 3.01 dB. We use spectrum analyzer at the input and the output of amplifier to measure the power of the fundamental signal and power of third-order product signal.

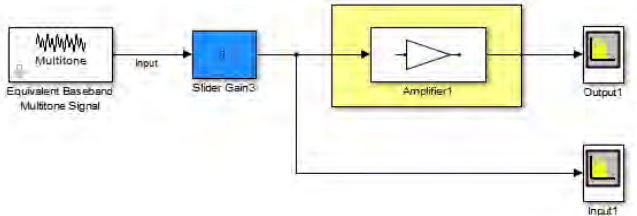


Fig.3. Matlab model for practical calculation.

To achieve that there are two options:

*Use two signals at equal level:* The power of input signals is equal and range between  $(-27 \sim 48) dBm$ , by using spectrum analyzers to measure powers at the output, the  $IP_3$  could be shown in Fig.4.

*Use two signals at different level:* Power of the first input signal will be constant at 5 dBm, and the second input signal power will range from  $-37 dBm$  to  $49 dBm$ , and with spectrum analyzers measure output powers for the fundamentals signals and third-order intermodulation product. But here to have a correct calculation, the equivalent input and output power are calculated by using Eq. (8). Then by plotting fundamental equivalent signal with third-order intermodulation product, in Fig. 5, could be found that  $IP_3$  is equal to 30 dBm.

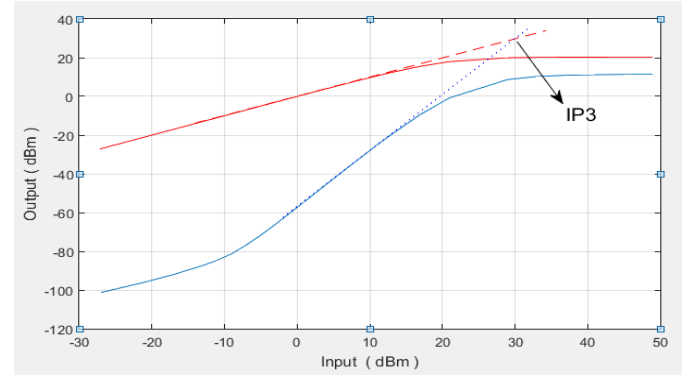


Fig.4. Practical Measurements 1. (blue line for the third intermodulation product)

As shown in Fig.5 the third-order intercept point  $IP_3$  is equal to 30 dBm.

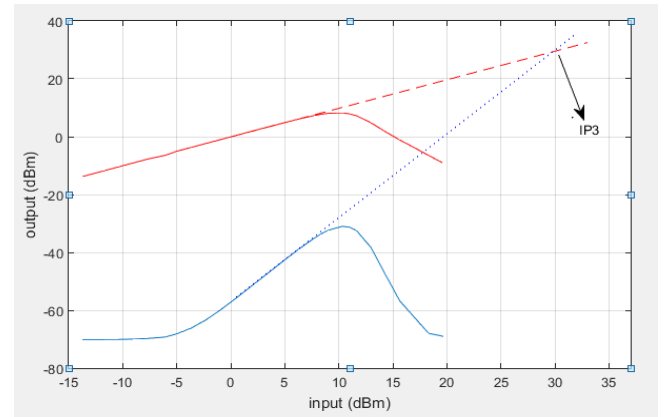


Fig.5. Practical measurements 2. (blue line for the third intermodulation product)

#### V. SIMULATION RESULTS USING MODIFIED SINR MODEL

For GSM receiver with noise floor  $N = -111 dBm$ , the minimum SINR value required for GSM receiver is 9 dB. That's mean for desired signal at  $-99 dBm$ , if the third-order input intercept point is  $IIP = -18 dBm$  and there are just two interfere signals have intermodulation component at desired signal's frequency, the maximum acceptable power for these signals is  $-49 dBm$  to achieve wanted SINR value [11] [12].

In this simulation, we will examine these conditions for GSM receivers and study the impact of  $IIP_3$  and power of interfere signals on SINR according to our modified SINR model using tow interfere signals.

##### A. Impact of $IIP_3$ on SINR

We use the modified SINR model as expressed in Eq. (17) with GSM receiver specifications from [12]. Fig.6 characterizes relationship between third-order input intercept point  $IP_3$  and SINR for different interfere signals power  $P_{e-in}$  where desired signal at  $-99 dBm$  and noise floor  $-111 dBm$ .



The simulation results are: in term to achieve desired SINR value for RF receivers, the IIP3 should increase with increment of the interfere signals power which may produce intermodulation products in the receiver. As given in [11] [12], third-order input intercept point IIP3 for GSM receiver is about  $-18\text{ dBm}$ . That's mean when desired signal at  $-99\text{ dBm}$  and interfere signals with power more than  $-49\text{ dBm}$ , SINR will fall under accepted value  $-9\text{ dB}$ .

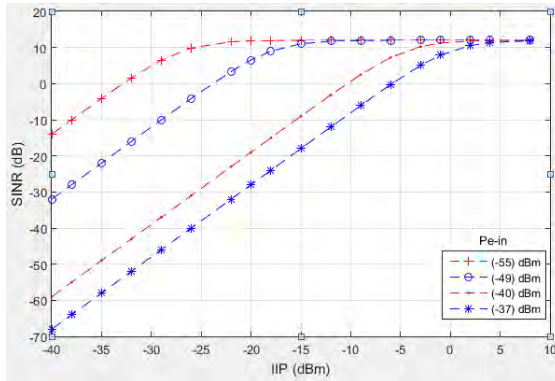


Fig.6. Impact of third-order input intercept point IIP3 on SINR.

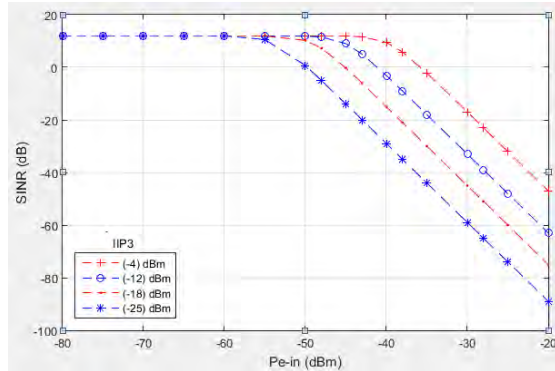


Fig.7. Impact of  $p_{e-in}$  on SINR.

### B. Impact of interfere signals on SINR

By using our SINR modified model with desired signal at  $-99\text{ dBm}$  and noise floor at  $-111\text{ dBm}$ , relationship between power of interfere signals  $p_{e-in}$  and SINR is depicted in Fig.11. To achieve the required SINR for GSM receiver, acceptable interfere signals power is about  $-54\text{ dBm}$  when IIP3 is  $-25\text{ dBm}$ ,  $-49\text{ dBm}$  when IIP3 is  $-18\text{ dBm}$  and about  $-40\text{ dBm}$  when IIP3 is  $-4\text{ dBm}$ . This results show that with increment of the IIP3 value acceptable value for interfere signals power is get increase. From Fig.7, the RF receiver reject for intermodulation products is get higher when IIP3 increase and the linearity of the receiver is increase.

## VI. CONCLUSION

As a result of growth of wireless systems and the existence of many service providers at same area, the intermodulation products distortion rise up as significant issue which should be mitigated to insure required SINR for receivers.

In this paper, we study the effect of third-order intermodulation product on signal-to-interference-and-noise ratio. the results show that when IIP3 increases, the receiver could handle with higher power of interfere signals (producers of third-order intermodulation product) with acceptable value of SINR. Modified SINR model could be used to get control on receiver nonlinearity effect, on the other hand it could be used to modify mobile networks planning technique and frequency allocation algorithms to avoid generating these unwanted interference signals beside of receiver nonlinearity. Thus, there are a lot of challenges for RF receivers designers to achieve all of these requirements.

## REFERENCES

- [1] Nuno Borges de Carvalho and Raquel Castro Madureira, "Intermodulation Interference in the GSM/UMTS Bands," , Journal Proc. 3a Conf. de Telecomunicaoes (CONFTELE 2001), October 2000.
- [2] Shabbir Ahmed, " INTERFERENCE MITIGATION IN COLOCATED WIRELESS SYSTEMS," doctoral research, Melbourne, Australia, December 2012.
- [3] Rec. ITU-R SM.1134-1, "Intermodulation interference calculations in the land-mobile service," ITU, 1995-2007.
- [4] Jean-Marc Kelif, Stephane Senecal, Constant Bridon, Marceau Coupechoux, " A Fluid Approach for Poisson Wireless Networks",. Cornell University, New York, USA, January 2014.
- [5] Moses Ekpenyong and Joseph Isabona, "AN ENHANCED SINR-BASED CALL ADMISSION CONTROL IN 3G NETWORKS". *International Journal of Wireless & Mobile Networks (IJWMN)* Vol. 3, No. 5, October 2011.
- [6] Olga Goussevskaia, Yvonne Anne Oswald and Roger Wattenhofer, "Complexity in Geometric SINR". *MobiHoc'07*, September 9–14, 2007, Montreal, Quebec, Canada, 2007.
- [7] Faramarz Hendesi, Asrar Shiekh and Roshady M.hafiz, "CoChannel and Adjacent Channel Interference in Wireless Cellular Communications". *Wireless Personal Communications*, March 2000, Volume 12, Issue 3, pp 239-253.
- [8] Nura Muhammad Shehu, "Coverage and Capacity Improvement in GSM Network". *International Journal of Novel Research in Electrical and Mechanical Engineering*, Vol.2 , Issue.3 , pp: 57-62, September-December 2015, [Online].Available: [www.noveltyjournals.com](http://www.noveltyjournals.com).
- [9] Q. Gu, *RF System Design of Transceivers for Wireless Communications*. Springer, 2005.
- [10] Lina Lan, Xuerong Gou, Jingli Mao, Wenyuan Ke, "GSM Co-Channel and Adjacent Channel Interference Analysis and Optimization". *TSINGHUA SCIENCE AND TECHNOLOGY* ISSN 1007-0214 04/12pp583-588 Volume 16, Number 6, December 2011.
- [11] Jacques C. Rudell, Jeffrey A. Weldon, Jia-Jiunn Ou, Li Lin, and Paul Gray, "An Integrated GSM/DECT Receiver: Design Specifications". UCB Electronics Research Laboratory Memorandum, April 1998.
- [12] ETSI TS 100 910 V5.12.0 (2001-02), "Digital cellular telecommunications system (Phase 2+); Radio transmission and reception". [Online].Available: [http://www.etsi.org/deliver/etsi\\_ts/100900\\_100999/100910/05.12.00\\_60/ts\\_100910v051200p.pdf](http://www.etsi.org/deliver/etsi_ts/100900_100999/100910/05.12.00_60/ts_100910v051200p.pdf).

# Omnidirectional Double Biquad Omni-Antenna for 2.4 GHz Wireless Link Application

Kalina Kalinovska<sup>1</sup> and Peter Z. Petkov<sup>2</sup>

**Abstract** – In this paper a modified biquad antenna is proposed. The modified antenna has features, relatively high gain and small physical dimensions. The directivity is omni in horizontal direction. The antenna is tuned for 2.4GHz-2.45GHz bandwidth. The antenna can be used for Wi-Fi communications and for drones communications.

**Keywords** – Biquad, biQUAD, QUAD, antenna, Omni-antenna, Wi-Fi, drone, omnidirectional

## I. INTRODUCTION

The trend in recent years shows that the number of Internet users is increasing and well as the number of devices used by every individual. The realization of user's devices full functionality requires Internet access. More and more devices are mobile and able to connect to the web via a wireless transmission environment. The development of apps that require transmission media is increasing the need of good coverage rapid speed of data transmission over Wi-Fi HotSpots. These Wi-Fi HotSpots are located in different places in our homes, shopping centers, various conference rooms, etc. These antennas should be mostly omni-directed graph with the greatest possible gain. The antennas, which are mainly used for Wi-Fi routers are omni-directed, with a gain of more than 2 dBi, with a minimum vertical width of the radiation pattern. The aim of the article is to develop an antenna for Wi-Fi access by the standards of IEEE 802.11b/g [1], for use in public places that have many users and need antennas with omni-directed radiation pattern. These antennas should provide better coverage in indoor and outdoor environments. Also the aim of article is to prove the ability to use the antenna for communication between the photo-drone and the navigator.

## II. BIQUAD ANTENNA CONCEPTS

### A. IEEE 802.11 standards

IEEE 802.11 is a standard that is used for wireless local area networks (WLAN). 802.11 has several widely used protocol - 802.11a, 802.11b, 802.11g and 802.11n, which are mostly in 2.4 - 2.5 GHz range.

Despite advances in technology standards in 5.25 – 5.35 GHz,

2.4 - 2.5 GHz frequency band is still preferred. 802.11b/g/n [2] standards are preferred over others using higher frequencies. Greater attenuation with higher frequency in free. Development in the article can be used for all standards that are 2.4 - 2.5 GHz frequency band. Through the optimized model of the current development can be used in different frequency standards.

### B. Quad Antenna

Quad antennas are type of loop antenna. These antennas are specific to their closed loop which may have a specific geometric shape. An example is the square loop antenna. The length of each wire is a multiple of the wavelength  $\lambda$  with multiplicity  $\lambda / 4$ . (1.1) (Figure 1) [3]

$$l = \lambda / 4 \quad (1.1)$$

The single square loop antenna has two beams with a maximum gain about 4 dBi. (Figure 2) This antenna is powered by one of its corners. The various loop antennas are preferred for mobile communications at higher frequencies. They also have easy workmanship. Materials for these antennas can come from many places. The conductor of which can produce the antenna can be copper wire. Its thickness can vary depending on the desired bandwidth.

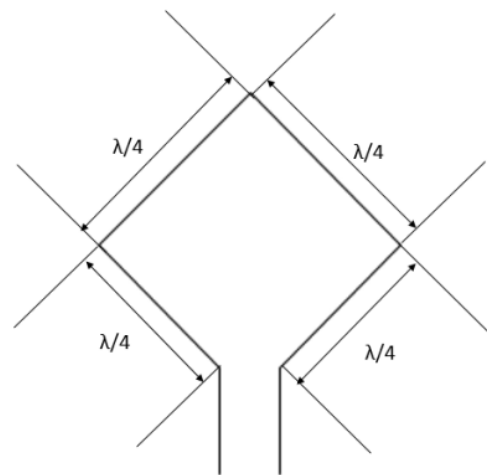


FIG. 1 – QUAD ANTENNA - PRINCIPLED SCHEME

<sup>1</sup> Kalina Kalinovska is a student at Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria. kalina.kalinovska@gmail.com

<sup>2</sup> Peter Z. Petkov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria. pjpetkov@tu-sofia.bg

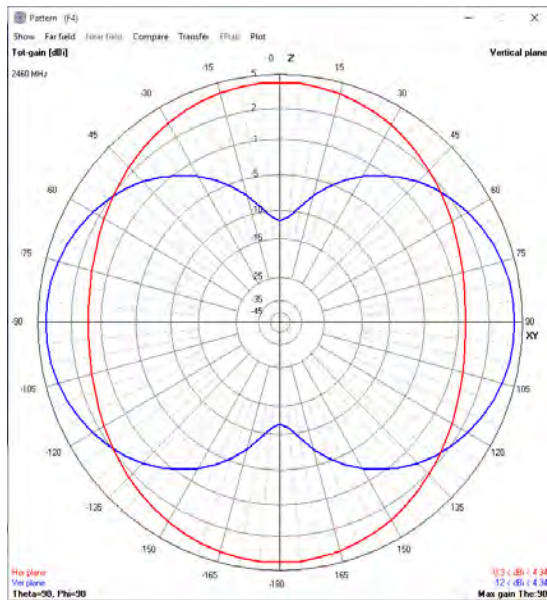


FIG. 2 – QUAD ANTENNA – DIRECTIVITY IN POLAR COORDINATE SYSTEM

### C. Biquad antenna

If two square loop antennas join the indicated in (Figure 3), it gets biQUAD antenna. It is easy to construct an antenna which consists of two squares on which can be placed reflector. It provides greater amplification compared to the square loop antenna. The polarization of the emitted signal depends on the spatial location of the antenna. Horizontal - horizontal, vertical - vertical polarization. [4] The horizontal position represents when both squares are next to each other, the vertical position is when both squares are placed one above the other.

The radiation pattern is similar to square loop antenna but biQUAD antenna provides greater gain of almost 5 dBi. The width of the main sheet is approximately 70 degrees. (Figure 4) [5]

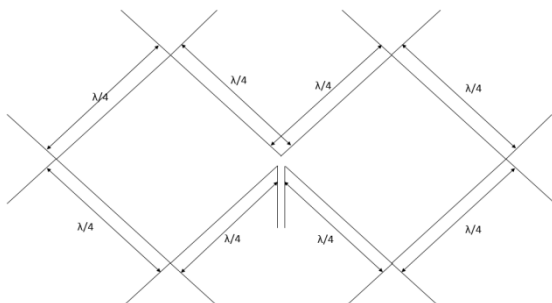


FIG. 3 – BIQUAD ANTENNA – PRINCIPLE SCHEME

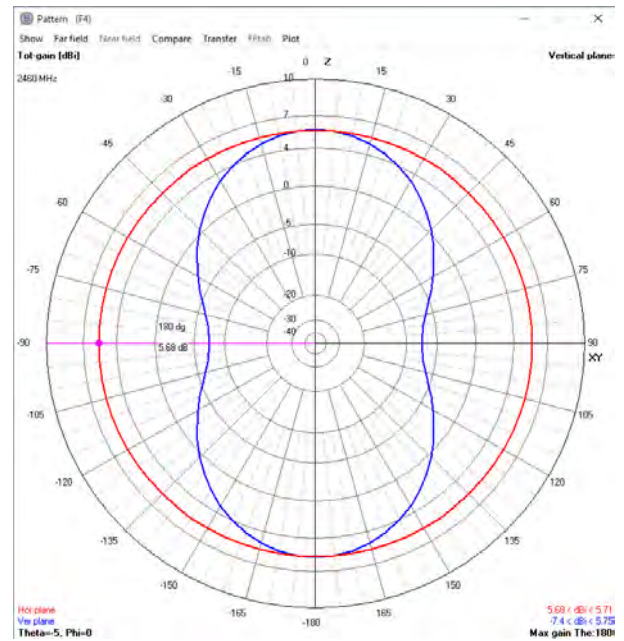


FIG. 4 – BIQUAD ANTENNA – DIRECTIVITY PATTERN IN POLAR COORDINATE SYSTEM

### D. Double Biquad

When connecting two Biquad antennas, as described in Figure 5 scheme gets Biquad. The antenna is with horizontal omni-directional radiation pattern. (Figure 6) This type of antennas are not widespread. In recruiting literature examination of the problem were not found formulas and previous developments. The theoretical data comparable and similar structures show that each side of it must have a length of  $\lambda/4$ , but the simulations and practical measurements do not match the theoretical model. When optimizing antenna simulation with 4nec2 software, there is a tendency in extending the wires. In real conditions there is an extension of the wires by  $\sim 20\%$ . The prototype (Figure 6) confirmed these simulations.

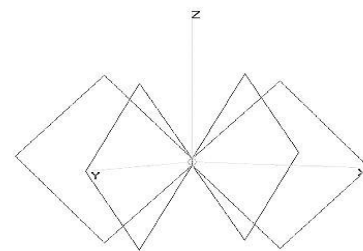


FIG. 5 – DOUBLE BIQUAD ANTENNA – PRINCIPLE SCHEME

This extension is due to surface currents distribution on the wires and the angle between the wires. The directivity of the antenna is omni in both horizontal and vertical direction (Figure 6)



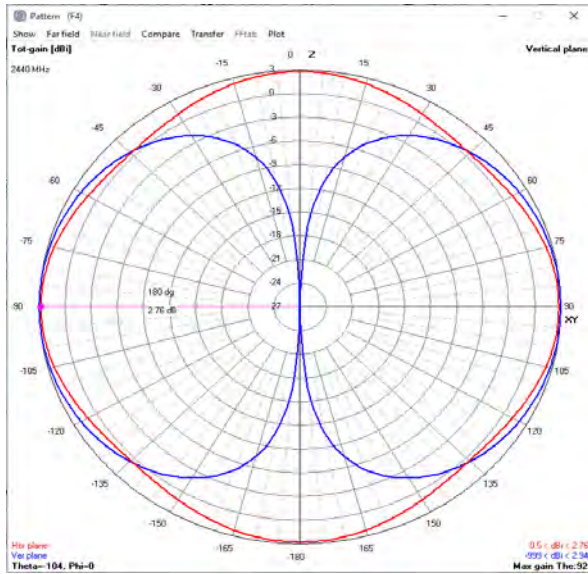


FIG. 6 – DOUBLE BIQUAD ANTENNA – DIRECTIVITY IN POLAR COORDINATE SYSTEM

The simulation model showed bandwidth is 2200-2700MHz . The prototypes bandwidth is wider than the theoretical model (Figure 7).

The prototype of the antenna was made from copper wire with thickness of 2mm and SMA female connector. The center frequency was shifted because of the complicated geometric shape.

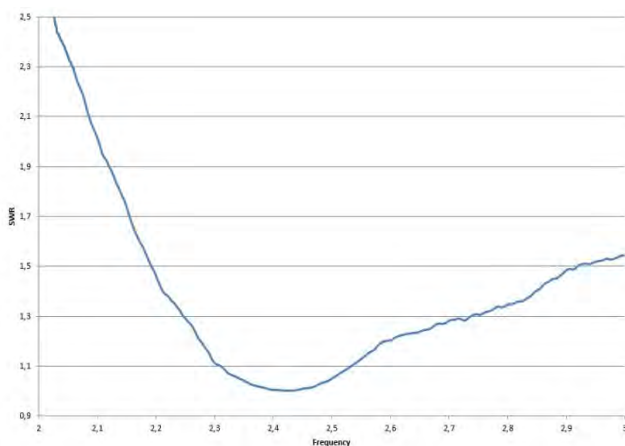


FIG. 7 – DOUBLE BIQUAD ANTENNA – PROTOTYPES BANDWIDTH, SWR, EXPERIMENTAL MEASUREMENTS

#### E. Double biquad with reflector

When placing a reflector at a distance of  $\lambda/4$  from the center of the structure. (Figure 8) The antenna with reflector is not omni in vertical direction, but in horizontal is. The maximum gain is 4.5dBi. (Figure 9, Figure 10) Reflector can be placed

below the antenna. The best position is  $\lambda/4$  from the geometrical center of the structure.

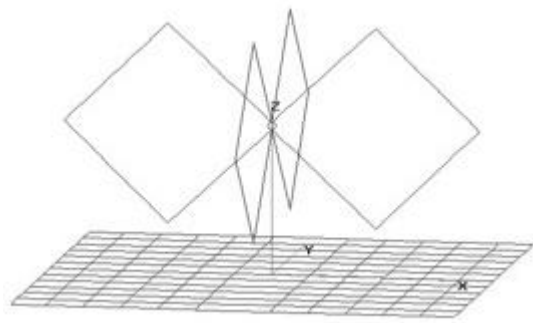


FIG. 8 – DOUBLE BIQUAD ANTENNA WITH REFLECTOR– PRINCIPLED SCHEME

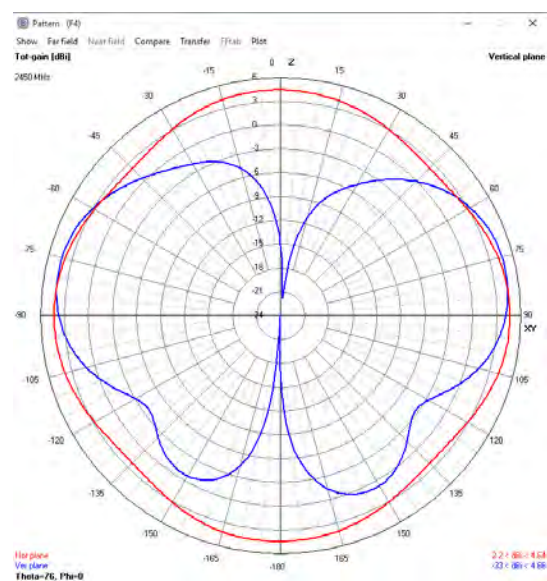


Fig 9 – Double biquad antenna with reflector– directivity in polar coordinate system

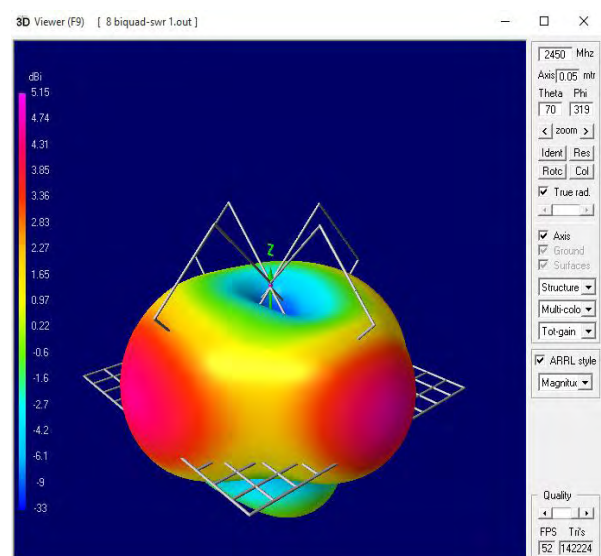


FIG. 10 – DOUBLE BIQUAD ANTENNA WITH REFLECTOR – DIRECTIVITY IN 3D

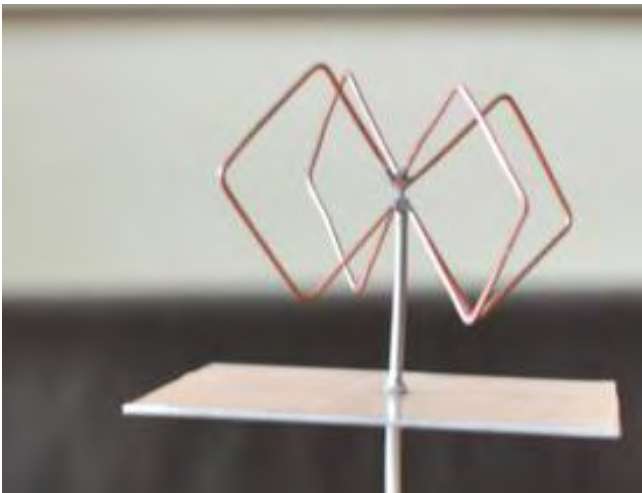


FIG. 11 – BIQUAD ANTENNA WITH REFLECTOR – PROTOTYPE

The simulation model showed bandwidth is 2350-2700MHz. The prototypes bandwidth is wider than the theoretical model (Figure 12). That can be explained with the non-idealistic characteristics of the copper wire that was used for the prototype. The prototype has higher gain then the antennas that are used in middle and low-cost Wi-Fi routers. The antenna can be placed upside-down on a sealing for better coverage in shopping centers and conference halls.

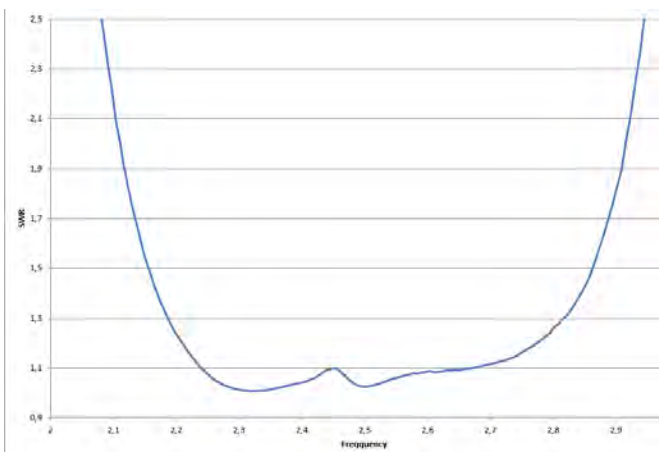


FIG. 12 – BIQUAD ANTENNA – PROTOTYPES BANDWIDTH, SWR, EXPERIMENTAL MEASUREMENTS

### III. CONCLUSION

A modified double biquad antenna with reflector has been developed and examined in this study. The article shows a new type of loop antenna, modified with reflector. Double biquad antenna with reflector has the highest gain from every

other simulation of QUAD type antennas (Table 1). Modified double biquad antenna with reflector shows satisfactory electrical performance, compact size and easiness to manufacture, which makes it a strong candidate for practical applications. It can be used as standalone antenna and in different arrays, depending of the specifical structure of a building. The innovative approach chosen for antenna modification can lead to future developments in this still relatively undeveloped area of antenna design.

TABLE 1  
COMPARATIVE ANALYSIS

Type of the antenna	Gain	Directivity in horizontal direction	Conclusion
Quad	2 dBi	Dipole type	It is not relevant to the task
Biquad	2 dBi	Dipole type	It is not relevant to the task
Double biquad	2 dBi	Omni	It is not relevant to the task
Double biquad with reflector	4.5 dBi	Omni	The directivity is

### REFERENCES

- [1] "IEEE-SA Standards Board Operations Manual". IEEE-SA. Retrieved 2015-09-13.
- [2] "ARRLWeb: Part 97 - Amateur Radio Service". American Radio Relay League. Retrieved 2010-09-27
- [3] Hall, Gerald (1991). The ARRL Antenna Book (15th edition). ARRL. p. 24. ISBN 0-87259-206-5.
- [4] evelaqua, Peter J. "Types of Antennas". Antenna Theory. Antenna-theory.com Peter Bevelaqua's private website. Retrieved June 28, 2015
- [5] Johnson, R, Jasik, H, ed. (1984). Antenna Engineering Handbook. McGraw Hill.



# Synthesis of Wideband Edge Coupled Microstrip Filters with Defected Ground Structures

Marin Nedelchev, Alexander Kolev, Ilia Iliev<sup>1</sup>

**Abstract** – The paper presents design, simulation and measurement of an edge coupled microstrip filter with a dumbbell type defected ground structure (DGS). For the design purposes, the coupling coefficients of the coupled resonators with defected ground structure were investigated using full wave electromagnetic simulator. The design graphs for the coupling coefficient used for the filter synthesis are presented. Based on the simulation results a five pole filter with a Chebyshev response was simulated, manufactured and measured. There is a very good agreement between the simulation and the measured results.

**Keywords** – microstrip coupled, resonator, coupling coefficient, defected ground, dumbbell

## I. INTRODUCTION

The emerging communication systems, allow users to communicate at ever higher speeds, which result in stringent requirements on the increase of bandwidth in which the devices have to operate. Filters play an important role in many microwave applications. But the filter response requirements become more stringent to the insertion loss, wider bandwidth, and steeper response in the stopband. In most cases the microstrip filters are designed using Chebyshev approximation.

The main problem designing wideband and ultrawideband coupled resonator filters are the tight gaps between the resonators. Then the manufacturing tolerances have serious impact on the coupling coefficients and the filter's response. One of the possible solution of this problem is using interdigital or combline filters, but their common drawback that they require a short circuit in microstrip technology. Another possible solution is using etched topologies in the ground plane of the microstrip, known as defected ground structures in edge coupled halfwave filter. This filter is proposed by S. Cohn in his famous paper [11]. It is preferred by the designer because of its clear theoretical and practical design procedure, easy adjustment, descent performance in the stopband and relatively large size in the lower GHz band. The realizable relative bandwidth of the filter is between 5 and 35%.

DGS is realized by introduction of a topology etched in the ground plane of microstrip as shown on Fig.1 [9]. It disturbs

the shielded current distribution depending on the shape and dimension of the defect. The disturbance at the shielded current distribution will influence the input impedance and the current flow of the microstrip lines. The positions and lengths of the dumbbell sections control the response and insertion loss of the designed device. DGS increases the effective values of dielectric constant of substrates ( $\epsilon_{eff}$ ), thus, decreases the wavelength and the overall length of the design.

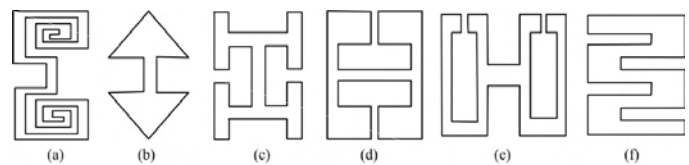


Fig.1 Types of DGS elements-(a) spiral head, (b) arrowhead-shot, (c) "H"-shape slot, (d) a square open-loop with a slot in middle section, (e) open-loop dumbbell and (f) interdigital DGS

Defected ground structures has many practical applications in microwave planar devices. They can be applied in the design of microstrip filters, microstrip antennas, power dividers, matching networks. The main effects connected to introduction of DGS are size reduction of the device [5], harmonic control [4], sharp rejection and wide stopband[2], better matching of the feed lines [6]. It is shown in [1] that the DGS can improve and control the radiation property of microstrip antenna and suppress the cross-polarized radiation from microstrip array [7].

Introduction of DGS in coupled resonator microstrip filters can increase the bandwidth and to cause multiband response. This is the reason to investigate in more details the possibilities of design of wideband microstrip filter using DGS.

The coupling enhancement scheme is based on the difference between the dielectric constants in even and odd modes of the coupled lines. The defected ground structure can control the effective dielectric constant. In even mode, the wave travels longer way and the signals slow down. In that case the phase velocity for even mode decreases. Therefore, the effective dielectric constant increases. In odd mode E-field pattern is asymmetric i.e., E-field is continuous even in presence of a DGS. So the signal path in odd mode is exactly the same as without a DGS. Hence waves do not experience any slow-wave effect and thus the dielectric constant for odd mode remains unchanged. Therefore, the overall effective dielectric constant increases with the inclusion of a DGS and thus, the coupling coefficient enhances [3,8].

This paper presents a synthesis procedure of wideband edge coupled halfwave microstrip resonators with a dumbbell type defected ground structure. The procedure follows the Chebyshev approximation and calculation the coupling

<sup>1</sup>Marin Veselinov Nedelchev, Alexander Kolev and Ilia Iliev –are with Dept. of Radiocommunication and Videotechnologies in Faculty of Telecommunication in TU –Sofia, N8, Kliment Ohridski bul., 1700 Sofia, Bulgaria. E-mail: [medelchev@tu-sofia.bg](mailto:medelchev@tu-sofia.bg)

coefficient between the resonators. A fullwave planar EM simulator is used in order to obtain the corresponding coupling coefficients. The synthesized filter is simulated, manufactured and its frequency responses are measured. There is a very good agreement between the simulated and measured results.

## II. HALF WAVE COUPLED RESONATOR WITH DUMBELL TYPE DEFECTED GROUND STRUCTURE

All the simulations, design procedures in the paper are performed for dielectric substrate FR-4 with height 1.5mm, relative dielectric constant  $\epsilon_r=4.4$  and loss tangent  $tg\delta=0.02$ .

The form of the dumbbell type defected ground structure and its dimensions is shown on Fig. 2. It consists of a main slot connected on both ends to rectangular dumbbell heads.

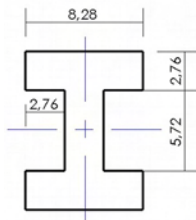


Fig.2. Form and size of the DGS (all sizes in mm)

The edge halfwave coupled resonators have width corresponding to the width of a line with a characteristics impedance of  $50\Omega$ . For the dielectric substrate FR-4, it is computed that the width of the line is  $w=2.76\text{ mm}$ .

Fig.3 shows two coupled resonators with length  $l=32.82\text{ mm}$  for center frequency  $f_0=2.4\text{ GHz}$ . The length is determined using a simulation in a full wave EM simulator.



Fig.3 Two coupled resonators with length 32.82mm and width 2.76mm

Fig.4 shows two coupled resonators with length  $l=32.82\text{ mm}$  for center frequency  $f_0=2.4\text{ GHz}$  with the DGS shown on Fig.2 in the ground plane.

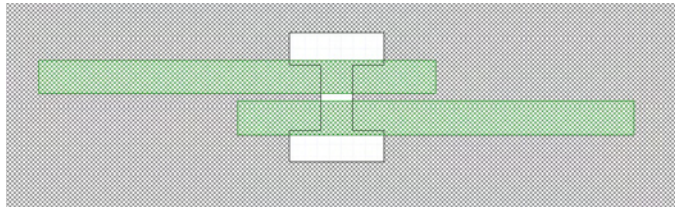


Fig.4 Two coupled resonators with length 32.82mm and width 2.76mm with DGS in the ground plane

The corresponding coupling coefficient for synchronously tuned resonators can be calculated easily by the resonance frequencies of even and odd mode [10], when the coupled resonators are overcoupled:

$$k = \frac{f_e^2 - f_0^2}{f_e^2 - f_0^2} \quad (1)$$

A full wave EM simulator based on the Method of the Moments (MoM) is used to identify the resonance frequencies in the response. Figure 5 shows the calculated coupling coefficient of the coupled resonators, when there is no DGS and when we place a dumbbell DGS with the form and dimensions shown in Fig.4.

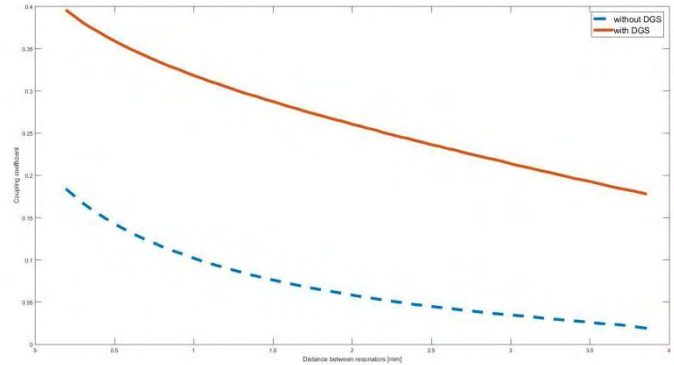


Fig.5 Coupling coefficient between halfwave resonators with (dashed line) and without DGS (solid line)

## III. DESIG OF A WIDEBAND EDGE COUPLED FILTER WITH DGS

The designed filter will have a frequency response based on the Chebishev approximation of the 5 order. Therefore the required coupling coefficients must be computed using [10]

$$k_{n,n+1} = \frac{FBW}{\sqrt{g_n \cdot g_{n+1}}} \quad (2)$$

Where FBW is the fractional bandwidth and  $g_n$   $n=0-5$  are the values of the element of the lowpass filter prototype. There are various sources of precomputed values for the elements for different pass band ripple. The current design uses  $L_r=0.1\text{ dB}$  and the values for the elements [10] are

$$g_0 = g_6 = 1$$

$$g_1 = g_5 = 1.1468$$

$$g_2 = g_4 = 1.3712$$

$$g_3 = 1.9750$$

The needed coupling coefficient are computed using (2) and are

$$k_{0,1} = k_{5,6} = 0.19563$$

$$k_{1,2} = k_{4,5} = 0.16707$$

$$k_{2,3} = k_{3,4} = 0.12731$$

Next step is to determine the needed separation between the edge coupled resonators. This is done using a full wave EM simulator and extracting the coupling coefficient when the resonators are overcoupled. After a parametric study of the DGS the extracted separation between the resonators are as follows

$$s_{0,1}=s_{5,6}=0.7\text{ mm}$$

$$s_{2,3}=s_{4,5}=1.05\text{ mm}$$

$$s_{3,4}=1.75\text{ mm}$$

For comparison without the DGS etched in the ground the plane the distances between the resonators has to be

$$s_{0,1}=s_{5,6}=0.13\text{ mm}$$

$$s_{2,3}=s_{4,5}=0.30\text{ mm}$$

$$s_{3,4}=0.65\text{ mm}$$

It is obvious how much the separation between the resonators can be increased then a DGS are inserted under the coupled lines of the filter.

Based on the previous calculation the resulted microstrip edge coupled coupled filter is shown on figure 5 a). Figure 5 b) depicts the filter after tuning with an EM simulator.

Figure 6 show the simulated return loss and insertion loss of the designed filter

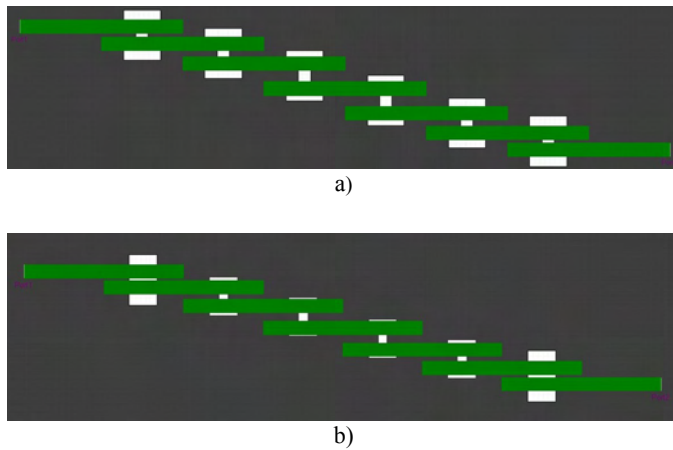


Fig 5 Topology of the designed filter a) based upon calculation b) tuned using a electromagnetic simulator

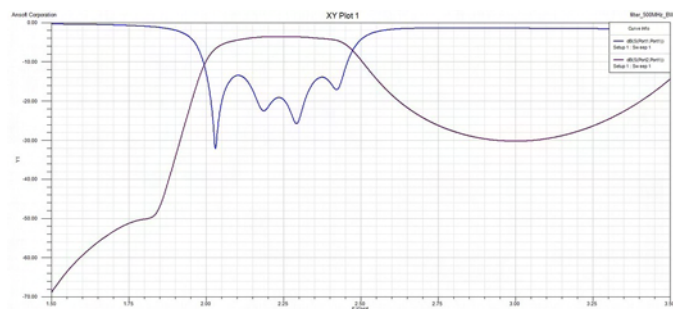
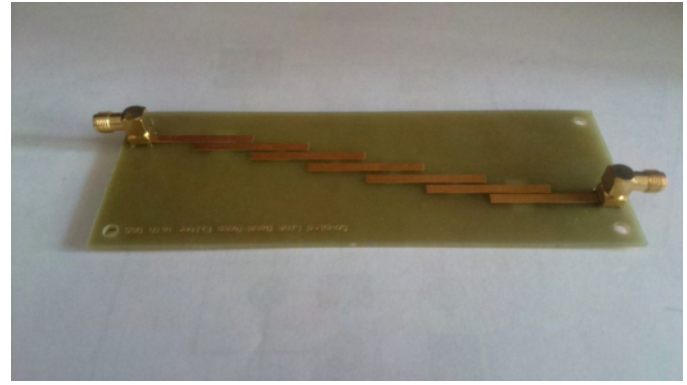
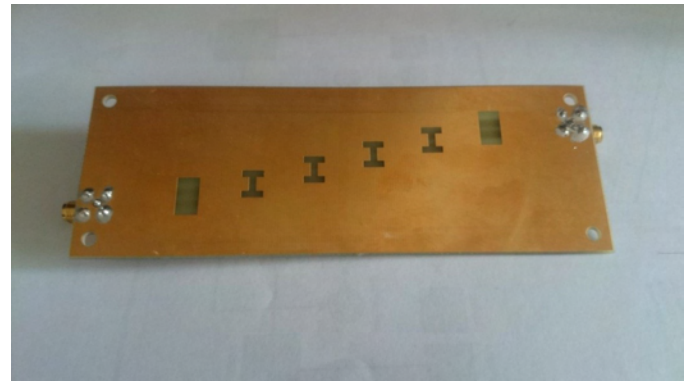


Fig.6 Simulated frequency response of the return loss and inserted loss.

After simulation the filter was fabricated. Figure 7 depicts a top and bottom view of the manufactured filter. Afterwards the return loss and the inserted loss of the manufactured filter were measured. The measurements are shown on figure 8.



a)



b)

Fig. 7 Top a) and bottom b) side of the manufactured filter.

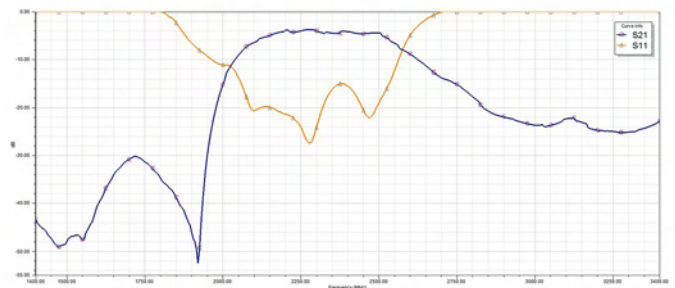


Fig.8 Measured return loss /S11/ and inserted loss /S21/

## IV. CONCLUSION

This paper presents the procedures for designing a wideband edge coupled filter using DGS. The filter was designed according to the Chebyshev approximation. A full-wave electromagnetic simulator was used to extract the required distances with regards to the coupling coefficients and to model the frequency response of the designed filter. The modeled filter was produced and the measured results are in good agreement with the ones acquired from the EM simulator.

# REFERENCES

- [1] F.Y. Zulkifi, E.T. Rahardjo and D.Hartanto, "Radiation Properties Enhancement Of Triangular Patch Microstrip Antenna Array using Hexagonal Defected Ground Structure", *Progress In Electromagnetics Research M*, Vol. 5, 101-109, 2008.
- [2] Ke LU, Guang-Ming WANG, He-Hui XU, "Compact and Sharp-Rejection Bandstop Filter Using Uniplanar Double Spiral Resonant Cells", *Missile Institute of Air Force Engineering University, Sanyuan, Shaanxi, China*.
- [3] Si-Si Chen, Lin Li and Zhi-Yong Chen, "A Novel Coupled-Defected Ground Structure with Enhanced Coupling-Coefficient and Its Application in Filter Design", *Progress In Electromagnetics Research Letters*, Vol. 52, 93-97, 2015
- [4] R.A. Pandhare, P.L. Zade and M.P. Abegaonkar, "Harmonic Control by Defected Ground Structure on Microstrip Antenna Array" *Indian Journal of Science and Technology*, Vol.8 (35), December 2015
- [5] R.A. Pandhare, P.L. Zade and M.P. Abegaonkar "Miniaturized Microstrip Patch Antenna Array at 3.8 GHz for WiMax Application" *IOSR Journal of Electronics and Communication Engineering*, Volume 10, Issue 6, Ver.I, (Nov-Dec, 2015), PP 20-24
- [6] S.Kakkar, A.P. Singh and T.S. Kamal, "Application of DGS in Microstrip Patch Antenna", *International Journal of Computer Applications* (0975-8887), *International Conference on Advancements in Engineering and Technology* 2015
- [7] D. Guha, C. Kumar and S.Pal, "Improved Cross-Polarization Characteristics of Circular Microstrip Antenna Employing Arc-Shaped Defected Ground Structure", *IEEE Antennas and Wireless Propagation Letters*, Vol.8, 2009
- [8] N. Militaru, M.G. Banciu and G.Lojewski, "Enhanced Couplings in Broadband Planar Filters with Defected Ground Structures", *Romanian Journal of Information, Science and Technology* Volume 10, Number 2, 2007, 199-212
- [9] L.H. Weng, Y.C.Guo, X.W. Shi and X.Q.Chen, "An Overview on Defected Ground Structure", *Progress In Electromagnetics Research B*, vol. 7, pp. 173-189, 2008.
- [10] J.Hong, M.J Lancaster, "Microstrip Filters for RF/Microwave Applications", John Wiley and Sons Inc, 2001.
- [11] S. B. Cohn, "Parallel Coupled Transmission Line Resonator Filters," *Microwave Theory and Techniques*, *IRE Transactions on* , vol. 6, no. 2, pp. 223-231, April 1958



# Modified Fractal Super J-pole Antenna

Boncho Bonev<sup>1</sup>, Peter Petkov<sup>2</sup> and Nina Spassova<sup>3</sup>

**Abstract** – In this paper a super J-pole antenna with two modified fractal Koch curve shaped elements is proposed. In order to increase the bandwidth and to provide proper phasing of the currents in the antenna elements, the length of some elements of the fractal has been changed. The antenna performances (radiation patterns, VSWR and gain) are simulated and optimized for GSM-1800, UMTS-2100 and Wi-Fi 2,4 GHz bands. Proposed Koch curve super J-pole antenna features relatively small size and high gain.

**Keywords** – Super J-pole, Fractal, Koch curve, Wideband, Multiband.

## I. INTRODUCTION

One of the biggest problems in wireless communications is the design of the antennas. They should work in wide frequency band while being small physically. The rule in such cases, as states in [1] is that the bandwidth of an antenna enclosed in a sphere of radius  $r$  can be improved only if the antenna utilizes efficiently, with its geometrical configuration, the available volume within the sphere. Several antenna types meet this requirement and fractal antennas are one of them [2-4]. They bear such name after the concept of fractals.

In our previous works [5,6] these concepts have been presented on other types of fractal antennas - research on fractal antennas performance, fractal antennas for some specific applications - Wi-Fi hotspots, digital TV reception, etc. In this paper are proposed and analyzed a super J-pole modified fractal antenna based on Koch curve. This approach represents a wideband and multiband performance, acceptable total gain and relatively small size. The main antenna parameters - radiation pattern, return loss and antenna gain were simulated.

## II. THEORETICAL ANALYSIS

### A. J-pole and Super J-pole Antenna

The J-pole antenna (Fig. 1) is an omnidirectional antenna, so called because of its similarity with the letter "J". It is formed by a vertical half-wavelength radiator end-fed by a quarter-wavelength matching stub. One of the advantages of this antenna comes from the fact, that the lower end of the

matching stub can be used for direct grounding. The impedance matching of the J-pole antenna is achieved by moving the feed point between the radiator and the stub short, back and forth, until point where they resonate with the

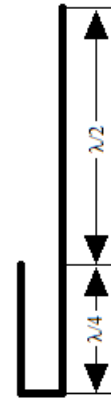


Fig. 1. J-pole antenna

transmitter frequency is reached[7,8].

When designing a J-pole antenna there are several particularities that should be taken into account [2]. First of all to achieve negligible radiation from the matching stub, the spacing between its two conductors must be commensurable with the wavelength. The current in the two parallel lines must be with phase difference of  $180^\circ$  and of course the open end of the stub must have infinite impedance while the connected one should be very high, but not infinite.

The Super J-pole antenna (Fig. 2) is a variation of the J-pole. It still has the basic construction of the J antenna, but also has an additional half-wavelength radiator which is connected to the first one with a quarter wave-length phasing stub. The phasing stub is used to unify the phases of the currents in the two radiators. This type of antenna provides more gain compared to the classical J-pole version.

The design of the Super J-pole antenna doesn't differ very much from the J-pole's. Same conditions, mentioned above, must be met to obtain the maximum performance from this antenna construction. Also the space between the two collinear radiators should be very small in order to derive the biggest profits of their collinear placement.

### B. The Koch curve fractal

Geometrical figures which are formed by miniaturized identically scaled copies of themselves are called fractals. The operation used to obtain their form is iteration – endlessly repetition of certain steps. The more times we do these steps, the more accurate the fractal will be, but also it will become more complex [3].

<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>Peter Petkov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria. E-mail: pjpetkov@tu-sofia.bg.

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

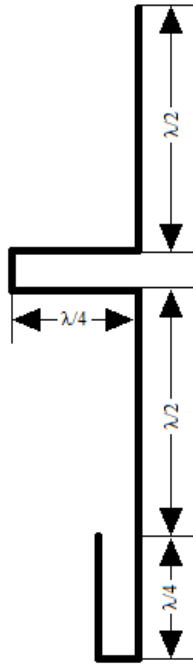


Fig. 2. Super J-pole antenna

One of the fractals used in antenna theory and design to miniaturize the antenna elements and help utilize the frequency band more efficiently is the Koch curve. To form it we start with a straight line, initiator, and split it into three equal parts. After that we build an equilateral triangle over the middle part and remove its foundation (the middle section of the line) - first iteration ( $n=1$ ) in Fig.3. We do the same thing with the remaining two sides of the triangle and the two parts of the initiator, as shown below to obtain the 2-nd order Koch curve -  $n=2$ . We can continue these steps to obtain a fractal with the desired form, precision and complexity.

The total length of the Koch curve can be given with equation [4]

$$L = L_0 \left( \frac{4}{3} \right)^n, \quad (1)$$

where  $L_0$  is the 0-th order length, also called initiator, and  $n$  is the iteration number.

### III. DESIGN AND SIMULATIONS

As initial step a super J-pole antenna is designed using a known calculation approach [2]. The two elements of the antenna is shaped as 1-st order Koch curve. Then the length of the elements in two triangle parts of each radiator of the antenna has been changed to provide the proper phasing of the currents in the antenna elements. The optimizations and simulations of antenna parameters (VSWR and antenna peak gain) are performed with 4nec2 moments-method software ([www.qsl.net/4nec2](http://www.qsl.net/4nec2)) for the frequency band 1,71-2,49 GHz. The simulated and optimized antenna model is presented on

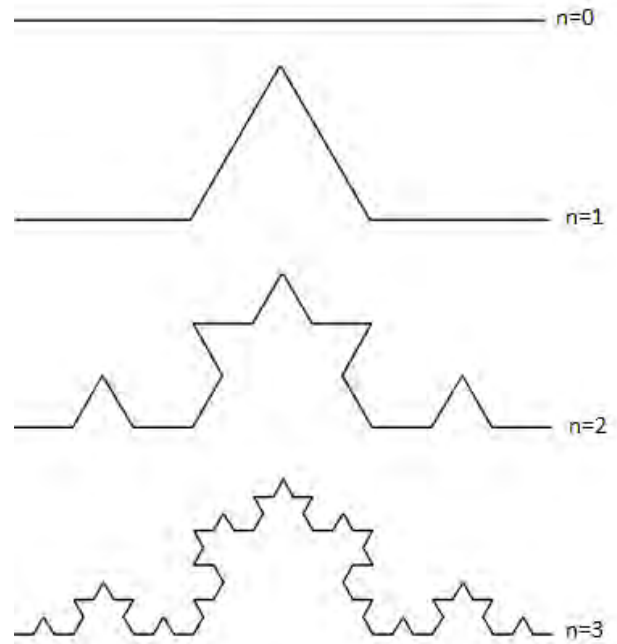


Fig. 3. Generation of the Koch curve

Fig. 4 and the elements length in mm are given in Table I. The antenna elements are round wires with radius 1.75mm.

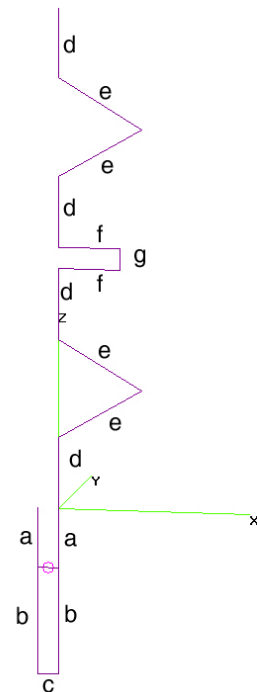


Fig.4. Proposed design of super J-pole fractal antenna

TABLE I  
ANTENNA ELEMENTS LENGTH

a	b	c	d	e	f	g
30	54	11	36	50	32	11

In simulations up to 200 segments per half wave length are used which provides the necessary accuracy of the simulations.

The proposed antenna has two frequency bands of

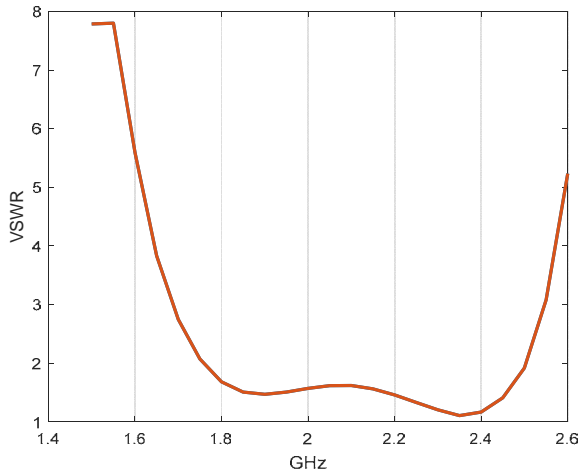


Fig. 5. Simulated VSWR for 1-st frequency band

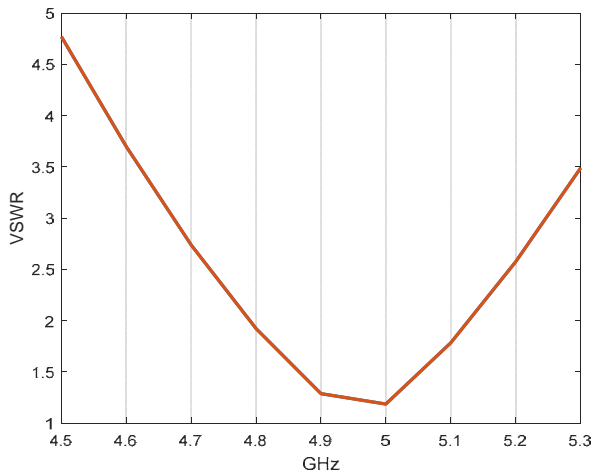


Fig. 6. Simulated VSWR for 2-nd frequency band

operation where the VSWR is lower than two (see Fig. 5 and Fig. 6). The first one (Fig. 5) is from 1.735 GHz to 2.52 GHz and the second one (Fig. 6) is from 4.785 GHz to 5.125 GHz. The bandwidth for the 1-st frequency band is 37,4 % and 6,9 % for the 2-nd one for the proposed fractal antenna while the classical antennas (non fractal) in most cases has not usable 2-nd frequency band of operation.

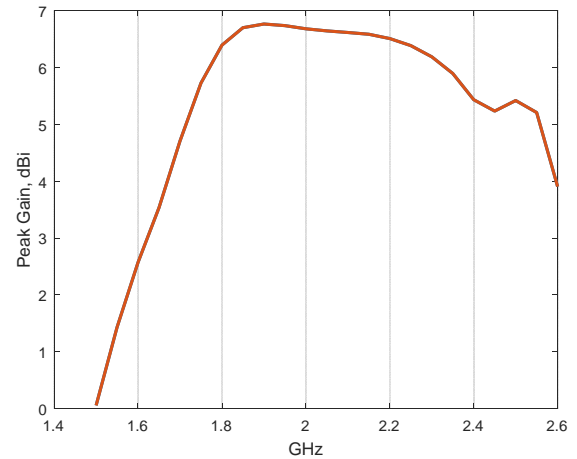


Fig. 7. Simulated antenna peak gain for 1-st frequency band

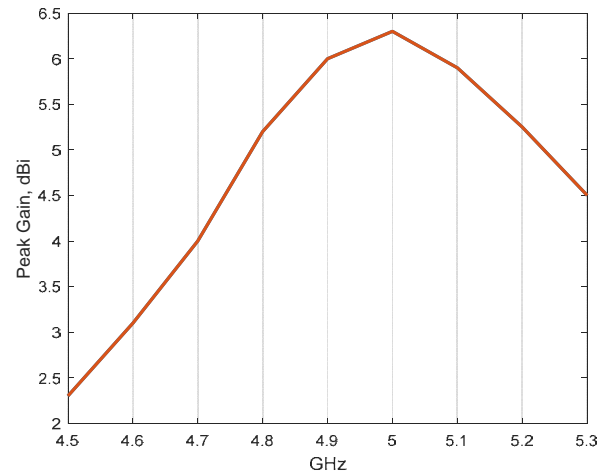


Fig. 8. Simulated antenna peak gain for 2-nd frequency band

The peak antenna gain for both bands of operation is displayed on Fig. 7 and Fig. 8. It is noticeable that the gain easily exceeds 5.2dBi, which is almost the maximum achievable with a single wire antenna. The maximal gain for first operation band is 6.8dBi and for the 2-nd one is 6.3dBi. It is obvious that the peak gain stays relatively constant.

In other hand the antenna shows directional properties (Fig.9 and Fig.10), because of the specific current distribution, especially in the upper band.

It has almost omnidirectional horizontal radiation pattern for the lower band of operation and directional vertical pattern. In the vertical plane, the antenna acts as a linear antenna array formed by several dipoles. The phasing section and fractal elements length modification provides a proper phase shift in a large frequency band, therefore the vertical pattern doesn't vary a lot with the frequency.

For the upper frequency band the antenna construction acts also as linear array, however the identical phase distribution of the currents is provided by more not tightly controlled manner (i.e not optimized).

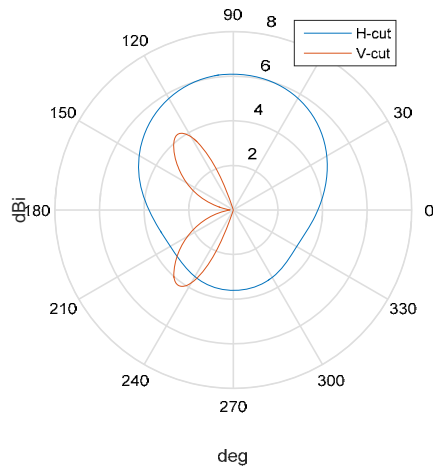


Fig. 9. Simulated radiation patterns for 2GHz.  
(Z-axis of the antenna is aligned with 180-0 deg line)

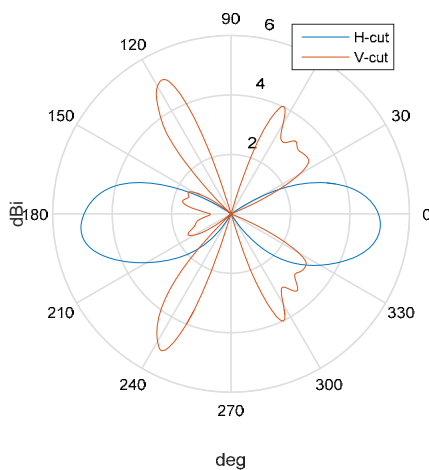


Fig. 10. Simulated radiation patterns for 5 GHz.  
(Z-axis of the antenna is aligned with 180-0 deg line)

#### IV. CONCLUSION

A fractal super J-pole antenna with radiation elements shaped as Koch curve is introduced in this paper. Fractal shape is used in order to reduce antenna size and to achieve the multiband performance. For increasing the main operation frequency band fractal elements with different length is used. In this case the 37.4 % frequency band is achieved. The second advantage of the proposed fractal antenna is the dual-band performance.

With additional frequency tuning the proposed antenna can be used, among many others, for cellular repeaters - GSM 1800, UMTS 2100, Wi-Fi hotspots, etc.

#### REFERENCES

- [1] C. Balanis, "Antenna theory" (Third edition), John Wiley & Sons, 2005.
- [2] E. A. El-khouly, H. A. Ghali, "High Gain Fractal Based Antenna", Proc of Antennas and Propagation Conference 2008, Loughborough, United Kingdom, pp. 405-408, March 2008.
- [3] Balanis, C., "Modern antenna handbook", John Wiley & Sons, 2008
- [4] M. N. A. Karim, M. K. A. Rahim, M. Irfan, T. Masri, "Fractal Koch Curve for UHF Band Application", Asia-Pacific Conference on Applied Electromagnetics - APACE 2007, pp. 1-4, December 2007.
- [5] P. Petkov, M. Kolev, B. Bonev, "Fractal Yagi Antenna", 2015 Conference of Microwave Techniques (COMITE), pp. 1-3, Pardubice, Czech Republic, April 2015.
- [6] P. Petkov, B. Bonev, "Analysis of a Modified Sierpinski Gasket Antenna for Wi-Fi Applications", Proc. of 24th International Conference RADIOELEKTRONIKA'2014, pp. 1-3, Bratislava, Slovak Republic, April 2014.
- [7] Beggerow, Hans (1909). "Zeppelin Antenna" (PDF)
- [8] Richardson, D. (K6MHE) "The J-Pole Revisited", CQ Magazine, pp 34-41, March 1998.



# Experimental Determination of 2biQUAD Antenna Radiation Pattern Using CRM Method

Kalina Kalinovska<sup>1</sup> and Kliment Angelov<sup>2</sup>

**Abstract** – In this article was made experimental determination of radiation pattern to 2biQUAD antenna with reflector, designed for untargeted applications in ISM bands. An interest in the article is the radiation pattern in vertical direction. The results of this type of measurements can be used to provide radio coverage for different radio networks.

**Keywords** – 2biQUAD, radiation pattern experimental mesurments, antenna, Omni-antenna,

## I. INTRODUCTION

In today's world of unlimited communication options are available many technologies using radio signals. They are providing coverage to subscribers with relevant services, while giving the opportunity to be mobile. Examples of such technologies are mobile cellular networks of second and third generation, Wi-Fi, WiMAX and others. Parallel to that the radio communications are used in different technologies for access control, RFID and others using radio signals in the ISM bands. The transmission of information in all of these technologies is made by providing radio in separate zones, aiming at the optimum use of radio resources. For example, Wi-Fi is using different access points (HotSpot), served by antennas that can be with narrow radiation pattern beam or with omni-directional radiation pattern. The directivity of the antenna must be appropriate to comply with the necessary radiation pattern and consistent with the structure of the building that the Wi-Fi coverage will be developed.

Solution for shopping centres and conference halls with many Wi-Fi users is 2biQUAD antenna with reflector. It can realize relatively uniform radiation pattern in a horizontal direction and the beam is more narrow in vertical plane. A member of the team of authors of this article had done a simulation design of 2biQUAD antenna with reflector for Wi-Fi bandwidth, but it is still in need for experimental determination of the radiation pattern to prove the theoretical model.

This article describes the experimental determination with Continuous Rotation Method (CRM) [1] of the radiation pattern of a 2biQUAD antenna with reflector for the purposes of Wi-Fi technology or other applications in different ISM frequency bands. CRM method is used to capture the radiation

pattern of an antenna in determent plane. The antenna is put on a specialized stand which is rotated several times in this plane. In this way it is realized repeatedly reading the signal from the antenna, which describes the radiation pattern. With the aid of mathematical processing the received periodically repeated data is processed to minimize the random errors and improve the accuracy of the experiment.

## II. DESCRIPTION OF THE PROBLEM

### A. The 2biQUAD antenna conception

It is known different types of loop antennas with different loop shapes [2]. Some of them have a shape close to a square, which can be used in the range of HF. These antennas have uneven radiation pattern and can not satisfy the condition of omni-directional antenna for broadcasting in one plane. Combining two biQUAD antennas located in two perpendicular planes would contribute to the alignment of the omni-directional radiation pattern. On the other hand, by placing the reflector above the structure we can obtained a limited radiation pattern in vertical plane. Such an antenna can be used to realize the Wi-fi HotSpot in a conference halls or shopping centres, using a ceiling mounting. Figure. 1 shows a general view of 2biQUAD antenna with reflector.

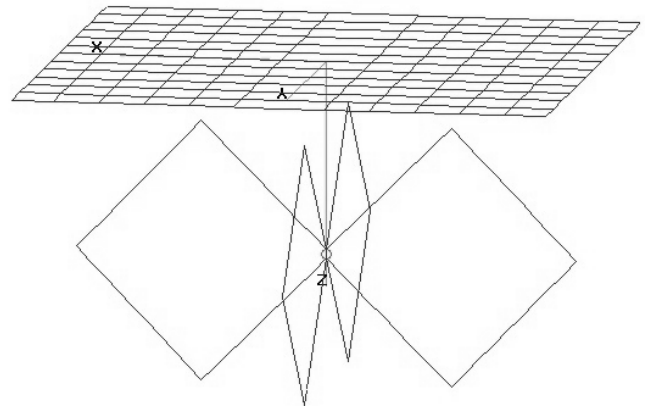


Figure 1 - 2biQUAD antenna common view

Based on the idea of two 2biQUAD antenna with reflector were made simulations of two prototypes. One is for operating frequencies matching Wi-Fi technology, and the other - intended for use in ISM band at 433 MHz. Figure 2 shows a simulation resulting radiation pattern in three-dimensional space, and Figure 3 - diagrams directional horizontal and vertical plane.

<sup>1</sup>Kalina Kalinovska is a student at Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria. kalina.kalinovska@gmail.com

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

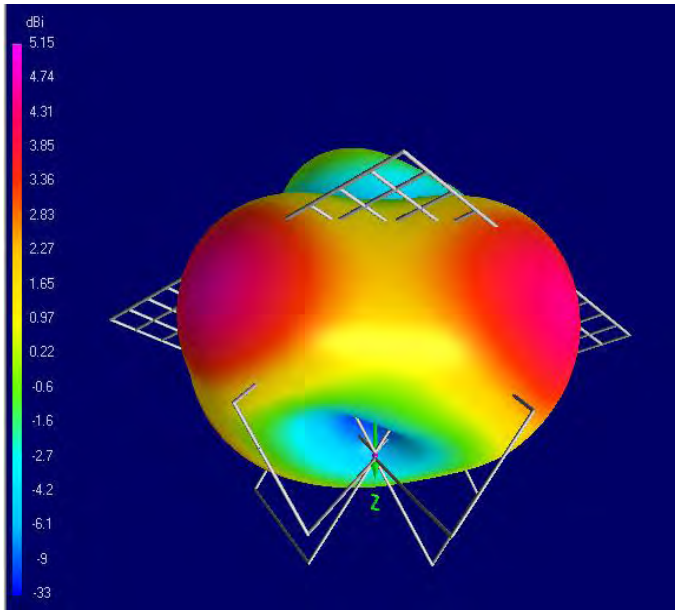


Figure 2 - 2biQUAD antenna simulated 3D radiation pattern

The defining of horizontal and vertical plane of the antenna is based on the above mentioned reasons for ceiling mounting in case of HotSpot realizations.

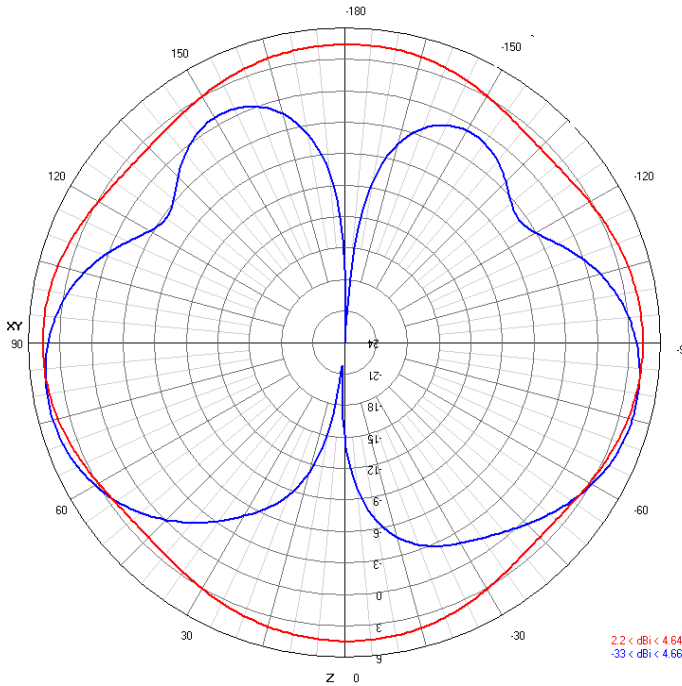


Figure 3 - 2biQUAD antenna simulated radiation pattern (horizontal and vertical)

After the simulation optimizing were realized two prototypes. One for Wi-Fi technology (2,45GHz) and second one for 433.05-434.79MHz frequency band. Their radiation pattern was reserched using CRM method.

### B. The CRM method

Members of the team of authors of this article have developed a method for measuring diagrams of directional antennas in which is used a long multi rotation with angular velocity  $\omega$  of the test antenna (AUT). This method accumulates cyclic data with sampling time  $T_s$ , through constant periodic received signal. The subsequent mathematical processing can obtain the main period of repetition and reduce random error in determining the radiation pattern. An advantage of the method is also that the current angular position is set as the relative rotation of the AUT with accuracy that can be determined by [1]:

$$\theta_{\min} = 12.T_s.\omega. \quad (1)$$

The accuracy of the method is increasing with the number of turns of the stand. If the sampling frequency is not a multiple of the rotation frequency with angular velocity  $\omega$  then each discrete reporting occurs in different point in the radiation pattern. With this way every turn will measure unique point of the polar coordinate system set for the radiation pattern. After mathematical processing the different discrete measurements will form one radiation pattern. A similar approach is appropriate for measuring the radiation pattern of 2biQUAD antenna with reflector. Figure 4 shows the principled scheme of the experimental set. In this experiment an interest is the radiation pattern in vertical plane, since the horizontal is expected to differ minimally from omidirectional.

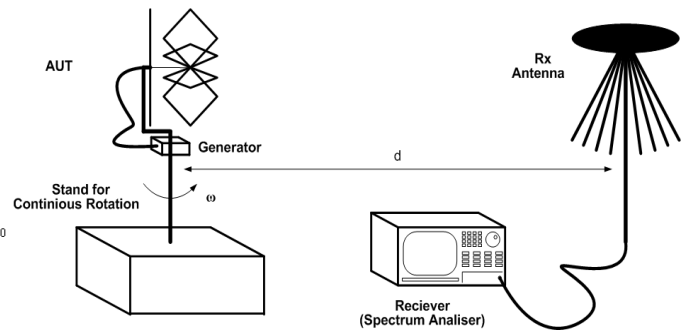


Figure 4 - Scheme of Test Stand for Radiation Pattern Measurements of 2biQUAD antenna in vertical plane

In Table 1 is given initial parameters for capturing the radiation pattern of the physical prototypes. The distance between the Tx and Rx antenna must be more than the near field of the antenna (2) [4][5]:

$$d > \frac{2.D^2}{\lambda}, \quad (2)$$

In Figure 5 is shown a photograph of the actual implementation of the test. In it 2biQUAD antenna with reflector is placed on a stand witch provide constant rotaition.

To reduce external interference the test with prototype on 433,92MHz was held on the roof of one of the buildings of the Technical University - Sofia.



Figure 5 – Photograph of Test Implementation

TABLE I  
TEST INITIAL PARAMETERS

AUT	Type	2biQUAD
	Maximal Dimension	0,6 m
	Theoretical Directivity	5,15 dBi
Transmitter	Frequency	433,92 MHz
	RF Power	0 dBm
Measuring Distance		10 m
Rotation speed		~54 rpm
Receiving Equipment	Spectrum Analyser	GSP-830
	Receiving Antenna	Omnidirectional
	Sampling Speed	1s <sup>-1</sup>

### III. RESULTS

Using the above described method was reported CRM data. This data is describing 10 rotations of the 2biQUAD antenna with reflector. Figure 6 shows the periodic variation of the received signal for the rotations of the prototype with working frequency 433,92MHz.

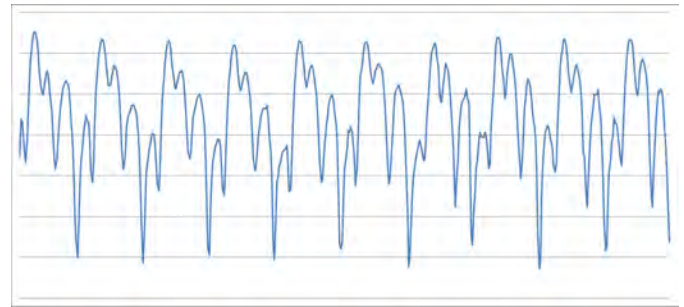


Figure 6 - Received Signal on Periods of Ten Revolutions of AUT

There was made a mathematical determination of the period of repetition of the received signal. The method is also used to reduce random errors that occurs in the experimental determined radiation pattern. For the calculations was used the programming environment of MatLab [6]. Figure 7 shows the experimentally obtained radiation pattern in vertical plane for the prototype in 433,92 MHz frequency band.

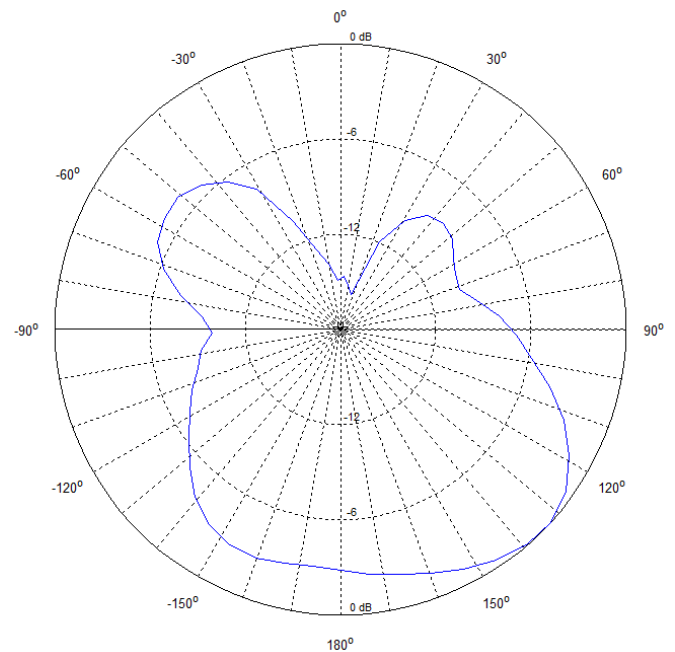


Figure 7. Experimental Radiation Pattern of AUT @ 433,92 MHz

Similarly is the experimental radiation pattern of the prototype for Wi-Fi frequency band. A graphic is shown in Figure 8. The radiation pattern varies because of the different conditions that the experiment was made. This experiment was held in a laboratory because of the weather conditions. There were many reflections from the walls and etc.

#### IV. CONCLUSION

Based on the experimental results the following conclusions can be made:

- 2biQUAD antenna with reflector has a omni-directional radiation pattern in horizontal plane;
- Despite some irregularities in experimental readings the results confirmed the simulations of the diagram of this type of antenna in a vertical plane,
- 2biQUAD antenna reflector is suitable in case of installing HotSpot with ceiling mounting;

#### ACKNOWLEDGEMENT

In the realization of the experimental research shown in this paper, the measurement equipment was provided by TEKOMA ENGINEERING Ltd and was acquired with the support of the Operational Program „Development of the Competitiveness of the Bulgarian Economy” 2007-2013, co-financed by the European Union through the European Regional Development Fund and the state budget of Republic of Bulgaria under contract BG161PO003-1.1.05-0226-C0001.

#### REFERENCES

- [1] Gechev M., K. Angelov, B. Kehayov, S. Denev, S. Kremenski, Continuous Revolution Method for Antenna Radiation Pattern Measurements, L International Scientific Conference On Information, Communication And Energy Systems And Technologies ICEST 2015, Proceedings of Papers Vol. 1 (pp 143-146), 24 - 26 June 2015, Sofia, Bulgaria.
- [2] Jonson, R. C., Antenna Engineering Handbook, 3-rd ed., McGraw Hill, 1993.
- [3] Collin R. E., Antennas and Radiowave Propagation, McGraw-Hill, 1985.
- [4] Antenna Measurement Theory, Introduction to Antenna Measurement, Reprinted with the permission of ORBIT/FR Inc., [www.orbitfr.com](http://www.orbitfr.com)
- [5] Foegelle M. D., Antenna Pattern Measurement: Concepts and Techniques, Reprinted from Compliance Engineering, Annual Reference Guide 2002 • Copyright © 2002 Canon Communications LLC
- [6] MATLAB IMAGE PROCESSING TOOLBOX. User's Guide, "The Math - Works Inc.", 2000. [www.mathworks.com](http://www.mathworks.com)

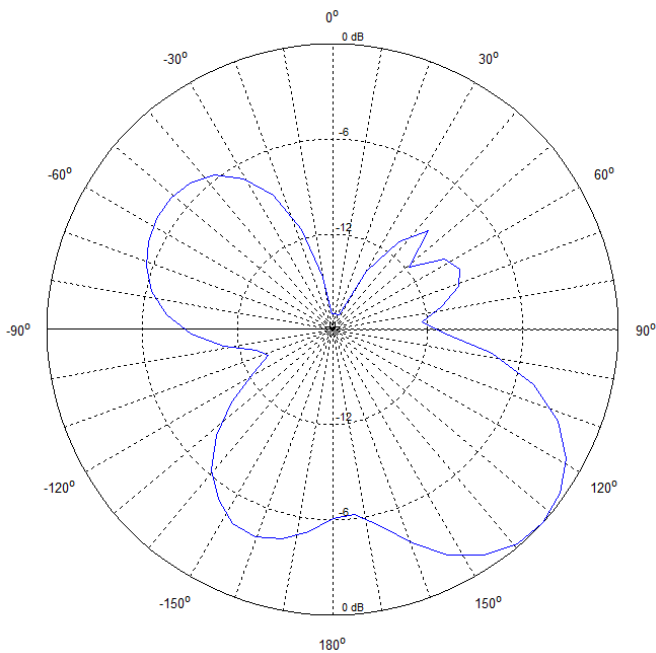


Figure 8 - Experimental Radiation Pattern of AUT @ 2,45 GHz

The obtained experimental radiation pattern of 2biQUAD antenna with reflector shows that it is suitable for the proposed use, which provides radio coverage of a HotSpots. Figure 9 shows clearly the idea of using this type of antenna to provide Wi-Fi access.

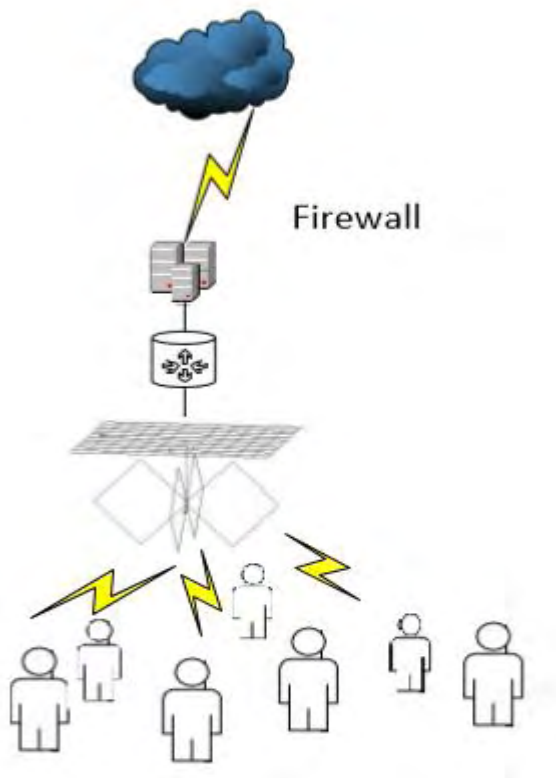


Figure 9 – An example of practice application of 2biQUAD antenna with reflector in Wi-Fi frequency band



# Infrared investigation on the thermal field in the case of influence of low frequency magnetic signals on live tissue

Kalin Dimitrov <sup>1</sup>

**Abstract** – This paper presents a study on human skin surface warming during physiotherapy with electromagnetic field. The case is for low-frequency magnetic signal therapy with frequencies lower than 100Hz. For the experiment was used the connection between the optical radiation of the objects themselves and their temperature. The study was conducted under clinical conditions during regular physiotherapy sessions with patients. The measurements were done simultaneously at many points by using an infrared camera. The conditions in the rooms and the condition of the patients were taken into account. The emissivity coefficients of the objects studied were also considered.

**Keywords** – Infrared thermography, medical applications, magneto therapy.

## I. INTRODUCTION

With the development of technology every day we are facing new perspectives in health care – both diagnostically and therapeutically. What is observed in the last few years is an increasingly more significant implementation of infrared thermal-imaging-based medical apparatus, generally in the diagnostic sphere. However, the potential of this technology in contemporary medicine is still to be explored in further detail. The current paper is an attempt to broaden the view and deepen the understanding of some of the existing therapies (magneto therapy with low frequency magnetic field) through the analysis of data provided by a thermal camera [7-13].

Physiotherapy is a branch of medicine. Its power lies in using energy sources directed to particular areas of the body. This energy is directed in a specific way, for the prophylactics or treatment of different diseases. The sources can be natural or artificially created by people. In some cases, physiotherapy is applied as an accompanying treatment in aid of the main one. It is conducted by doctors (therapists) which define the parameters of the procedures and the course of the treatment very precisely. Physiotherapy is applied for different ages and diseases and it provides wide possibilities for influencing the human organism (electrotherapy, inductothermia, ultra-high frequencies treatment, aero-ionisation therapy, magnetic-therapy, ultra-sound therapy, lasers, magnet and laser, light-treatment, electrophoresis, diadynamic therapy, amplipulse

therapy in ophthalmology, transcranial and transvertebral micropolarization, muscle toning, cryotherapy, inhalers, oxygen therapy, water treatment, warm-treatment, etc).

The effects on the human body caused by high intensity electric fields, such as muscle stimulation, electroporation and heating, are relatively well studied. We are interested in the effects caused by low intensity magnetic fields. [14-24]. The biological processes which occur are mainly related to signal transduction at the cell membrane and subsequent biochemical processes. The physical mechanisms of conversion from an electric or magnetic field to biochemical process are however subject of science studies at the present time. It is also a subject of study which aspect of the applied or environmental field is important. Properties such as duration, polarization and harmonic content of the signal, may turn out to be as important as amplitude. The effectiveness of the applied field will also be influenced by the biological, chemical and mechanical condition of the organism. A few experimentally observed results show specific biological effects which are caused directly by the magnetic field or by the electric field that is induced in the tissue by a time varying magnetic field. For fields larger than about 100  $\mu$ T direct effects have been suggested and the presence of magnetite particles in some tissues may also play a biological role. Experiments suggest that effects of induced electric fields are particularly likely when their magnitudes at the tissue or cell level are relatively large and when multiple cells are connected by gap junctions. Most proteins have a small net electrostatic charge and nucleic acids are polyelectrolytes with large net charge and are surrounded by counterions. Induced fields mainly affect the moving ions on the cell surface and possibly the structural fluctuations of nucleic acid molecules. The mechanisms for direct interaction of time varying magnetic fields with biological processes are the subject of many studies.

Other experimental results have suggested that the relative magnitude and direction of a static magnetic field (such as that of the Earth) can determine the biological effectiveness of a simultaneously present alternating field in some biological systems under controlled laboratory conditions.

Very few scientific studies consider a simultaneous investigation of the effects caused by a low frequency magnetic field with infrared thermography [25-27].

## II. CURRENT STATE OF THE PROBLEM

In this paper, we are going to focus on the application of magnetic-therapy. More specifically, we are going to consider the extent to which we can neglect the surface warming of the

<sup>1</sup>Kalin Dimitrov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria., kld@tu-sofia.bg

skin during the conducting of this therapy [15,17,18,20,21,23,24].

The mechanism of this type of therapy is based on a flow of electrical charges causing a net flow of ionic current for basic cellular restoration activities. Many studies show that healing effects are observed: cytoprotection of cells and the stimulation of growth factor synthesis. Contemporary devices can generate signals with a different shape, frequency and length. The frequencies used are between 1 and 100 Hz, magnetic flux density being up to 100 mT. There are 3 established physical mechanisms through which pulsed magnetic fields interact with living matter: magnetic induction, magnetomechanical effects and electronic interactions, which is thought to be responsible for the vasodilatation, analgesie, anti-inflammatory, anti-edematous and spasmolytic activity and healing acceleration. There is an overall improvement of the microcirculatory processes and local blood circulation.

Depending on the cases (diseases, people), suitable inductors are used.

Usually the period of the therapy includes 20 procedures every day or every other day. In the case of special medical recommendation it is possible to do two procedures per day (in the morning and in the afternoon).

The magneto therapy is well combined with other physical factors (laser therapy, interfered currents, ultrasound therapy, water and mud treatments). The magnetic field can be well combined with drug treatment, too.

### III. EXPERIMENTS

The thermal camera used for data collection was FLIR E40, with thermal sensitivity of  $< 0.07^{\circ}\text{C}$  and temperature range of  $(-20^{\circ}\text{C}$  to  $650^{\circ}\text{C})$  [28]. For maximum accuracy, the camera was fixed on a stand and movement of the object was avoided. All examinations were performed in a sitting position in a quiet room at a constant room temperature of  $20 \pm 0.5^{\circ}\text{C}$  following an acclimatization period of 20 min keeping the hands free of any contact to the rest of the body or other objects [29]. Equally, the relative humidity showed stable values over time (50%). All temperature maps were archived using the internal SD memory. For the final analysis temperature values were determined and given in degrees Celsius ( $^{\circ}\text{C}$ ). All images were corrected using an emissivity factor of 0.98 for the human skin and appropriate emissivity factors for the fabric and plastic, respectively 0.77 and 0.91. For the human hair emissivity factor we assume the value of 0.98.

In this study, a series of images was made between certain intervals of time. The infrared pictures were taken at intervals of 30 sec. for 15 minutes. The therapy duration was determined by the doctor in charge. The infrared pictures do not influence the main therapy, so there are no special requirements on the part of doctors apart from the general requirements for hospital hygiene and rules.

The processing was made with the original software of the camera - version FLIR Tools+ 5.3.1 (5.3.15320.1002). The setting is shown in fig.1.

During the experiment were taken 30 pictures. The first one

is shown in fig. 2.



Fig. 1. Therapeutic procedure

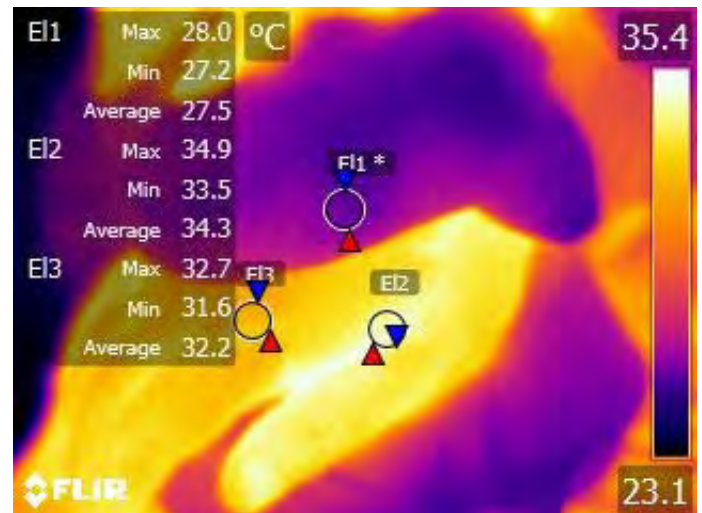


Fig. 2. An image from the magneto therapy series

Three ellipses of interest were selected: E1 – on the textile wrapping of the magnet; E2 – in the area of the most intense thermal radiation of the hand; E3 – in the area closest to the metacarpophalangeal joint. We selected ellipses, not points, to eliminate probable human movements and noises. For the areas where the human skin is observed, is selected a radiation coefficient 0.98 and for the fabric-wrapped magnet is selected a radiation coefficient 0.77 [12].

### IV. RESULTS

The first calculations, of course, are related to the mathematical expectation, dispersion and standard deviation. By using the respective formulae [30,31], we derive (Table I):

## A. MAGNETO THERAPY

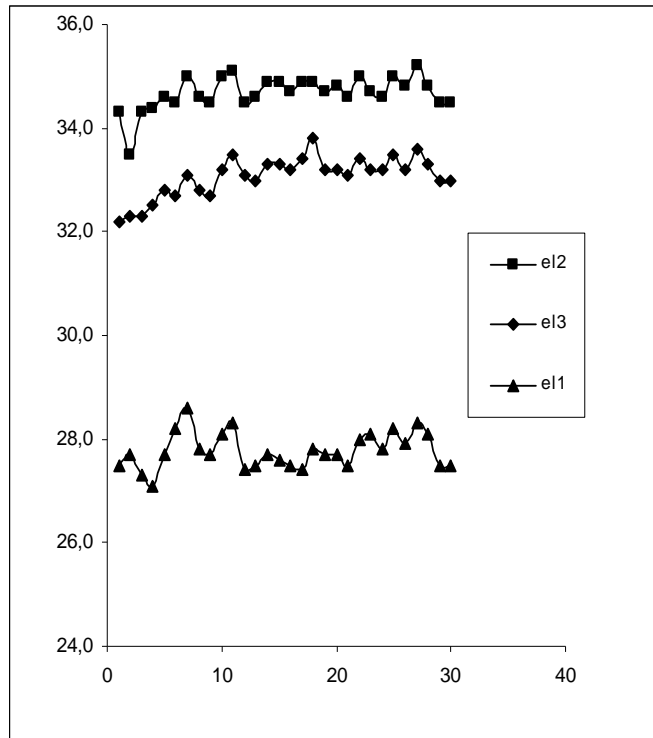


Fig. 3. Graphic representation of measurements during the experiment with magneto therapy (x-axis – number of experiment; y-axis – temperature)

TABLE I. CALCULATION OF THE AVERAGE VALUES, DISPERSIONS AND STANDARD DEVIATIONS

	EI1	EI2	EI3
AVG	27,773333	34,68	33,07
VAR	0,1192644	0,105103	0,151138
sqrt(VAR)	0,3453467	0,324197	0,388765

## V. CONCLUSION

The analyzed data gives one more perspective of the countless applications of thermal imaging cameras for medical purposes. Such analysis can be very useful in providing a new understanding of widely used technology in terms of safety and effectiveness in physiotherapy. Since IR imaging is non-invasive and does not bear any risks, it can be implemented wherever it is useful and possible. It also carries a great potential for diagnostic medicine to search for different connections between the points of the temperature filed and completely different phenomena.

## ACKNOWLEDGEMENT

This research was supported by the personnel of Studentska Polyclinic and Technical University of Sofia. We would like to thank them for facilitating this research and providing the appliances to analyze.

We would also like to thank Prof. Dimitar Dimitrov for assisting whenever needed and for providing comments that greatly improved this paper.

## REFERENCES

- [1] P. Childs, (et. all), *Practical Temperature Measurement*, Butterworth Heinemann, Elsevier, 2001.
- [2] I. Fernández-Cuevas, J. C. Bouzas Marins, J. A. Lastras, P. M. G. Carmona, S. P. Cano, M. Á. García-Concepción, M. Sillero-Quintana, "Classification of factors influencing the use of infrared thermography in humans: A review", *Elsevier, Infrared Physics & Technology*, Elsevier, 71, pp.28–55, 2015.
- [3] R. Fuksis, M. Greitans, O. Nikisins, M. Pudzs, "Infrared Imaging System for Analysis of Blood Vessel Structure", *Electronics and Electrical Engineering*, No.1(97), pp.45-48, 2010.
- [4] A. Szentkuti, H. S. Kavanagh, S. Grazio, "Infrared thermography and image analysis for biomedical use", *Periodicum Biologorum*, Vol. 113, No. 4, pp.385–392, 2011.
- [5] S. Pleshkova, A. Bekyarski, K. Peeva, "Testing Thermal Images Characteristics for Thermal Images Quality Estimation", *Latest Trends on Systems, 18th International Conference on Systems*, Vol.1, pp.251-256, 2014.
- [6] S. Pleshkova, A. Bekyarski, "Motion Detection in Thermal Images Sequence Using Wigner Distributions", *Latest Trends on Systems, 18th International Conference on Systems*, Vol.1, pp.153-156, 2014.
- [7] J. Serup, G. B. E. Jemec, G. L. Grove, *Handbook of Non-Invasive Methods and the Skin*, Second Edition, CRC Press, 2006.
- [8] B. Jones, "A reappraisal of the use of infrared thermal image analysis in medicine", 17(6), *IEEE Trans Med Imaging*, pp.1019-27, 1998.
- [9] F. Ring, A. Jung, J. Žuber, *Infrared Imaging*, IOP, 2015.
- [10] S. Nudelman, *Nuclear Medicine, Ultrasonics, and Thermography*, Springer Science & Business Media, 2013.
- [11] M. Diakides, J.D. Bronzino, D. R. Peterson, *Medical Infrared Imaging: Principles and Practices*, CRC Press, 2012.
- [12] Y. Houdas, E.F.J. Ring, *Human Body Temperature: Its Measurement and Regulation*, Springer Science & Business Media, 2013.
- [13] E. Salazar-López, E. Domínguez, V. Juárez Ramos, J. de la Fuente, A. Meins, O. Iborra, G. Gálvez, M.A. Rodríguez-Artacho, E. Gómez-Milán, "The mental and subjective skin: Emotion, empathy, feelings and thermography", *Elsevier, Consciousness and Cognition*, 34, pp.149–162, 2015.
- [14] Ueno S., *Biological Effects of Magnetic and Electromagnetic Fields*, Springer Science & Business Media, 2007.
- [15] Petreus D., Patarau T., Etz R., Kiraly E., Baru P.E., Low frequency pulsed magneto therapy microsystem, *Electronics Technology (ISSE)*, 36th International Spring Seminar on, pp.372 – 377, 2013.
- [16] M. S. Markov, *Electromagnetic Fields in Biology and Medicine*, CRC Press, 2015.
- [17] D. Dimitrov, A. Dimitrov, "Visualisation of Space-Temporal Configuration of Low Frequency Magnetic Signals", *Electronics and Electrical Engineering*, No.3(109), pp.46-48, 2011.
- [18] D. Dimitrov, "Visualization of a Low Frequency Magnetic Field, generated by Girdle Coil in Magnetotherapy", *Electronics and Electrical Engineering*, No.6(78), pp.57-60, 2007.
- [19] A. Presman, *Electromagnetic Fields and Life*, Springer Science & Business Media, 2013.
- [20] A. del Morala, María J. Azanzab, A model explaining synchronization of neuron bioelectric frequency under weak alternating low frequency magnetic field *Journal of Magnetism and Magnetic Materials*, Vol.377, pp.298–307, Elsevier, 2015.
- [21] J. Wilson, J. Haines, Z. Sienkiewicz, Y. Dubrova, The effects of extremely low frequency magnetic fields on mutation induction

- in mice, Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, Vol.773, pp.22–26, Elsevier, 2015.
- [22] C. Polk, Biological Effects of Magnetic and Electromagnetic Fields, (chapter) pp. 63-83, Biological Effects of Magnetic and Electromagnetic Fields, Springer, 1996.
- [23] C.L.M. Bauréus Koch, M. Sommarin, B.R.R. Persson, L.G. Salford, J.L. Eberhardt, Interaction between weak low frequency magnetic fields and cell membranes, Bioelectromagnetics, Vol. 24, Issue 6, pp.395–402, Wiley-Liss, Inc., 2003.
- [24] F. Prato, M. Kavaliers, A.W. Thomas, Extremely low frequency magnetic fields can either increase or decrease analgesia in the land snail depending on field and light conditions, Bioelectromagnetics, Vol. 21, Issue 4, pp. 287–301, Wiley-Liss Inc. 2000.
- [25] E. Boerner, J. Bauer, B. Ratajczak, E. Deren, H. Podbielska, “Application of thermovision for analysis of superficial temperature distribution changes after physiotherapy”, Springer, J Therm Anal Calorim, 120, pp.261–267, 2015.
- [26] A. Domingues, E. M. Pereira, J. Gabriel, R. Vardasca, “Case Study in Thermal Monitoring of Physiotherapy Treatments to Ankle Sprains in Rugby Athletes”, Pan American Journal of Medical Thermology, vol.1, No.1, pp.3-10, 2014.
- [27] A. Seixas, S. Rodrigues, V. Soares, T. Dias, R. Vardasca, J. Gabriel, “Thermographic assessment of a physiotherapy related upper limb task: Preliminary study”, Occupational Safety and Hygiene III, pp.273-277, 2015.
- [28] <http://www.flir.com/cs/emea/en/view/?id=41372> - FLIR Systems, Inc.
- [29] K. Ammer, “The Glamorgan Protocol for recording and evaluation of thermal images of the human body”, Thermology International, 10, 18(4), 2008, pp.125-129.
- [30] Morten Brøns, Martin Philip Bendsoe, Mads Peter Sørensen, Progress in Industrial Mathematics at ECMI 96, Springer-Verlag, 2013.
- [31] H.C. Taneja, Advanced Engineering Mathematics, I. K. International Pvt Ltd, 2007.



# Statistical Characteristics of Optimal Divergence of Laser Beam at FSO Systems

Tsvetan Mitsev<sup>1</sup>

**Abstract** – In this paper a methodology for calculating the statistical characteristics of the optimal divergence of the laser beam at free space optics systems is developed. It estimates the probability density function of meteorological visibility through the use of statistical material on the likelihood of different weather conditions. The statistical characteristics of the optimal divergence of the laser beam using a deterministic relationship between it and the meteorological visibility are calculated. A simplified model for calculating the volume coefficient of the atmospheric extinction is proposed. Its spectral dependence is taken into account

**Keywords** – FSO, optical wireless communication systems, probability density function, laser beam optimum divergence.

## I. INTRODUCTION

The main problem with the reliability of the work of free space optics (FSO) systems occurs in the presence of fog and haze. In a dense fog meteorological distance of visibility  $V$  [km] may be reduced and reach values of a few tens of meters to several meters. Relative to attenuation  $b$  (respectively to volume extinction coefficient  $\alpha$  [1/km]) in the communication channel, this corresponds to the values of the order of hundreds of dB/km.

In most works related to the model description of the communication channel, it is assumed homogeneous channel, i.e. constant attenuation along the length of the communication channel. This is due mainly to convenient for use in theoretical analyzes semi-empirical relationship between attenuation and meteorological visibility (the meteorological visibility reflects integrally the status of the communication channel, along it's entire length) [1], [2]. In modeling the communication channel attenuation  $b$  is determined mostly by empirical data on meteorological visibility [3], [4]. Most works devote attention to the wavelength of the optical radiation  $\lambda$ , i.e. to the spectral attenuation depending on the weather conditions [5], [6]. There are also models for prediction and analysis of the effects of fog on the operation of FSO communications system. [7], [8], [9].

In recent years, much attention is paid to the stochastic nature of fading of optical radiation into fog and to the study of statistical characteristics of the atmospheric channel. In contrast to the statistical characteristics of the communication channel of radio frequency wireless communication systems, which long ago are well studied, only recently more attention is paid to the study of statistical characteristics of communica-

tion channel of FSO systems [10]. Taking into account the stochastic nature of attenuation of optical radiation in the atmospheric channel of communication is the main factor that contributes to increasing the reliability of operation of these systems. Of particular importance is the modeling of optical communication channel in terms of fog and smoke or haze. This involves a lot of experimental work and theoretical research. An important step towards overcoming the impact of weather conditions and increasing opportunities for data transmission with FSO systems is statistical modeling of atmospheric communication channel.

In our previous work we consider the possibility of increasing the quality indicators of FSO systems through correct choice of divergence of optical radiation  $\theta_t$  [rad] of the transmitter during process of instalation [11], [12]. Also it is possible automatic adjustment of the divergence of optical radiation of the system during the working process. In [13] formula for calculating the optimum value of divergence angle  $\theta_{t,opt}$  depending on various parameters of the system and of the atmospheric communication channel is derived. Studies were made on the assumption of uniform and stationary atmospheric communication channel.

In this work we start building of statistical description of the communication channel of the FSO system. We use statistical data on the histogram of meteorological distance of visibility. The data used are for years. They refer to the northern hemisphere. A model of probability density function (PDF) of meteorological distance of visibility is built. The resulting PDF is used to calculate the statistical characteristics of meteorological visibility. A deterministic relationship between the optimal divergence of optical radiation of the transmitter  $\theta_{t,opt}$  of FSO system and meteorological distance of visibility is presented. Through this connection expressions to calculate the statistical characteristics of  $\theta_{t,opt}$  are derived and are calculated for a particular case.

## II. THEORY

As we noted in the introduction, in recent years more and more attention is paid to the statistical nature of the work of FSO systems. This is in direct relation to the increase of their reliability of work.

In our works [14] and [15] we showed that serious importance to increasing the time during which the system will operate with values of  $BER$  under specified maximum value is setting the divergence of optical radiation of the transmitter of its optimal value  $\theta_{t,opt}$  [rad]. Its value depends on the value of the random variable meteorological visibility distance  $V$  [km]. Therefore  $\theta_{t,opt}$  also is a random variable. To determine its statistical characteristics we need to know those on  $V$ .

<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.

In the literature is published for a given area and for a chosen season (or time period, usually annual) average statistical data on the percentage of the total time during which there is a particular weather conditions. Because surroundings are varied, mists, fogs, rain or snow, general characteristics, which can be summarized is meteorological visibility  $V$ . In fact, this is the characteristic important for the work of the FSO system. It is associated with volume extinction coefficient  $\alpha$  [ $\text{km}^{-1}$ ], and through it to the transparency of the communication channel  $\tau_a$ . The last characteristic is fundamental to energy balance (the change of power from the transmitter to the receiver) of the FSO systems.

#### A. Visibility probability density function (PDF)

Usually we have a histogram of the distribution of meteorological distance of visibility  $V$ . These are presented in tabular form statistical data on average time as a percentage of total in the year  $p$  [%] in which there is a particular weather conditions. Any weather conditions are characterized by an appropriate range of values of meteorological distance of vision (visibility)  $\Delta V_i$  [ $\text{km}$ ] =  $V_{i+1} - V_i$ ,  $i = 1, 2, 3 \dots$ .

Based on these data can be calculated statistical characteristics of meteorological visibility  $V$ , viewed as a random variable: probability density function (PDF)  $f(V)$  [ $\text{km}^{-1}$ ], mathematical expectation  $m_V$  [ $\text{km}$ ], dispersion  $\sigma_V^2$  [ $\text{km}^2$ ], average square deviation  $\sigma_V$  [ $\text{km}$ ]. To make a step shaped line  $p(V)$  at a probability density function  $f(V)$  we must comply with the condition [16]

$$\int_{-\infty}^{\infty} f(V) \cdot dV = 1. \quad (1)$$

In the case of a stepped (step shaped) function with finite range (limited by physical considerations above and below) the condition (1) is converted to

$$\sum_i f_i \Delta V_i = 1, \quad f_i = f(V \in [V_i, V_{i+1}]) = \text{const}. \quad (2)$$

The last equality and the condition

$$\sum_i K \cdot p_i \Delta V_i = 1$$

give us the relationship

$$f_i = K \cdot p_i, \quad K = \frac{1}{\sum_i p_i \cdot \Delta V_i}. \quad (3)$$

By determining the probability density function  $f_i$  we easily determine other statistical characteristics of  $V$  [15],

$$m_V = \int_{-\infty}^{\infty} V \cdot f(V) dV \quad (4)$$

and

$$\sigma_V^2 = \int_{-\infty}^{\infty} [V - m_V]^2 \cdot f(V) dV = \int_{-\infty}^{\infty} V^2 \cdot f(V) dV - m_V^2. \quad (5)$$

Of course for the practical calculations of  $m_V$  and  $\sigma_V^2$  we also replace integrals in (4) and (5) with sums like in (1).

#### B. Relationship between volume extinction coefficient and meteorological distance of vision (visibility)

Deterministic relationship between  $\theta_{t,\text{opt}}$  and  $V$  is implemented through transparency of the communication channel  $\tau_a$  and volume extinction coefficient  $\alpha$ . In the literature, the most commonly used semi-empirical relationship is

$$\alpha [\text{km}^{-1}] = \frac{3,92}{V [\text{km}]} \left( \frac{\lambda [\mu\text{m}]}{0,55} \right)^{-0,585 \sqrt[3]{V [\text{km}]}}. \quad (6)$$

This relationship is characterized by: it has relatively inconvenient form for theoretical analysis; it applies to the  $V \leq 10$  km. Therefore, the work offer more convenient for mathematical calculations relationship

$$\alpha = A \cdot V^{-Q}, \quad (7)$$

where the coefficients  $A$  and  $Q$  are spectrally dependent. In Table I are attached values of  $A$  and  $Q$  for three values of the wavelength of the optical radiation  $\lambda$ . The last column is a maximum relative error between the values of  $\alpha$  calculated by formulas (7) and (6). Calculations were made for a range of values of meteorological distance of visibility  $V$  [ $\text{km}$ ]  $\in [0,02; 100]$ .

TABLE I  
VALUES OF SPECTRAL COEFFICIENTS  $A$  AND  $Q$

$\lambda$ [ $\mu\text{m}$ ]	$A$	$Q$	$\Delta\alpha_{\text{max}}$ [%]
0,85	2,87	1,08	7,28
1,31	2,11	1,16	15,53
1,55	1,87	1,2	22,97

#### C. Calculation of the statistical characteristics of $\theta_{t,\text{opt}}$ by numerical integration

In our work [11] is taken out the deterministic relationship between random quantity  $\theta_{t,\text{opt}}$  and random quantity  $V$ , namely

$$\theta_{t,\text{opt}} \equiv \theta = \frac{1}{z} \sqrt{\frac{2 \cdot \tau_t \cdot \Phi_L}{\pi \cdot e \cdot I_{\text{min}}}} \sqrt{e^{-\alpha(V) \cdot z}}, \quad (8)$$

where  $z$  [ $\text{km}$ ] is the length of the communication channel;  $\tau_t$  is the transmission of transmitter optical antenna;  $\Phi_L$  [W] is the power of the laser pulses (i.e. bit entity in IM / DD and

OOK modulation);  $e$  is the base of natural logarithm;  $I_{\min}$  [ $\text{W}/\text{m}^2$ ] is the minimum intensity of the optical radiation into the aperture of the photodetector.

The minimum intensity  $I_{\min}$  depends on many parameters of the receiver and different physical constants. It is selected so that satisfies the information capacity of the systems  $C_1$  [bps] and the specified maximum value for bit error  $BER_{\max}$ , which will operate the system.

If it is known a PDF of  $V$  and using equation (3) and connection (8), we can determine the statistical characteristics of  $\theta$  [16]

$$m_{\theta} = \int_{-\infty}^{\infty} \theta(V) \cdot f(V) \cdot dV, \quad (9)$$

$$\sigma_{\theta}^2 = \int_{-\infty}^{\infty} [\theta(V) - m_{\theta}]^2 \cdot f(V) \cdot dV = \int_{-\infty}^{\infty} \theta^2(V) \cdot f(V) \cdot dV - m_{\theta}^2. \quad (10)$$

Of course, here, similarly to (1), due to the step shaped dependence  $f(V)$ , replacing the solution of the integrals with the sums. In this case the inner integrals for the different members of the sum, i.e., within the ranges  $\Delta V_i$ , we solve numerically.

### III. RESULTS AND DISCUSSIONS

To test developed methodology for calculating the statistical characteristics of optimum divergence of the laser beam at FSO systems we will use enclosed in [17] tabulated statistical average data on the percentage of time during the year in which there is a different weather conditions. The data refer to the northern hemisphere of the Earth (Tab. II, columns 2 and 4). The values of meteorological distance of visibility  $V$  are divided into 11 spaces. The probability of a fall of  $V$  into the final two spaces is zero. On the basis of these data has been obtained the probability density function of  $V$  in column 6.

TABLE II  
PROBABILITY DENSITY FUNCTION OF METEOROLOGICAL DISTANCE OF VISIBILITY  $V$

$i$	$V$ [km]	$\Delta V$ [km]	$P$ [%]	$\Delta V \cdot P$	$f(V)$ [1/km]	$\Delta V \cdot f(V)$
1	2	3	4	5	6	7
1	< 0,05	0,05	0	0	0	0
2	0,05÷0,2	0,15	0,4	0,06	3,0256E-4	4,538E-5
3	0,2÷0,5	0,3	0,6	0,18	4,538E-4	1,362E-4
4	0,5÷1	0,5	1,6	0,8	1,21E-3	6,051E-4
5	1÷2	1	4,2	4,2	3,177E-3	3,177E-3
6	2÷5	3	0,6	1,8	4,538E-4	1,362E-3
7	5÷10	5	43,4	217	3,282E-2	1,641E-1
8	10÷20	10	19,8	198	1,498E-2	1,498E-1
9	20÷50	30	28,5	855	2,156E-2	6,467E-1
10	50÷100	50	0,9	45	6,808E-4	3,404E-2
11	> 100	$\infty$	0	0	0	0
$\Sigma$	—	—	100	1322,04	—	1

On the Fig. 1 is shown  $f(V)$ , as step-like character of the  $f(V)$  is replaced with a line. The line connects neighboring values  $f_i$ . The values  $f_i$  are built for the midpoint of each interval  $\Delta V_i$ .

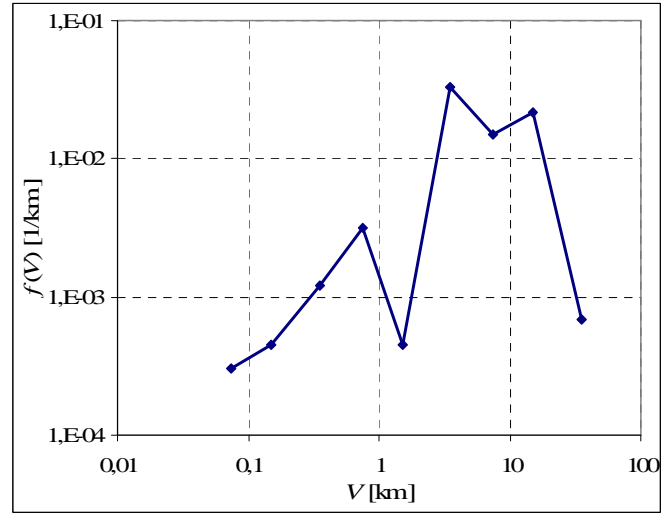


Fig. 1. Probability density function  $f(V)$  of meteorological distance of visibility  $V$

Fig. 2 shows used in the work dependencies  $\alpha(V)$  calculated for the three lengths of  $\lambda$ . Calculation based on the expression (7) and corresponding coefficients from the Table I.

It is seen a big difference in the values of the volume extinction coefficient  $\alpha$  for various wavelengths with increasing the value of meteorological distance of visibility  $V$ .

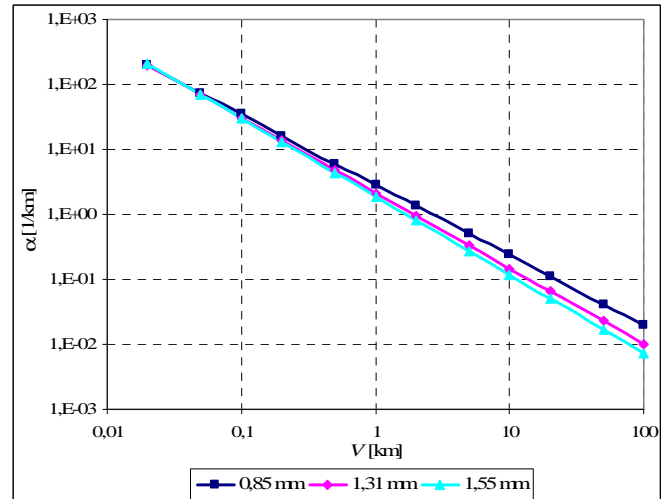


Fig. 2. Dependence on the volume extinction coefficient  $\alpha$  of meteorological distance of visibility  $V$

For the calculation of the statistical characteristics through (8) and (9) we choose typical values of variables involved:  $z = 1$  km;  $\lambda = 1,55$   $\mu\text{m}$ ;  $\tau_t = 0,85$ ;  $\Phi_L = 20$  [mW];  $I_{\min} = 1,023 \cdot 10^{-4}$   $\text{W}/\text{m}^2$ . The value of  $I_{\min}$  is calculated for the

average statistical parameters of the receiver. In the calculation is selected  $C_1 = 100$  Mbps and  $BER_{\max} = 10^{-8}$ . In the value of  $I_{\min}$  the presence of background optical radiation is reported. Calculated by the method above described values of statistical characteristics of  $V$  and of  $\theta_{t,\text{opt}}$  are mapped in Table III.

TABLE III

VALUES OF STATISTICAL CHARACTERISTICS OF VISIBILITY AND OF OPTIMAL DIVERGENCE OF OPTICAL RADIATION OF TRANSMITTER

	$m$	$\sigma^2$	$\sigma$
$V$	28,7 km	261,5 km <sup>2</sup>	16,2 km
$\theta_{t,\text{opt}}$	6,046 mrad	36,62 mrad <sup>2</sup>	6,052 mrad

#### IV. CONCLUSION

Over the past decade has increased the number of scientific publications devoted to the possibilities of predicting and improving the reliability of operation of FSO systems. The reliability of the work is directly dependent on the parameters of the atmospheric channel of communication and the choice of the main parameters of the systems. The random character of the the parameters of the communication channel leads to random character of the basic parameters of the system. One of the main parameters is the optimal divergence of optical radiation of the transmitter.

The first step to a statistical analysis of reliability of FSO systems is to build real statistical model of atmospheric channel of communication, finding its statistical characteristics and related statistical characteristics of optimal divergence of optical radiation of the transmitter.

The work shows that in typical for northern hemisphere weather conditions in a communication channel with a length of 1 km and average statistical parameters of the system, the average optimal divergence is 6,05 mrad with dispersion 36,62 mrad<sup>2</sup>.

#### REFERENCES

- [1] Oliver Bouchet, Herve Sizum, Christian Boisrobert, Frederique de Foruel, *Free-Space Optics: Propagation and Communications*, John Wiley&Sons, 2010.
- [2] Roberto Ramirez-Iniguez, Sevia M. Idrus, Ziran Sun, *Optical Wireless Communications - IR for Wireless Connectivity*, CRC Press, An Auerbach book, 2008.
- [3] F. Nadeem, T. Javornik, E. Leitgeb, V. Kvicera, and G. Kandus, "Continental fog attenuation empirical relationship from measured visibility data", *Radioengineering*, vol. 19, p. 596-600, Dec. 2010.
- [4] Heinz A. Willebrand, Baksheesh S. Ghuman, *Free-Space Optics: Enabling Optical Connectivity in Today's Networks*, Indianapolis, Sams Publishing, 2002.
- [5] I. Kim, B. McArthur, E. Korevaar, "Comparison of laser beam propagation at 785 nm and 1550 nm in fog and haze for optical wireless communications," *Proc. of SPIE*, vol. 4214, pp. 26-37, 2001.
- [6] K. Fischer, M. Witiw, and E. Eisenberg, "Optical attenuation in fog at a wavelength of 1.55 micrometers", *Atmospheric Research*, vol. 87, pp. 252-258, March 2008.

- [7] J. Gebhart, M. Leitgeb, E. Bregenzer, "Atmospheric effects on optical wireless links", in 7-th International Conference on Telecommunication, vol.2, pp. 395-401, Zagreb, 2003.
- [8] M. Al Naboulsi, H. Sizun, F. de Fornel, "Fog attenuation prediction for optical and infrared waves", *Journal of SPIE, Optical Engineering*, vol 43, pp. 319-329, 2004.
- [9] M. Ijaz, Z. Ghassemlooy, J. Pesek, O. Fiser, H. Le Minh, and E. Bentley, "Modeling of fog and smoke attenuation in free space optical communications link under controlled laboratory conditions", *IEEE Journal of Lightwave Technology*, vol. 31, pp. 1720-1726, April 2013.
- [10] M. Khan, M. Awan, E. Leitgeb, F. Nadeem, and I. Hussain, "Selecting a distribution function for optical attenuation in dense continental fog conditions", *International Conference on Emerging Technologies*, pp. 142-147, Oct. 2009.
- [11] Tsvetan Mitsev, Kalin Dimitrov, Hristo Ivanov, Nikolai Kolev, Optimum divergence of laser radiation in FSO systems, 7-th International Conference on Communications, Electromagnetics and medical Applications (CEMA'12), pp. 42-45, Athens, Greece, 2012.
- [12] Mitsev Ts., N. Kolev, Hr. Ivanov, K. Dimitrov, Optimum Divergence of the Transmitter Optical Radiation in FSO Systems, XLVII Intern. Scientific Conf. on Inform., Communication and Energy Systems and Technol. (ICEST 2012), pp. 55-58, Veliko Tarnovo, Bulgaria, 2012.
- [13] Yordan Kovachev, Tsvetan Mitsev, FSO Availability Depending on the Meteorological Conditions, 9-th International Conference on Communications, Electromagnetics and Medical Applications (CEMA'14), pp. 19-23, Sofia, 2014.
- [14] Tsvetan Mitsev, Nikolay Kolev, Optimal Divergence of Laser Beam in Optical Wireless Communication Systems, *Elektrotechnika & Elektronika*, Vol. 49, No 11-12, pp. 15-20, 2014.
- [15] Tsvetan Mitsev, Yordan Kovachev, Availability of MFSO Using Optimal System Parameters, *Proceedings of 14<sup>th</sup> Conference on Microwave Techniques "COMITE 2015"*, April 22-23, Pardubice, Czech Republic, pp. 131-134, ISBN 978-1-4799-8121-2, 2015.
- [16] E. Ferdinandov, B. Pachedjieva, Probabilistic and statistical methods in communications – Part I, Sofia, Siela, 2005. (in bulgarian)
- [17] E. Ferdinandov, Laser radiation in Radio-technics, Sofia, Technics, 1981. (in bulgarian)



# ANN Approach for Return Loss Analysis of Ohmic Series RF MEMS Switches

Tomislav Ćirić<sup>1</sup>, Zlatica Marinković<sup>1</sup>, Marija Milijić<sup>1</sup>, Olivera Pronić-Rančić<sup>1</sup>,  
Vera Marković<sup>1</sup>, Larissa Vietzorreck<sup>2</sup>

**Abstract** – RF MEMS switch applications in communication systems have been increased in the recent time, creating a need for reliable and accurate switch models. The model exploited for the analysis presented in this paper is based on the artificial neural networks. In particular, a model relating the lateral dimensions of the switch bridge and the switch S-parameters is considered. The model allows fast and efficient analysis of S-parameters behavior, requiring significantly shorter time for the same analyses than standard electromagnetic simulators. The main purpose of the paper is analysis of the switch return loss changes with small the changes of the bridge dimensions. The most illustrative results of the analysis are discussed and corresponding recommendations for the switch optimization are given.

**Keywords** –Artificial neural networks, Return Loss, Insertion Loss, RF MEMS switch.

## I. INTRODUCTION

RF MEMS switches are surface-micromachined devices which use mechanical movements to achieve a short circuit or an open circuit in the RF transmission-line [1]. These devices are extremely small and light and are applicable in many communication devices and systems. Also, they are highly linear and can operate up to higher frequency bands providing high isolation and low insertion loss in general. The dimensions of RF MEMS switch components have to be optimized carefully in order to fulfill the desired RF specifications. Generally, the optimization is performed by full-wave EM simulations [2]. However, simulations in standard simulators are time consuming and require considerable computing resources because the switch generally consists of several thin layers and vias with high complexity. Moreover, since the optimization process mostly involves a series of EM simulations with varying parameters, the computing load is further increased. The common method to improve simulation speed and efficiency is to use a lumped element circuit model in a circuit simulator. However, to derive an accurate circuit model is demanding, especially at high frequencies, because the circuit model should involve more parasitic components at higher frequencies, which

increases the model complexity [3]. A solution to these needs can be the application of artificial neural networks (ANNs) for deriving proper models for different switch types and to relate different sets of input parameters to the desired output quantities. ANNs have been widely used as a fast and efficient tool for modeling of different electronic devices [4]-[5], among them RF MEMS devices [6]-[7]. Use of ANN based models has a great advantage in reducing the computational cost, especially when implemented within a circuit simulator that has integrated tuning and optimization options. Namely, ANNs are capable of modeling nonlinear mappings of multiple input/output parameters, providing models which enable an accurate device characterization and efficient prediction of unknown input-output relationships with low computational overhead.

ANN based modeling and optimization of an ohmic series RF MEMS switch is presented in [8]. On the base of several full-wave numerical simulations of the switch, ANNs are developed to directly relate the switch geometrical parameters to the scattering parameters over frequency. Besides the lateral dimensions of the bridge, the lateral dimensions of the gap surrounding the bridge are concerned as well. The gap size has to be adjusted in order to achieve a proper matching of the bridge section with input/output line impedance. It is reported that with the developed model an optimization of the structure can be done within seconds.

In the design of an RF MEMS switch one of the requirements to be met in the optimization is to keep return loss below a specified value in the frequency range of interest. In this paper the model presented in [8] is exploited for analysis of the switch return loss. In particular, special attention is put on analysis how the return loss deviates from the nominal values if the dimensions of the bridge membrane are changed due to their small changes caused by the fabrication tolerances. In addition, the corresponding insertion loss changes are reported in the paper.

The paper is organized as follows: After the introduction, in Section II the considered RF MEMS switch is described. The ANN model used in this work is presented in Section III. The results of the analysis of the switch return and insertion losses and the corresponding discussion are given in Section IV. Finally, Section V contains the concluding remarks.

## II. MODELED DEVICE

The device considered in this work is a CPW (coplanar waveguide) based RF MEMS ohmic series switch, depicted in

<sup>1</sup> T. Ćirić, Z. Marinković, M. Milijić O. Pronić-Rančić, and V. Marković are with the University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia  
E-mail: cirict78@gmail.com, zlatica.marinkovic@elfak.ni.ac.rs, marija.milijic@elfak.ni.ac.rs, olivera.pronic@elfak.ni.ac.rs, vera.markovic@elfak.ni.ac.rs

<sup>2</sup> L. Vietzorreck is with the TU München, - Lehrstuhl für Hochfrequenztechnik, Arcisstr. 21, 80333 München, Germany, e-mail: vietzorreck@tum.de.

Fig. 1a. The switch is fabricated at FBK in Trento on a 525 $\mu\text{m}$  thick silicon wafer by standard surface micromachining and CMOS technologies, which requires eight masks, [9]. A serious switch should have a low return loss in ON state. Therefore, not only the bridge length ( $L_b$ ) and width ( $W_b$ ), but also lateral ( $L_g$ ) and longitudinal ( $W_g$ ) gaps between a bridge anchor and CPW ground have to be adjusted. This means, during optimization, those dimensions are to be simultaneously swept and determined to the optimums.

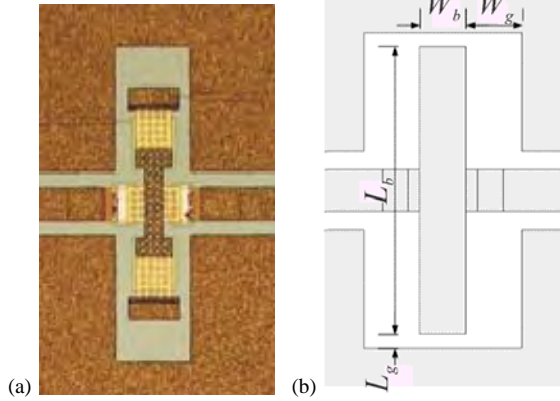


Fig. 1. Fabricated ohmic series RF MEMS switch [9] (a) and the simplified configuration (b)

### III. EXPLOITED ANN MODEL

The ANN model used to perform the analysis consists of four multilayer perceptron networks modeling magnitudes ( $|S_{ij}|$ ) and for the phases ( $\angle S_{ij}$ ) of  $S_{11}$  and  $S_{21}$  scattering (S-) parameters against the switch lateral dimensions ( $L_b, W_b, L_g, W_g$ ) and frequency  $f$ , as shown in Fig. 2. As the switch is a symmetrical and reciprocal component, it follows that  $S_{22} = S_{11}$  and  $S_{12} = S_{21}$ , therefore the presented model can be used to predict all four S-parameters. Each of the four ANNs has five input neurons corresponding to four mentioned lateral dimensions of the switch and the frequency. All the ANNs have only one neuron in the output layer corresponding to one of the modeled parameters ( $|S_{11}|(\text{dB})$ ,  $\angle S_{11}(\text{deg})$ ,  $|S_{21}|(\text{dB})$  and  $\angle S_{21}(\text{deg})$ ).

The ANN model of the considered device was developed and validated with the S-parameters data obtained from full-wave simulations with CST Microwave Studio (CST MWS) [10]. Instead of the original device layout, a simplified layout as shown in Fig. 1(b) was used. The details about training and validation of the model can be found in [8]. Trained ANN models are capable to predict the S-parameters for the given bridge geometrical parameters and frequency, enabling fast and efficient further analyses related to the bridge size adjustment in order to fulfill the design requirements.

For the analysis of reported in this paper, the ANNs modeling magnitudes of  $S_{11}$  and  $S_{21}$  have been used, as

magnitude of  $S_{11}$  represents return loss and the magnitude of  $S_{21}$  represents the insertion loss. As reported in [8], the both ANNs have two-hidden-layers: the ANN modeling  $|S_{11}|(\text{dB})$  has 17 hidden neurons in each of the two hidden layers, whereas the ANN modeling  $|S_{21}|(\text{dB})$  has 18 hidden neurons in the first hidden layer and 16 neurons in the second hidden layer.

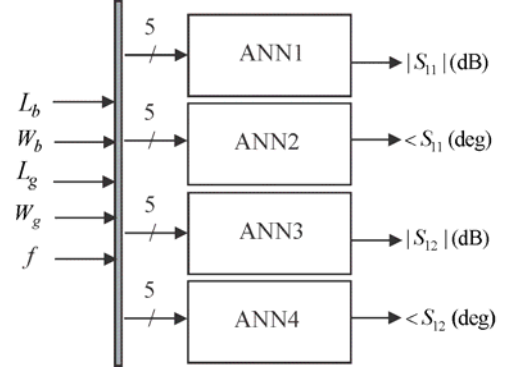


Fig. 2. ANN model of ohmic series RF MEMS switch

### IV. RESULTS AND DISCUSSION

The mentioned two ANNs have been used for further analysis of the behavior of the switch return loss and switch insertions loss with changes of the switch lateral dimensions. The analysis has been performed for the following range of the lateral dimensions:  $L_g$  - from 30 to 110  $\mu\text{m}$ ,  $L_b$  - from 200 to 1000  $\mu\text{m}$ ,  $W_g$  - from 70 to 150  $\mu\text{m}$ ,  $W_b$  - from 20 to 220  $\mu\text{m}$ . The considered frequency range has been from 0 to 40 GHz. It should be mentioned that the above given ranges correspond to the validity range of the model, i.e. to the range of input parameters used for the model development.

For the purpose of analysis, eight differently sized devices, i.e., eight combinations of the bridge lateral dimensions have been considered.

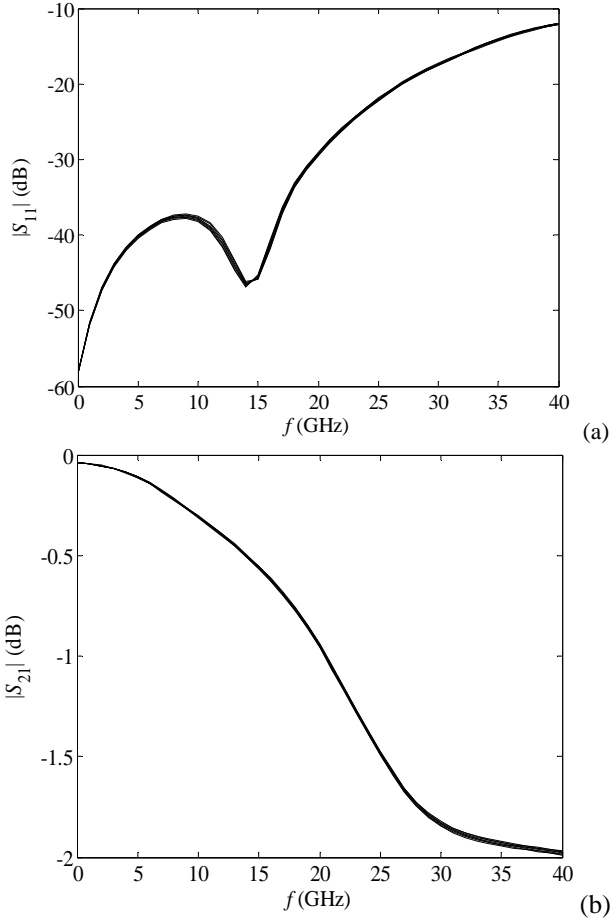
As mentioned in Introduction, the aim of work is to analyze changes of the switch return loss caused by small changes of the dimensions in fabrication of the bridge. The considered deviations have been in the range  $\pm 3 \mu\text{m}$  (taken with the step of 1  $\mu\text{m}$ ), as that can be considered as the range of bridge fabrication tolerances in the considered technology. Moreover, besides return loss, analyses of the insertion loss changes are given. Therefore, the analysis has been performed for the deviation of the bridge dimensions  $L_b$  and  $W_b$ . First, the return and insertion loss changes when one of the two considered bridge lateral dimensions has been changed, while the others have been kept constant have been calculated by the neural model and the corresponding deviations from the expected nominal values were calculated.

The calculated deviations, i.e. maximum deviation values (absolute values) in the considered frequency range are given

**TABLE I**

 RETURN AND INSERTION LOSS MAXIMUM DEVIATIONS -  $L_b$  CHANGES

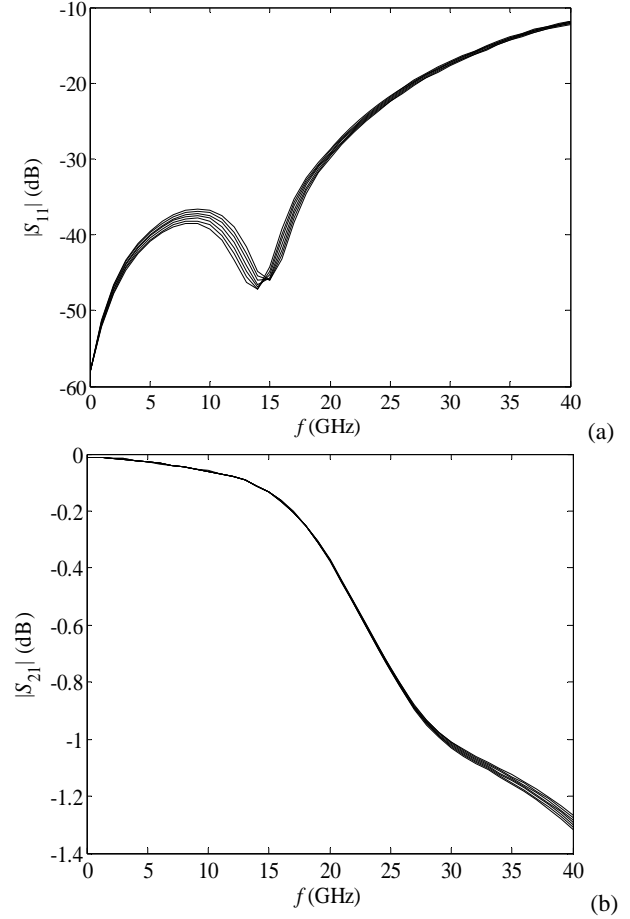
$L_b$ [ $\mu\text{m}$ ]	$W_b$ [ $\mu\text{m}$ ]	$L_g$ [ $\mu\text{m}$ ]	$W_g$ [ $\mu\text{m}$ ]	$\text{Max} \Delta S_{11} $ [dB]	$\text{Max} \Delta S_{21} $ [dB]
200	20	30	70	0.5	0.0026
400	20	50	70	0.29	0.0025
300	70	90	70	0.42	0.0038
800	70	70	90	0.27	0.0054
800	20	50	110	0.54	0.0075
400	20	30	150	0.73	0.0045
1000	220	30	20	0.08	0.012
600	120	70	110	0.58	0.0063


 Fig. 3. Maximum influence of  $L_b$  deviation: (a) Return loss: device (400, 20, 30, 150), (b) Insertion loss: device (1000, 220, 30, 20).

**TABLE II**

 RETURN AND INSERTION LOSS MAXIMUM DEVIATIONS -  $W_b$  CHANGES

$L_b$ [ $\mu\text{m}$ ]	$W_b$ [ $\mu\text{m}$ ]	$L_g$ [ $\mu\text{m}$ ]	$W_g$ [ $\mu\text{m}$ ]	$\text{Max} \Delta S_{11} $ [dB]	$\text{Max} \Delta S_{21} $ [dB]
200	20	30	70	1	0.0016
400	20	50	70	0.95	0.0092
300	70	90	70	0.83	0.0064
800	70	70	90	1.1	0.011
800	20	50	110	2.4	0.024
400	20	30	150	2.4	0.015
1000	220	30	20	0.24	0.017
600	120	70	110	1.3	0.0079


 Fig. 4. Maximum influence of  $W_b$  deviation: (a) Return loss: device (400, 20, 30, 150), (b) Insertion loss: device (800, 20, 50, 110).

in Tables I and II for changes of  $L_b$  and  $W_b$ , respectively. The results show that the maximum deviation of the return loss is less than 0.73 dB for the  $L_b$  deviation and less than 2.4 dB for the  $W_b$  deviation. For the insertion loss, the deviations are less than 0.024 dB, which is significantly smaller comparing to the values of insertion loss. For illustration, in Fig. 3 and 4, the plots corresponding to devices which exhibited the maximum deviations are given. Devices are denoted as  $(L_b, W_b, L_g, W_g)$ , where dimensions are given in  $\mu\text{m}$ .

In order to examine what happens when both dimensions are changed simultaneously in the range  $\pm 3 \mu\text{m}$ , keeping the

gap dimensions constant, the same eight sized bridges have been considered. The maximum deviations of the return and insertion losses are depicted in Table III. The maximum deviation of the return loss is less than 3.1 dB and 0.032 dB for the insertion loss. The frequency dependence of the return and insertion loss for the devices exhibiting max deviations is shown in Fig. 5.

In all analysed cases, the insertion loss is most sensitive in frequency range from 25 to 40 GHz, where in the case of return loss the most sensitive frequency range varies with the bridge size. As expected, deviations are higher in the case of smaller sized devices.

TABLE III  
RETURN AND INSERTION LOSS MAXIMUM DEVIATIONS FOR  
SIMULTANEOUS CHANGES OF THE  $L_b$  AND  $W_b$  UP TO  $\pm 3 \mu\text{m}$

$L_b$ [ $\mu\text{m}$ ]	$W_b$ [ $\mu\text{m}$ ]	$L_g$ [ $\mu\text{m}$ ]	$W_g$ [ $\mu\text{m}$ ]	$\text{Max} \Delta S_{11} $ [dB]	$\text{Max} \Delta S_{21} $ [dB]
200	20	30	70	1.5	0.0036
400	20	50	70	1.2	0.012
300	70	90	70	1.3	0.01
800	70	70	90	1.4	0.016
800	20	50	110	3	0.032
400	20	30	150	3.1	0.02
1000	220	30	20	0.31	0.029
600	120	70	110	1.9	0.014

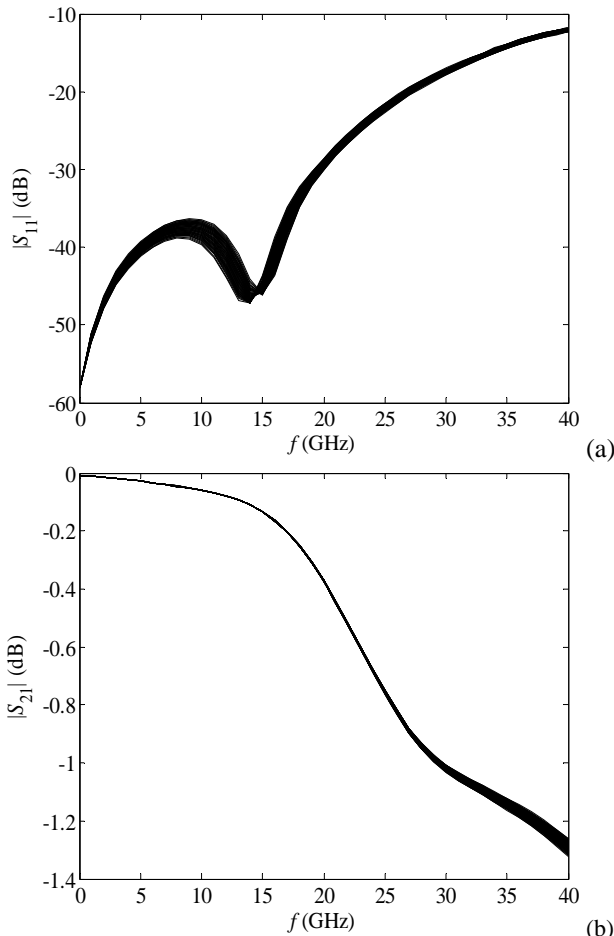


Fig. 5. Maximum influence for simultaneously deviation of  $L_b$  and  $W_b$ : (a) Return loss - device (400, 20, 30, 150), (b) Insertion loss - device (800, 20, 50, 110).

## V. CONCLUSION

In this paper an ANN model of ohmic series RF MEMS switch has been applied to obtain and analyze the switch return and insertion losses versus the switch bridge lateral dimensions. Since the model has very short response time, the analysis was performed in significantly shorter time that it would be needed in standard electromagnetic simulators. The influence of small changes of the bridge lateral dimensions caused by the fabrication process on the return and insertion

losses has been studied. It is shown that the deviation of the switch lateral dimensions during fabrication process cause deviations of the return and insertion loss. However, in all considered case, the deviations of the insertion loss are almost negligible, while in the case of return loss the deviations can be up to around 3 dB. Therefore, taking care about the frequency range of interest, it is recommended to set the optimization goal for the return loss to be a few dB smaller than the requested value in order to ensure that, even with the deviation of the bridge dimensions during the fabrication, the return loss is below the desired value.

## ACKNOWLEDGEMENT

This work was funded by the bilateral Serbian-German project "Smart Modeling and Optimization of 3D Structured RF Components" supported by the DAAD foundation and Serbian Ministry of Education, Science and Technological Development. The work was also supported by the projects III-43012 and TR-32052 of the Serbian Ministry of Education, Science and Technological Development.

## REFERENCES

- [1] G. M. Rebeiz, *RF MEMS Theory, Design, and Technology*, New York, Wiley, 2003.
- [2] L. Vietzorreck, "EM Modeling of RF MEMS," 7th International Conference on Thermal, Mechanical and Multiphysics Simulation and Experiments in Micro-Electronics and Micro-Systems, EuroSime, pp.1-4, Como, Italy.
- [3] M. A. Llamas, D. Girbau, E. Pausas, L. Pradell, S. Aouba, C. Villeneuve, V. Puyal, P. Pons, R. Plana, S. Colpo, F. Giacomozzi, "Capacitive and Resistive RF-MEMS switches 2.5D & 3D Electromagnetic and Circuit Modeling", Spanish Conference on Electron Devices, 2009. CDE 2009, pp.451-454, 11-13 Feb. 2009.
- [4] Q. J. Zhang and K. C. Gupta, *Neural Networks for RF and Microwave Design*, Boston, MA, Artech House, 2000.
- [5] H. Kabir, L. Zhang, M. Yu, P. Aaen, J. Wood, and Q. J. Zhang "Smart Modelling of Microwave Devices", *IEEE Microwave Magazine*, vol. 11, no. 3, pp.105-108, 2010.
- [6] Y. Lee, D. S. Filipovic, "Combined Full-wave/ANN based Modelling of MEMS Switches for RF and Microwave Applications," *Proc. of IEEE Antennas and Propagation Society International Symposium*, vol. 1A, pp. 85-88, Washington DC, USA, 2005.
- [7] Y. Gong, F. Zhao, H. Xin, J. Lin, Q. Bai, "Simulation and Optimal Design for RF MEMS Cantilevered Beam Switch," *Proc. of International Conference on Future Computer and Communication (FCC '09)*, pp. 84-87, Wuhan, China, 2009.
- [8] M. Milijic, Z. Marinkovic, T. Kim, O. Pronic-Rancic, L. Vietzorreck, V. Markovic, "Modeling and optimization of ohmic series RF MEMS switches by using neural networks," *8th German Microwave Conference, GeMiC 2014*, Aachen, Germany, March, 2014, pp. 1-4.
- [9] S. DiNardo, P. Farinelli, F. Giacomozzi, G. Mannocchi, R. Marcelli, B. Margesin, P. Mezzanotte, V. Mulloni, P. Russer, R. Sorrentino, F. Vitulli, L. Vietzorreck, "Broadband RF-MEMS based SPDT", *Proc. European Microwave Conference 2006*, Manchester, Great Britain, September 2006.
- [10] CST MICROWAVE STUDIO® 2012.



# Calculation of Capacitance of the Rectangular Coaxial Lines with Offset Inner Conductor by Strong FEM

Vladimir V. Petrovic<sup>1</sup> and Žaklina J. Mančić<sup>2</sup>

**Abstract** – In this paper, capacitance per unit length of rectangular coaxial transmission lines with offset nonzero-thickness inner conductor, having an isotropic and anisotropic dielectric, using strong FEM formulation is calculated. The results were compared with the results obtained by the weak FEM and commercial software FEMM, which uses node-based first-order basis function. Based on that, appropriate conclusions are made.

**Keywords** – Quasi-static analysis, Finite element method, Strong FEM formulation, lines with rectangular cross section, offset inner conductor, isotropic and anisotropic dielectric, capacitance per unit length.

## I. INTRODUCTION

Problem of capacitance per unit length of square or rectangular lines calculation, especially lines with offset inner conductor is topical in theory and practice. The paper [1] gives a review of the literature, dealing with this task and it performs the calculation of capacitance of the rectangular coax line with offset inner conductor by using the weak FEM formulation [2]. This paper deals with calculation of capacitance per unit length of square and rectangular coaxial lines filled with isotropic and anisotropic dielectric by using strong FEM formulation [3-6]. The results are compared with those obtained by weak FEM [1] and by commercial software FEMM [6]. FEM is a very suitable method for the analysis of closed polygonal structures and it can be simply used for analysis of geometries with anisotropic dielectrics, unlike the methods that use Green's function (e.g., MoM or EEM) for which an additional complicated step of anisotropic Green's function determination is needed [7]. Besides classifying FEM into strong and weak formulation, this method can be classified as a node-based [1,6,8,9] and non node-based (with hierarchical basis functions) [2-5, 10-12]. Node-based FEM can be found much more often than non node-based FEM. However, weak FEM formulation is usually presented in the literature, while strong formulation can rarely be found. In weak FEM formulation, only function's continuity condition is exactly satisfied, whereas in strong FEM formulation, boundary conditions for the both function and its first derivative are satisfied exactly [2-5,10-12]. In this paper are obtained for the third order basis functions ( $n = 3$ ).

## II. BRIEF DESCRIPTION OF THE STRONG FEM FORMULATION

FEM approach in this paper is based on hierarchical strong basis functions of higher (arbitrary) order that are constructed

by using mutual multiplication of 1D strong basis functions [13]. Consider a two-dimensional domain, uniform with respect to  $z$ -axis, Fig. 1, filled with linear inhomogeneous dielectric without free charges, in which the distribution of electrostatic potential,  $V(x, y)$ , is the unknown function. Let the problem be of the closed type: on one part of the domain boundary ( $C_1$ ), boundary conditions of the first kind (given  $V$ ), and on the rest of the boundary ( $C_2$ ), boundary conditions of the second kind (given  $\partial V / \partial n$ ), are imposed (Fig.1). (Boundary condition of the second kind here is equivalent to given  $D_n = -\epsilon \partial V / \partial n$ .) Differential equation for  $V(x, y)$  can be defined with:

$$\text{div}_S (\epsilon \text{grad}_S V) = 0, \quad (1)$$

In previous equation  $\text{div}_S$  and  $\text{grad}_S$  denote surface divergence and gradient, respectively. Calculation domain is divided into  $M$  sub-domains (elements) in FEM solution of Eq. (1).

Exact solution  $V(x, y)$  is expressed as a linear combination of basis functions with unknown coefficients,

$$V \approx f = \sum_{j=1}^N a_j f_j.$$

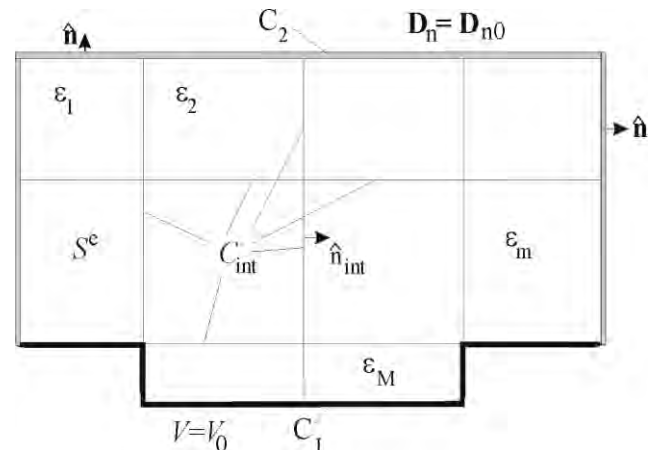


Fig. 1. Two-dimensional calculation domain divided into elements.

The system of linear algebraic equations for unknown coefficients is obtained by applying the weak Galerkin formulation [14, 15], and it is defined with:

$$[K_{ij}][a_j] = [G_i], \quad i, j = 1, \dots, N, \quad (2)$$

where

$$K_{ij} = \int_S \epsilon (\text{grad} f_i) (\text{grad} f_j) dS, \quad G_i = \int_{C_2} f_i D_{n0} dl. \quad (3)$$

<sup>1</sup>Vladimir V. Petrovic is with the Robert Bosch, GmbH, Reutlingen, Germany, e-mail [vvpetf@rcub.bg.ac.rs](mailto:vvpetf@rcub.bg.ac.rs)

<sup>2</sup>Žaklina J. Mančić is with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, e-mail [zaklina.mancic@elfak.ni.ac.rs](mailto:zaklina.mancic@elfak.ni.ac.rs).

In previous equation with  $D_{n0}$  is denoted a normal component of vector  $\mathbf{D}$  on the contour  $C_2$ , whereas  $i$  and  $j$  represent global serial numbers of basis functions. Furthermore,  $S$  represents the union of all the element's surfaces, defined with  $S = \bigcup_{e=1}^M S^e$ . Next, rectangular elements of arbitrary order are utilized for strong formulation. Strong basis functions automatically satisfy continuity of potential  $V$  ( $C^0$  continuity) and continuity of  $D_n$  (generalized  $C^1$  continuity) on interelement boundaries ( $C_{int}$  in Fig.1). Complete set of strong basis functions for 2-D problems in homogeneous (isotropic or anisotropic) media is presented in [13]. Instead of  $\epsilon$ , for anisotropic dielectrics it should be used  $\bar{\epsilon} = \begin{bmatrix} \epsilon_x & \epsilon_y \end{bmatrix}$  in equation (3).

### III. NUMERICAL EXAMPLES

#### I. Square coaxial line with offset inner conductor

For a square coaxial line with offset inner conductor, Fig. 2, for  $b/a=4$ , results for normalized capacitance per unit length,  $C'/\epsilon$ , are presented in Fig. 3. When the inner conductor is moved from the center and positioned closer to the outer conductor, the normalized capacitance increases. The results of  $C'/\epsilon$  in the case when  $b/a=4$  are compared with the corresponding results obtained by FEMM [6] and results obtained by weak FEM [1]. The results are shown in Fig. 3 and an excellent agreement can be observed. In this case, it is not possible to exploit symmetry for the problem solution. In all the cases the mesh that consists of 288 rectangular elements is used for strong and weak FEM. This resulted in 1152 unknowns for strong FEM and 2448 unknowns for weak FEM formulation. In order to obtain results of the similar accuracy by using FEMM software, the number of nodes (which is equal to the number of unknowns) was between 3980 and 4130 while the number of triangular mesh elements was between 7592 and 7830.

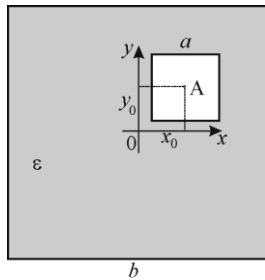


Fig. 2. Square coaxial line with offset inner conductor. Coordinate origin is in the center of the outer conductor.

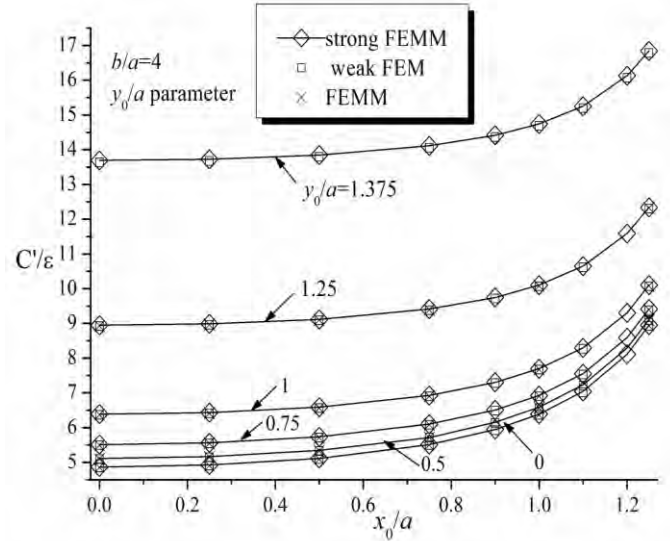


Fig. 3. Ratio  $C'/\epsilon$  depending on  $x_0/a$ , where  $y_0/a$  is a parameter,  $b/a=4$  and dielectric is isotropic.

#### II. Rectangular coaxial line with offset inner conductor

For rectangular coaxial line, Fig. 4,  $C'/\epsilon$  dependence of  $x_0/a$  is shown in Fig. 5.

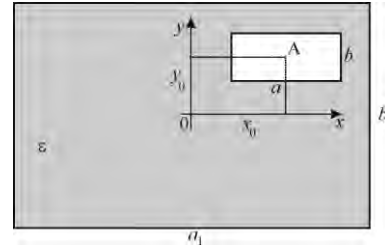


Fig. 4. Rectangular coaxial line with offset inner conductor

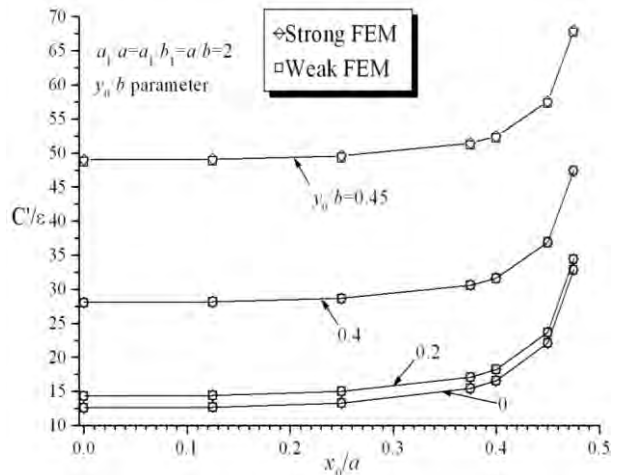


Fig. 5. Ratio  $C'/\epsilon$  depending on  $x_0/a$ , where  $y_0/b$  is a parameter,  $a/b=2$  and dielectric is isotropic, Fig. 4.

### III. Rectangular coaxial line with offset inner conductor and multilayered dielectric

Fig. 6 shows the structure with layered isotropic dielectric in which the inner conductor was moved in direction  $t$ .

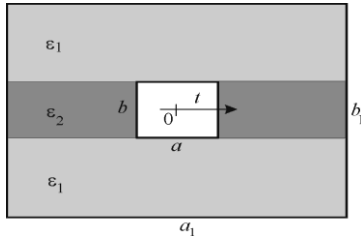


Fig. 6. Rectangular coaxial line with offset inner conductor and multilayer isotropic dielectric

In Fig. 7 dependence of the normalized effective permittivity  $\varepsilon_e / \varepsilon_1$  on  $\varepsilon_1 / \varepsilon_2$  for two different values of  $t/b$  for a square coaxial line from Fig.1 is shown, where  $a_1/a = a_1/b_1 = a/b = 2$ .

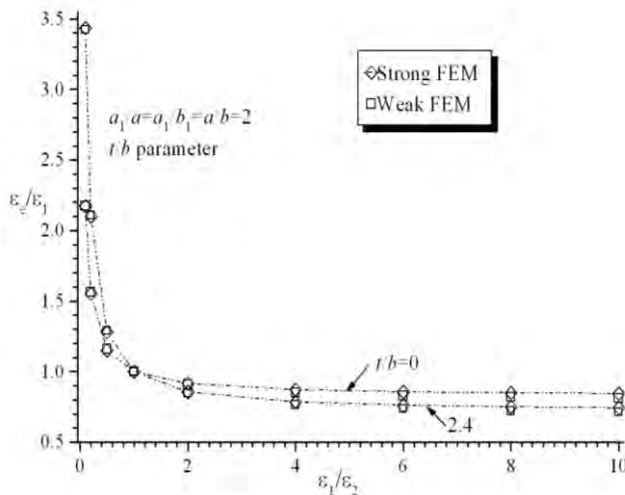


Fig. 7. Normalized effective permittivity  $\varepsilon_e / \varepsilon_1$  of a rectangular coaxial line with offset inner conductor and multilayered isotropic dielectric, Fig.6, for two different values of ratio  $t/b$ .

### IV. Square coaxial line with offset inner conductor and anisotropic dielectric

For a square coaxial line with offset inner conductor, Fig. 1, for  $b/a=4$ , filled with anisotropic dielectric Sapphire, where  $\bar{\varepsilon} = [\varepsilon_x \varepsilon_y]$ , results for relative permittivity  $\varepsilon_{re}$ , are presented in Fig. 8, for the following cases: a)  $\varepsilon_x = 9.4$ ,  $\varepsilon_y = 11.6$  and b)  $\varepsilon_y = 9.4$ ,  $\varepsilon_x = 11.6$ . The required number of unknowns for strong FEM formulation is 1152 and for weak FEM formulation is 2448, whereas the number of rectangular elements is 288. On the other hand, FEMM requires the number of unknowns between 3964 and 4088, whereas the number of triangular elements is between 7559 and 7804. From Fig. 8 both effects of the proximity and anisotropy can be noticed, as described in detail in [2, 4, 5].

Moreover, an excellent agreement with FEMM results can be noticed, which proves that the strong FEM can be successfully applied for an accurate and efficient calculation of rectangular coaxial line with offset anisotropic dielectric.

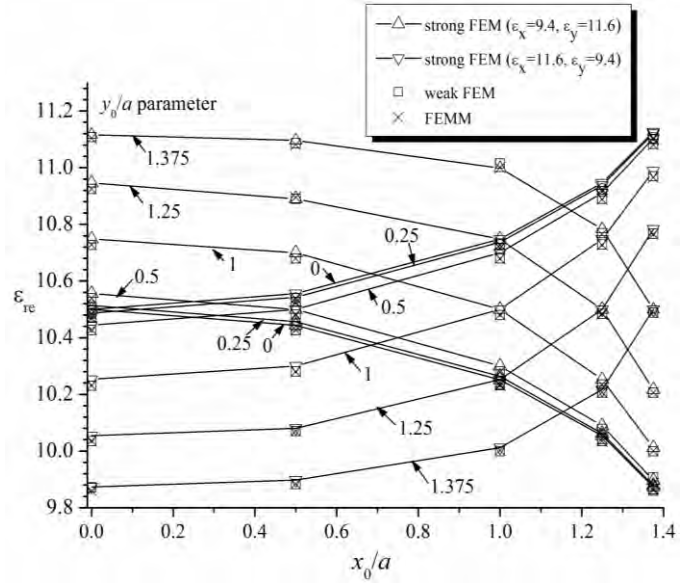


Fig. 8. Effective relative permittivity  $\varepsilon_{re}$ , of a rectangular coaxial line with offset inner conductor and anisotropic dielectric Sapphire, Fig. 2, for different ratios  $y_0/a$ .

### CONCLUSION

Based on numerical examples shown in section III it can be concluded that the strong FEM formulation of the higher order and hierarchical basis functions can successfully be applied for accurate and efficient analysis of transmission lines with offset inner conductor of finite thickness in the case of isotropic and anisotropic dielectrics. Excellent agreement of obtained results and those obtained by weak FEM and commercial software FEMM has been observed. The advantage of strong FEM formulation compared to weak FEM is approximately one half of the number of unknowns. The advantage of both strong and weak FEM, is more than 25 times smaller number of required finite elements with respect to FEMM.

### ACKNOWLEDGEMENT

This research is supported by Serbian Ministry of Education, Science and Technological Development (Project TR-32052 MNTR)

### REFERENCES

- [1] V. Petrović, Žaklina J. Mančić, Calculation of Capacitance of Rectangular Coaxial Line with Offset Inner Conductor bu Using Weak FEM Formulation, Telecommunication in modern Satellite, Cable and Broadcasting Services (TELSKS), 2015, Pages:342-345, DOI:10.1109/TELSKS.2015.7357803

- [2] Ž. J. Mančić, V. V. Petrović, "Strong FEM Calculation of the Influence of the Conductor's Position on Quasi-Static Parameters of the Shielded Stripline With Anisotropic Dielectric", In Proceedings of the ICEST conference, Niš, 2011, pp. 191-194, (ISBN 978-86-6125031-6).
- [3] Ž. J. Mančić, V. V. Petrovic, "Strong and Weak FEM Formulations of Higher Order for Quasi-Static Analysis of Shielded Planar Transmission Lines", *Microwave and Optical Technology Letters (MOTL)*, Vol. 53, No. 5, pp. 1114-1119, May 2011. (DOI [10.1002/mop.25917](https://doi.org/10.1002/mop.25917), online ISSN 1098-2760).
- [4] Žaklina J. Mančić, Vladimir V. Petrović, Analysis of a square coaxial line with anisotropic substrates by strong FEM formulation, *Facta universitatis - series: Electronics and Energetics*, vol. 28, br. 4, pp. 625-636, 2015.
- [5] Mančić, Ž. J. and Petrovic, V. V., "Strong FEM Formulation for Quasi-Static Analysis of Shielded striplines in Anisotropic Homogeneous Dielectric", *Microwave and Optical Technology Letters (MOTL)*, Vol. 54, No. 4, pp. 1001-1006, April 2012. (DOI [10.1002/mop.26676](https://doi.org/10.1002/mop.26676)).
- [6] [http://www.femm.info/Archives/bin/femm42bin\\_x64.exe](http://www.femm.info/Archives/bin/femm42bin_x64.exe)
- [7] A. Milovanović, B. Koprivica, "Calculation of Characteristic Impedance of Eccentric Rectangular Coaxial Lines", *PRZEGLĄD ELEKTROTECHNICZNY (Electrical Review)*, ISSN 0033-2097, R. 88 NR 10a/2012. (<http://pe.org.pl/articles/2012/10a/54.pdf>)
- [8] Z. Pantic, R. Mittra: „Quasi-TEM analysis of microwave transmission lines by the finite-element method“, *IEEE Trans MTT* 34 (1986), 1096–1103.
- [9] COMSOL\_\_\_\_Multiphysics Modeling Software, ([www.comsol.com](http://www.comsol.com))
- [10] A. B. Manić, S. B. Manić, M. M. Ilić, and B. M. Notaroš, "Large anisotropic inhomogeneous higher order hierarchical generalized hexahedral finite elements for 3-D electromagnetic modeling of scattering and waveguide structures," *Microwave and Optical Technology Letters (MOTL)*, vol. 54, no. 7, 2012, pp. 1644–1649.
- [11] M. M. Ilić, A. Ž. Ilić, and B. M. Notaroš, "Efficient Large-Domain 2-D FEM Solution of Arbitrary Waveguides Using *p*-Refinement on Generalized Quadrilaterals," *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, No. 4, April 2005, pp. 1377-1383.
- [12] Ž. J. Mančić, V. V. Petrović, "Strong FEM formulation for 2D quasi-static problems and application to transmission lines", *Invited Paper*, 22nd Telecommunications Forum, TELFOR 2014, Belgrade, 25-27.11.2014.
- [13] V. Petrović, B.D. Popović, "Optimal FEM solution for onedimensional EM problems", *Int. J. of Numerical Modelling* Vol. 14, No. 1, pp., 49-68, Jan-Feb 2001.
- [14] J. Jin, *The Finite Element Method in Electromagnetics*, New York: Wiley, 1993.
- [15] R. F. Harrington: *Field computation by moment method*, Macmillan, New York, 1968.



# Design and Analysis of Realistic Vehicle Traces Model based on the Evolutionary Algorithms

Danijel Čabarkapa<sup>1</sup> and Petar Pavlović<sup>2</sup>

**Abstract** – Vehicular ad-hoc networks (VANETs) are subclass of mobile ad-hoc networks (MANETs). VANETs use vehicles as mobile nodes to provide communication among nearby vehicles and between vehicles and nearby roadside equipment. Due to several constraints such as reproducibility, logistic and considerably economic cost of implementing, most research in VANETs relies on simulations. A key component for VANETs simulations is a vehicular mobility model. The quality of these simulations strongly depends on the degree of reality of the vehicular mobility model. The current trend in vehicular mobility modeling is the generation of vehicular traces at a citywide area or region scale. The main part of the paper focuses on the realistic simulation of vehicular traces generated by population-based metaheuristics Evolutionary Algorithms (EAs). In addition, featured solutions in EAs domain for generating realistic vehicular traces are briefly introduced and their pros and cons are analyzed.

**Keywords** – VANETs, vehicular traces, traffic simulation model, evolutionary algorithms, traffic network simulator

## I. INTRODUCTION

VANETs are advanced dedicated wireless networks that support cooperative driving among a large number of dynamically moving communicating vehicles on the road. Vehicles perform as communication nodes or relays, forming highly dynamic vehicular networks together with other nearby vehicles or with nearby roadside equipment. VANETs provide both Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) dedicated communication [1]. VANETs have specific characteristics that distinguish them from typical mobile ad hoc networks. Vehicles do not move at random and they are limited to known paths road topology while moving, often in a predictable manner. Additionally, a specific vehicle might have only predictable routes. If the road information is available, it is possible to predict the future position of a vehicle or get information about various risk traffic events and accidents. Generally, variable network traffic density mostly depends on the time and the area, and usually at rush hours the traffic is high and it is low in rural or suburb areas [2].

The majority of applications, protocols and communication algorithms proposed in VANETs are designed to improve active safety in driving, efficiency and travel convenience. Developing vehicular applications and protocols usually requires experimental expensive testbeds and real simulation

tools. Real-world simulations for VANETs require realistic network and mobility models. Due to several constraints such as reproducibility, economic costs and lack of scalability, simulation is one of the most often used methods for performance evaluation. The recent challenge in mobility modeling process is the synthetic generation of realistic vehicular traces (at geographical and temporal domain) as an input to a network simulator [3]. The current research trend in realistic vehicular traces modeling is based on evolutionary algorithms (EAs). The EA model uses freely available source data - geographical from online digital maps, and set of traffic volume counts corresponding to the region covered by the digital map. The automatic counting of the vehicle traffic comprises a set of counting roadside devices (induction loops, radars) installed on main roads and highways. Collected data describes the cumulated volume of the traffic flow over a particular spot and can be distinguished regarding time, direction or type of a vehicle.

The rest of the paper organized as follows. Section II describes the basic concepts of realistic VANET simulation, while Section III focuses on EAs approach used for optimizing vehicular mobility models. Section IV presents some of the related solutions in the field of EAs vehicular traces optimization, and we finally conclude in Section V.

## II. REALISTIC VANET SIMULATIONS

Vehicular traffic simulators generally can be classified into macroscopic, microscopic and mesoscopic. Macroscopic models consider traffic flow, density and velocity of vehicles. Microscopic approach considers the movement of each individual vehicle (acceleration-deceleration, line change...) and mesoscopic models consider some interactions among vehicles at an individual level.

There are three classes of VANET mobility models: trace-based, survey-based and traffic simulator-based [4]. In the first class, mobility patterns are extracted directly from real-world mobility movement traces. A collections of datasets can be generated from traces obtained by GPS tracking of vehicles or by commercial vehicles (public busses, taxis). Such traces have a limited availability and are limited to the type of tracked vehicles. In the survey-based models mobility patterns are derived from traffic statistics (arrival times at work, breaks, pedestrian and vehicular dynamics etc.) at the macroscopic level. Traffic simulator-based mobility models based on microscopic traffic simulators. It determines the movement of each vehicle at the microscopic level (breaking, acceleration, energy consumption, noise level monitoring etc.). This class of mobility models can realistically simulate road infrastructure and interactions between vehicles.

<sup>1</sup>Danijel Čabarkapa is with the Higher School of Professional Technological Studies Šabac, H. Veljkova 10, Šabac 15000, Serbia, E-mail: d.cabarkapa@gmail.com

<sup>2</sup>Petar Pavlović is with the Higher School of Professional Technological Studies Šabac, H. Veljkova 10, Šabac 15000, Serbia, E-mail: petarpavlovic@yahoo.com

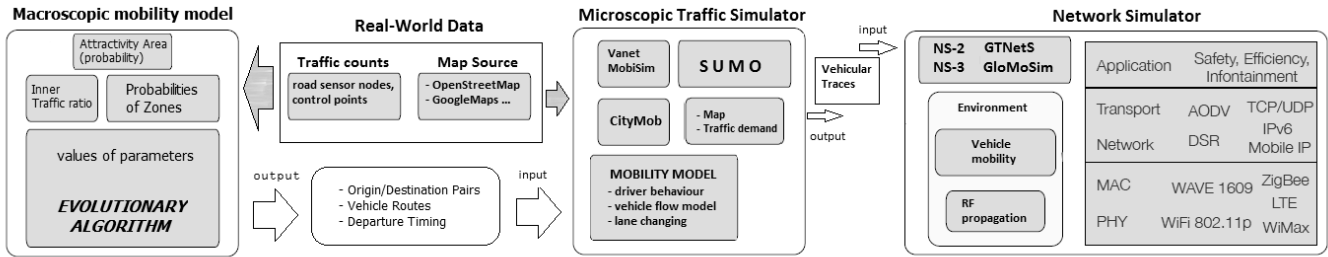


Fig. 1. Generation of vehicular traces and bidirectional coupling between microscopic traffic and network simulators

A recent research trend in mobility modelling is to combine real-world information such as digital maps, traffic counters and statistical data together with microscopic simulation. In order to obtain realistic vehicular mobility at macroscopic and microscopic levels, trace information needs to be used in relationship with microscopic traffic simulation. According to the concept picture in Figure 1, the EA model uses two sources of real-world data inputs (geographical map and traffic volume counts). As outputs, microscopic model generates three elements for each time slot: a prediction of the origin/destination (O/D) pairs for vehicles, a set of a routes for all generated O/D pairs and the estimation of the departure time for vehicles moving in the area covered by the digital map. The output data is then processed by a traffic generator that models traffic demand and generates synthetic traces as an input to a traffic simulator [5].

A microscopic traffic simulator moves vehicles in accordance to requested routes and physical rules. A network simulator based on new vehicles position update its own nodes positions and communications links in every time step. Interactions between particular elements present reciprocal impact and further increase the realism of a simulation. Traffic simulator can change vehicle routes as a result of VANET applications. Traffic generator fuses all real-world data that can be useful to determine the traffic demand and can use a feedback from traffic microsimulator. Information about current traffic situation can influence the traffic demand by changing traveler decisions and adjusting activity schedules. The separation of particular steps ensures modularity what makes easier the replacement of each module and testing of different scenarios. Although many microscopic simulators enable to specify traffic demand integrally, researchers tend to implement a separated module to gain the flexibility and modularity of the platform [6].

### III. EA ALGORITHM - OPTIMIZATION BASICS

Evolutionary algorithms (EAs) are a family of nature-inspired computational techniques and interactive heuristics that evolve a set of candidate solutions, represented as individuals that are grouped in a population. That candidate solutions are able to reproduce themselves to an additional selection procedure. Implementation of an EA begins with a definition of the search space as a finite bounded domain. Parameters are the population size ( $\alpha$ ) as well as the number of offspring ( $\beta$ ) that have to be created each generation (see

Algorithm 1). Additionally, a genotypic search space  $G$  must be determined together with a decoding function  $dec: G \rightarrow \Omega$  that determines to which phenotypic candidate solution a genotype is mapped. Ideally, a mapping from genotype to phenotype is bijective [7].

The crucial step is determining a fitness function. The value of the fitness function indicates the amount of closeness to the optimal solution. Using EA implementation is as good as the fitness or evaluation function. Generally, the fitness of an individual determines the probability of its survival to the next generation. The next step is the initialization or selection of the initial population  $PPL(t)$ . Through the next generations of the population, the existing solution is iteratively improved. This iterative process is called generation and stops after some termination condition is met (e.g. predefined number of iterations). Two basic operators are crossover and mutation. Crossover operator takes two individuals (parents), which are combined to form new chromosomes or offspring ( $PPL'$ ,  $PPL''$ ). Iteratively applying the crossover operator, genes of good chromosomes appear more frequently in the population, leading to convergence to the optimal solution. The mutation operator alters one individual to produce a single new solution and introduces random changes into the characteristics of chromosomes. Reproduction involves the selection of chromosomes for the next generation [7]. The processing scheme of the general EA is shown in Algorithm 1 in pseudocode.

#### ALGORITHM 1

##### GENERAL EVOLUTIONARY ALGORITHM EA (PSEUDOCODE)

---

```

1: INPUTS: parameters  $\alpha, \beta \dots$  quality function  $f: \Omega \rightarrow \mathbb{R}$ 
2: PARAMETERS: population size  $\alpha$ , number of offspring  $\beta$ ,
   genotype  $G$ , decoding function  $decod$ 
3:  $t \leftarrow 0$ 
4:  $PPL(t)$  //create a population of size  $\alpha$ 
5: evaluate individuals in  $PPL(t)$  using  $decod$  and  $f$ 
6: while termination criteria not fulfilled do
7:  $E$  //select parents for  $\beta$  offspring from  $PPL(t)$ 
8:  $PPL'$  //create offspring by recombination of individuals in  $E$ 
9:  $PPL''$  //mutate individuals in  $PPL'$ 
10: evaluate individuals in  $PPL''$  using  $dec$  and  $f$ 
11:  $t \leftarrow t + 1$ 
12:  $PPL(t)$  //select  $\alpha$  individuals from  $PPL''$  (and  $P(t-1)$ )
13: end while
14: OUTPUT: best individual in  $PPL(t)$ 

```

---



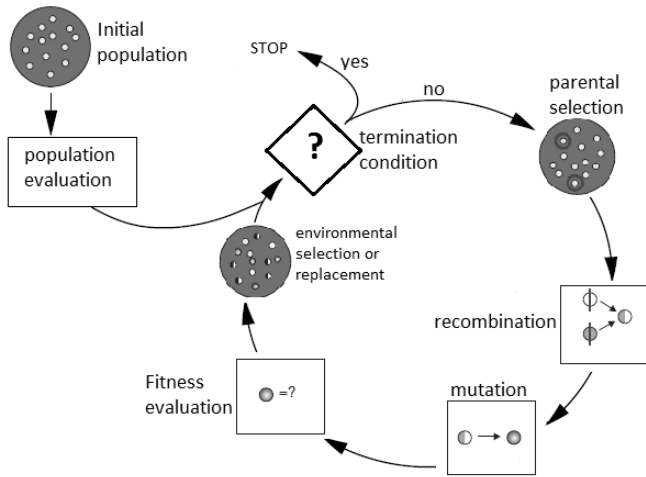


Fig. 2. Schematic description of the fitness evaluation in EAs

Figure 1 shows the evolutionary cycle of EA algorithm. In this section we review standard algorithms and paradigms that are relevant in the remainder of this paper, namely genetic algorithms (GAs), developed by Holland [8] and evolution programming (EP) [9].

#### IV. EVOLUTIONARY ALGORITHMS FOR VANETS

Evolutionary algorithms have been applied to vehicular networks for the last decade. There are still many optimization problems in these complex networks that can be solved using a suitable EA. New architecture of EAs are continuously proposed such as coevolutionary and parallel evolutionary algorithms. During this section, we followed the optimization solution proposed in [10, 11, 12] which defines a EAs generic framework for generating realistic mobility vehicular traces using real-world input traffic data.

The current trend in vehicular traces generation is to combine many approaches into a single process in order to obtain the required level of realism. EA model proposed in [10] generates a set of vehicular traces that consider temporal and spatial aspects of traffic distribution. Mobility model uses freely available source data - from digital OSM (*Open Street Maps*) maps [13], and set of traffic volume counts obtained from roadside control points. This model relies on probabilistic geographical zone surface and attraction points used to select the destination of each vehicle. The residential, commercial and industrial zone types are defined and extracted from OSM maps. Each of them is assigned with a probability of being selected as a destination type. EA model requires the following parameters: zone type, location of zones belonging to each type and location of attractivity areas. The probability for choosing a zone is influenced by the weight of its zone type or the weight of its attractivity area. In the first step EA selects probability of a zone type and then in second step selects the probability of an attractivity area for the selected zone. The third step is applied if within the selected attractivity area more than one zone of selected type exists. This EA model uses simple weighted Dijkstra shortest

path algorithm for the route generation between origin-destination vehicle pair. EAs are iterative heuristics that evolve a set of candidate solutions. Two individuals (parents) are chosen in the population using a given criteria. In the evolutionary cycle they are then recombined with fitness dependent probability to produce an offsprings. The obtained offsprings are mutated and they are evaluated and inserted back into the population following a given criteria [14]. As presented in Fig. 3, after procedure for selection of a destination zone, next step is optimization of EA parameters (fitness function, encoding and genetic operators).

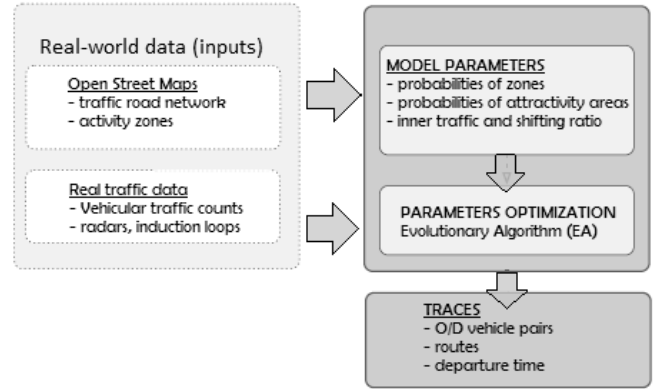


Fig. 3. Schematic EA model of the vehicle trace generation

The fitness of the individuals is the basis for the environmental selection, where for each individual a decision is met whether it will survive and be a potential parent in the next iteration (see Fig. 2). In this EA model fitness is a quality metric and indicates how the generated traces are consistent with real traffic volume counts. We can conclude that the best fitness is able to reproduce such a realistic traffic vehicle behaviour. The EA fitness function  $F$  is computed according to the following equation:

$$F = \sum_{c=1}^C \sum_{t=1}^T |r_c(t) - c_c(t)|$$

Here,  $C$  is the number of control points and  $T$  is the number of time slots. Parameter  $r_c(t)$  is the real traffic volume count at control point  $c$  in time slot  $t$ , and  $c_c(t)$  is the number vehicles at control point  $c$  derived from the generated traffic flows in time slot  $t$ . Generally, the objective is to minimize this sum  $F$  of absolute differences between the real traffic volume counts and the estimated ones for all the control points for the simulated period [11, 12].

Fig. 4 shows how the basic parameters of the model are encoded. This model uses integer gene representation where each gene represents one parameter. Zone probabilities are noted as  $P_T$  where  $T \in \{R, C, I\}$  denotes the zone type. The length of the chromosome depends on the numbers of zone types and attractivity areas. The sum of probabilities of each group must be 100 and it is basic constraint. For the fitness evaluation  $F$  (see equation) all of the values for each gene are scaled and must be in the range from 0 to 1. This model uses a

modified uniform mutation operator which replaces the value of the chosen gene with a random value selected between 0 and 100. As Fig. 4 shows, the gene with value 88 is mutated and replaced by 42. Therefore, gene with value 12 is also changed to 58. [10]

Global Type Probability			Residential Attractivity Areas		Commercial Attractivity Areas			Industrial Attractivity Areas		Inner Traffic Ratio	Shifting Ratio	
$P_R$	$P_C$	$P_I$	$D_R$	$P_{R1}$	$D_C$	$P_{C1}$	$P_{C2}$	$D_I$	$P_{I1}$	$P_{IR}$	$P_{SR}$	
Offspring	20	20	60	12	88	10	70	20	35	65	30	20
Offspring	20	20	60	58	42	10	70	20	35	65	30	20

Fig. 4. EA encoding scheme and uniform mutation operator

In order to obtain more realistic traffic distribution, Cooperative Coevolutionary GA (CCGA) gives more efficient optimization. CCGA proposed and discussed in [15] uses Gawron's algorithm [16]. Model modifications include time-frame reduction, geographical model decomposition and additional attractivity areas. CCGA consists of splitting the whole population into several subpopulations. Instead of evolving a population of similar individuals representing in classical EAs, CCGAs consider the coevolution of subpopulations of individuals representing different species. Each subpopulation runs a genetic algorithm.

The output of the proposed EA and CCGA mobility models is a set of vehicles with their route and ready to be used as an input for SUMO [17] traffic simulator. Finally, the newly generated traces will be compared to the original model accuracy.

## V. TRENDS AND CONCLUSION

This paper presented that realistic vehicular simulation is one of the biggest promising challenges in a VANETs research. We have acknowledged the need of realism in every aspect of simulation in order to obtain reliable results. The paper indicated that the future direction in research of inter-vehicle communication and applications is based on mobility traces. We have presented the main features and restrictions that should be taken into consideration for the use of evolutionary algorithms in generating traces for citywide area for which traffic volume counts exist. Additionally, we have reviewed the main works found in the research literature and we believe that the use of EAs in generating of realistic traces for vehicular mobility simulations is in a very dynamically stage of research.

The major concern of generating realistic vehicular traces is how to select the values of the probabilities associated with attraction points. A genetics operators and fitness function are proposed to model the problem, but in some cases the results notably deviate from real traffic count data. This is due to the route generation process of the EA model. The future research can be EAs expanding with the more tunable and time-variant probabilistic model of areas and zones.

## REFERENCES

- [1] CAR 2 CAR Communication Consortium Manifesto, Ver. 1.1 Overview of the C2C-CC System, 2007. [Online]: [http://elib.dlr.de/48380/1/C2C-CC\\_manifesto\\_v1.1.pdf](http://elib.dlr.de/48380/1/C2C-CC_manifesto_v1.1.pdf)
- [2] Sivasakthi, S. Suresh "Research on vehicular ad-hoc networks (VANETs): An overview" *Journal of Applied Sciences and Engineering Research*, vol. 2, no.1, pp. 23–27, 2013. [Online]: <http://www.ijaser.com/articles/vol2issue12013/lpages/jaser02010003.html>
- [3] J. Harri, F. Filali, C. Bonenet "Mobility Models for Vehicular Ad Hoc Networks: A Survey and Taxonomy" *IEEE Communications Surveys & Tutorials*, vol. 11, no. 4, pp. 19-41, 2009. [Online]: <http://dx.doi.org/10.1109/SURV.2009.090403>
- [4] Institut Eurecom "Mobility models for vehicular ad hoc networks: A survey and taxonomy " Research Report RR-06-168, pp. 3-7, 2006. [Online]: <http://www.eurecom.fr/en/publication/1951/download/cm-haerje-060320.pdf>
- [5] M. Seredynski, P. Bouvry "A survey of vehicular-based cooperative traffic information systems " 14<sup>th</sup> International IEEE Conference ITSC 2011, pp. 163-168 [Online]: <http://dx.doi.org/10.1109/ITSC.2011.6083055>
- [6] S. Uppoor, M. Fiore "Large-scale urban vehicular mobility for networking research" *IEEE Vehicular Networking Conference (VNC)*, pp. 62-69, 2011. [Online]: <http://dx.doi.org/10.1109/VNC.2011.6117125>
- [7] H. G. Beyer "An alternative explanation for the manner in which genetic algorithms operate" *Elsevier BioSystems*, vol. 41, pp. 1-15, 1997. [Online]: [http://dx.doi.org/10.1016/S0303-2647\(96\)01657-7](http://dx.doi.org/10.1016/S0303-2647(96)01657-7)
- [8] J. H. Holland "Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence" University of Michigan Press, Ann Arbor, MI, 1975.
- [9] L. J. Fogel, A. J. Owens, M. J. Walsh "Artificial intelligence through simulated evolution" John Wiley, New York, 1966.
- [10] G. Danoy, P. Bouvry, M. Tabatabaei "Generation of realistic mobility for VANETs using genetic algorithms" *IEEE Congress on Evolutionary Computation*, pp. 1-8, 2012. [Online]: <http://dx.doi.org/10.1109/CEC.2012.6252987>
- [11] A. Grzybek, G. Danoy, P. Bouvry "Generation of realistic traces for vehicular mobility simulations" *DIVAnet'12*, pp. 131-138, 2012 [Online]: <http://dx.doi.org/10.1145/2386958.2386978>
- [12] G. Danoy, P. Bouvry "A vehicular mobility model based on real traffic counting data" 3<sup>rd</sup> International Conference on Communication Technologies for Vehicles, pp. 131-142, 2011. [Online]: [http://link.springer.com/chapter/10.1007%2F978-3-642-19786-4\\_12](http://link.springer.com/chapter/10.1007%2F978-3-642-19786-4_12)
- [13] OpenStreetMap [Online]: [www.openstreetmap.org](http://www.openstreetmap.org)
- [14] B. Dorronsoro, P. Ruiz, Y. Pigne, P. Bouvry "Evolutionary algorithms for mobile ad hoc networks" John Wiley & Sons, 2014. [Online]: <http://eu.wiley.com/WileyCDA/WileyTitle/productCd-1118341139,subjectCd-MA91.html>
- [15] S. Nielsen, G. Danoy, P. Bouvry "Vehicular mobility model optimization using cooperative coevolutionary genetic algorithms" *GECO'13*, pp. 1349-1356, 2013. [Online]: <http://dx.doi.org/10.1145/2463372.2463539>
- [16] C. Gawron "An Iterative Algorithm to Determine the Dynamic User Equilibrium in a Traffic Simulation Model" *Int. Journal of Modern Physics*, 1970. DOI: 10.1142/S0129183198000303
- [17] D. Krajzewicz, C. Rossel "Simulation of Urban Mobility (SUMO)" German Aerospace Centre, 2007. [Online]: [http://www.dlr.de/ts/en/desktopdefault.aspx/tabid-9883/16931\\_read-41000/](http://www.dlr.de/ts/en/desktopdefault.aspx/tabid-9883/16931_read-41000/)

# Transmitting coherent carrier frequencies in modern CATV/HFC systems

Oleg Borisov Panagiev<sup>1</sup>

**Abstract** – In this paper are presented the researches made on the effort to improve the parameters of modern CATV/HFC systems for Downstream. Mathematical equations are drawn and an algorithm for calculation of the digital carrier in HRC/IRC systems for the standards B/G and D/K is proposed. In a table matter are also presented the values of the digital carrier for standard and coherent distribution, as well as the values of BER, MER and SNR.

**Keywords** – CATV/ HFC, HRC, IRC, DVB-C.

## I. INTRODUCTION

This paper is a scientific development for allocation of television channels for the standards B/G and D/K by using DVB-C in modern CATV/HFC systems. The digital carriers are coherent synchronized, which improves the BER, MER, SNR and other parameters. The transmission of coherent digital carrier is realized with the methods of harmonically related carrier/coherent (HRC) and incrementally related carrier/coherent (IRC). Those methods ensure improvement of

the parameters of 20% to 80% depending on the number of channels, the modulation, the frequency spectrum and the frequency allocation of the channels [1], [2], [3], [4]. In North America those methods have already found an application in a number of cable operators. Some of the main manufacturers have designed and are manufacturing television sets [5], [6], [7], [8] which can receive channels from both the standard distribution (STD), and the HRC and IRC distribution (Fig.1). The modulation of the analog carriers is AM-VSB and 8VSB/16VSB/m-QAM of the digital carriers [9], [10], [11]. These channel allocations are made for the FCC: standards ITU-T/J.83B/J.112, CEA-542-B, meanwhile for B/G and D/K there are none. Because in [3] are considered and presented the relations for HRC and IRC transmitting of analogue carriers, in this publication will be presented the mathematical relations and the results for digital carries for the standards B/G and D/K by DVB-C.

## II. STRUCTURAL SCHEMATICS OF A HEAD END WITH COHERENT CARRIERS

The method of transmission with coherent carriers consists of the usage of modified schemas for allocation of the carrier frequencies, which leads to the elimination of some components of the composite three-component beat.

In the Head End the oscillators in the modulators, the converters and the transmodulators are synchronised by a separate circuit from a Main oscillator, where by the synchronizing signal is feed by a shirmed/coaxial cable, and it's frequency  $f_R$  is in the range of megahertz, mostly 1MHz. The unstability of the generated signals must not be worst then  $10^{-6}$ .

Structural schematic of a Head End with coherent carriers is presented on Fig.2, where the generator for synchronisation is a quartz generator and the step  $f_0$  is defined by the type of the synchronization (HRC or IRC), whereby  $f_0 \leq f_R$ .

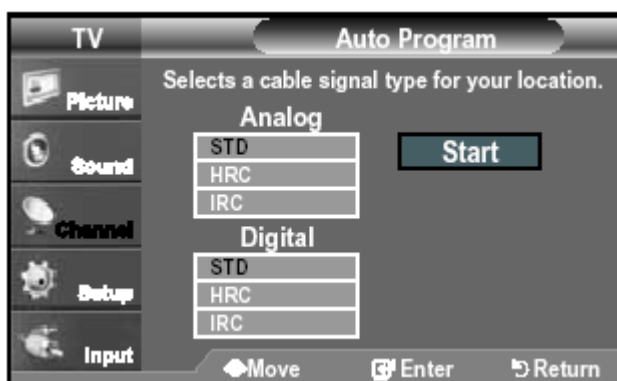
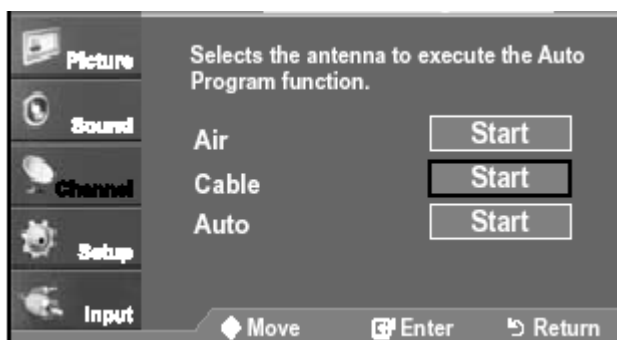


Fig.1. Menu tuning channels

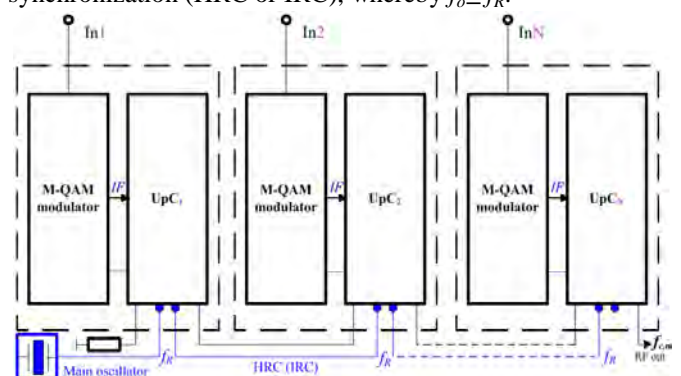


Fig.2. Structural schematic of Head End with coherent carriers

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

The influence of coherent synchronisation on the CATV/HFC system's spectrum when transmitting "channel to channel" is shown on Fig.3a, for decrease (Fig.3b) of the Main oscillator's frequency, as well as for its increase (Fig.3c). In this case the overlapping of the neighbouring channels with each other (which is possible for noncoherent carriers) is lacking, because by frequency alternation (increase or decrease), the whole spectrum is being shifted to the left/right with the value of unstability of the Main oscillator. This way the neighbouring channels cannot overlap with each other and respectively cannot get a throbbing between them, and hence also nonlinear distortions.

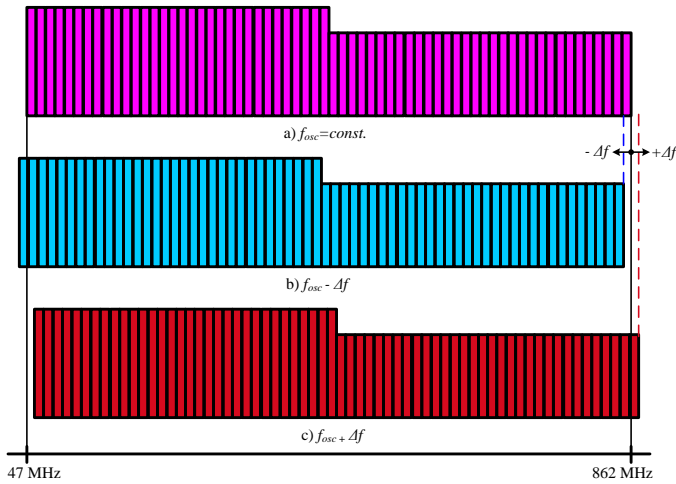


Fig.3. CATV/HFC system's spectrum

### III. TRANSMISSION OF HARMONICALLY RELATED CARRIERS

By this method the carrying frequencies synchronise with the phase of the harmonics of the generator with comb-spectrum, whose main frequency is chosen with value equal to the bandwidth of the channel for the relevant standard.

Since for the standard B/G in VHF range the bandwidth of every channel is 7MHz, and for UHF range 8MHz, the application of the HRC method is impossible. By D/K standard the bandwidth of every channel is 8MHz for both ranges. Because of those reasons are presented below the mathematical expressions for the standard D/K by DVB-C. The carrier frequency of each channel is described by the following formula:

$$f_{c,n} = (k + \mu) f_0, \text{ where} \quad (1)$$

$n = \text{RI, RII, ..., SR1, ..., RVI, ..., SR11, ..., SR21, ..., 21, ..., 69}$ ;  
 $k = 1, 2, 3, ..., 102$  is channel destination (Ch. Des.);  
 $\mu = 5$ ;  $f_0 = 8\text{MHz}$ .

The advantages of this method are:

- coherence of emerging beats of second and third order with  $f_{c,n}$ , in which no essential distortions and worsening of *BER*, *MER*, and *C/N* happen;
- Decrease of the transitional (interchannel) distortions;
- Use of one Main oscillator and others.

### Disadvantages

- Low flexibility;
- Difference in the values of digital carrier to the standard distribution (STD) - central frequency of the channel;
- The signals of NMS, pilot signals and other do not synchronize by phase.

### IV. TRANSMISSION OF INCREMENTALLY RELATED CARRIERS

By the IRC method most of the disadvantages of the HRC method are prevented. The principle of obtaining the digital carrier is the same as with the HRC method, but the step is different.

Here the step of synchronization is part/increment of the frequency of the synchronization signal, respectively of the width of the channel, as the value of  $f_0$  is an integer much smaller than them. Frequently  $f_0 = 100\text{kHz}$ ,  $125\text{kHz}$ ,  $250\text{kHz}$ .

As a disadvantage can be considered the uncoherence of the products of second order with digital carriers, respectively with picture carriers.

By the derivation of the mathematical dependencies will be considered the value of the synchronization step of the carrier frequencies and the parameters of the channels: width, quantity, order number and etc., which determine the number of the harmonic according to  $f_0$ .

For standards B/G and D/K formula (1) becomes [3]:

$$f_{c,n} = k_\mu \cdot f_0, \quad (2)$$

where  $k_\mu$  takes into account the values of  $k$  and  $\mu$  for every carrier frequency in the range  $47 \div 862\text{MHz}$ .

Every digital carrier is made by multiplying a fold integer  $k_\mu$  by the step.

In such case, the value of every digital carrier is an even number and simultaneously it represents the number of the harmonic according to the step.

Main oscillator can work on another frequency, different from the step and few times larger than it, and the value of the step is produced by division of the oscillation frequency (more often by 4, 8 and 10 by  $f_{osc} = 1\text{MHz}$ ).

#### A) Mathematical relations for the standard D/K

The channel distribution for DVB-C refers to the whole frequency spectrum from  $47 \div 862\text{MHz}$ , where equation (2) changes as follows:

$$f_{c,n} = (A + C \cdot k \cdot B) \cdot f_0, \text{ where} \quad (3)$$

$A = 42 \cdot C$ ;  $C = f_R / f_0$ ;  $k = 1 \dots 102$ ;  $B = 8\text{MHz}$ ;  
 $n = \text{RI, RII, ..., SR1, ..., RVI, ..., SR11, ..., SR21, ..., 21, ..., 69}$ .

#### B) Mathematical relations for the standard B/G

Because the channels' bandwidth in the frequency bands VHF and UHF is different ( $B_{\text{VHF}} = 7\text{MHz}$ ;  $B_{\text{UHF}} = 8\text{MHz}$ ) equation (3) changes as follows:



$$f_{c,n} = C \cdot [(42 - s) + k \cdot B] \cdot f_0, \text{ where} \quad (4)$$

$s=0; k=1 \dots 36; B=7\text{MHz};$   
 $s=32; k=37 \dots 106; B=8\text{MHz};$   
 $n=2, \dots, S1, \dots, 5, \dots, S11, \dots, S21, \dots, 21, \dots, 69.$

By comparing equations (3) and (4) we can deduce an aggregate equation for both standards, where when calculating the values for a digital carrier we need to take in a consideration the above mentioned conditions and dependencies for  $n, k, B$  and  $f_0$ .

$$f_{c,n} = C \cdot [\Delta + k \cdot B] \cdot f_0, \text{ or} \quad (5)$$

$$f_{c,n} = C \cdot [(f_b - B - s) + k \cdot B] \cdot f_0, \text{ where} \quad (6)$$

$$A = \Delta \cdot C; \Delta = (f_b - B - s) \text{ and } f_b = f_{c,RI} \text{ or } f_b = f_{c,2}.$$

Based on the deduced mathematical relationships is composed algorithm (Fig.4) for calculating the digital carrier. The numerical results are presented in Table 1 and Table 2.

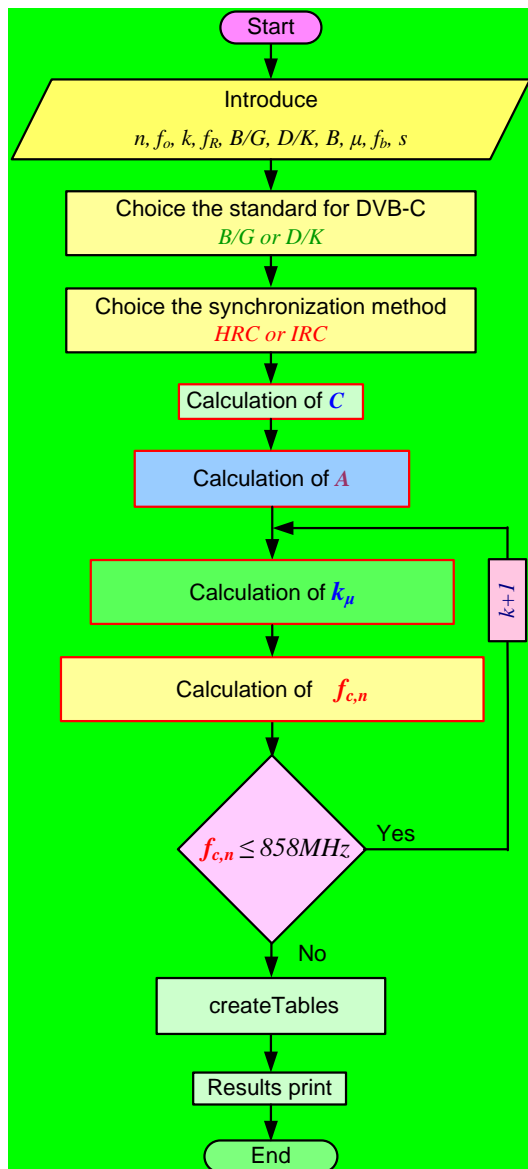


Fig.4. Algorithm for calculating the digital carrier

TABLE I  
CHANNEL DISTRIBUTION FOR D/K (STD, HRC, IRC)

Band	Channel	Channel BW MHz	$f_{c,n}$ MHz			
			STD DVB-C	HRC DVB-C all position Ch. Des.	HRC DVB-C	IRC DVB-C
1	2	3	4	5	6	7
<b>Standard D</b>						
VHF I	R I	48,5-56,5	52,5	48	48	50
	R II	58-66	62	56	56	58
	R III	76-84	80	64	80	82
VHF II	R IV	84-92	88	72	88	90
	R V	92-100	96	80	96	98
S Low	SR1	110-118	114	88	112	114
	SR2	118-126	122	96	120	122
	SR3	126-134	130	104	128	130
	.....	.....	.....	.....	.....	.....
	SR8*	166-174	170	144	168	170
VHF III	R VI	174-182	178	152	176	178
	.....	.....	.....	.....	.....	.....
S High	R XII	222-230	226	200	224	226
	.....	.....	.....	.....	.....	.....
S High	SR11	230-238	234	208	232	234
	.....	.....	.....	.....	.....	.....
S High	SR19	294-302	298	272	296	298
	.....	.....	.....	.....	.....	.....
<b>Standard K</b>						
S Extended (hyper)	SR21	302-310	306	280	304	306
	.....	.....	.....	.....	.....	.....
	SR30	374-382	378	352	376	378
	.....	.....	.....	.....	.....	.....
UHF IV/V	SR41	462-470	466	440	464	466
	21	470-478	474	448	472	474
	.....	.....	.....	.....	.....	.....
	29	534-542	538	512	536	538
	30	542-550	546	520	544	546
	.....	.....	.....	.....	.....	.....
	39	614-622	618	592	616	618
	40	622-630	626	600	624	626
	.....	.....	.....	.....	.....	.....
	49	694-702	698	672	696	698
	50	702-710	706	680	704	706
	.....	.....	.....	.....	.....	.....
	59	774-782	778	752	776	778
	60	782-790	786	760	784	786
	.....	.....	.....	.....	.....	.....
	69	854-862	858	832	856	858
	.....	.....	.....	840	.....	.....
	.....	.....	.....	856	.....	.....

In Table 3 are presented the data for improving the quality of transmitted signals in a IRC system in accordance to a system for cable television with noncoherent distribution of the channels.

TABLE II  
CHANNEL DISTRIBUTION FOR B/G (STD, HRC, IRC)

Band	Channel	Channel BW MHz	$f_{c,n}$ MHz	
			STD DVB-T/C	IRC DVB-C
1	2	3	4	5
<b>Standard B</b>				
<b>VHF I</b>	2	47-54	50,5	49
	3	54-61	57,5	56
	4	61-68	64,5	63
<b>S Low</b>	S2	111-118	114,5	112
	.....	.....	.....	.....
	S10*	167-174	170,5	168
<b>VHF III</b>	5	174-181	177,5	175
	.....	.....	.....	.....
	12	223-230	226,5	224
<b>S High</b>	S11	230-237	233,5	231
	.....	.....	.....	.....
	S20	293-300	296,5	294
<b>Standard G</b>				
<b>S Extended (hyper)</b>	S21	302-310	306	306
	.....	.....	.....	.....
	S30	374-382	378	378
	.....	.....	.....	.....
	S41	462-470	466	466
<b>UHF IV/V</b>	21	470-478	474	474
	.....	.....	.....	.....
	29	534-542	538	538
	.....	.....	.....	.....
	30	542-550	546	546
	.....	.....	.....	.....
	39	614-622	618	618
	.....	.....	.....	.....
	40	622-630	626	626
	.....	.....	.....	.....
	49	694-702	698	698
	.....	.....	.....	.....
	50	702-710	706	706
	.....	.....	.....	.....
	59	774-782	778	778
	.....	.....	.....	.....
	60	782-790	786	786
	.....	.....	.....	.....
	69	854-862	858	858

TABLE III  
BER, MER AND SNR

CATV/HFC system	Parameters		
862 MHz 51 channels DVB-C	<i>BER</i> (RS, Viterbi)	<i>MER</i> dB	<i>SNR</i> dB
noncoherent	$10^{-8}$	36	32
coherent	$10^{-12}$	44	37

## V. CONCLUSION

The results allow concluding that transmitting with HRC synchronization is possible only for the D/K standard, because the bandwidth of each television channels is identical, for both the VHF and UHF bands. The difference in bandwidth of the channels for the B/G standard in VHF and UHF bands do not allow HRC synchronization.

It is possible to use IRC synchronization for both standards by using steps with values divisible by 7MHz, respectively by 8MHz.

The distribution of the channels, the synchronization frequency, the step and the standard are not influenced by the signals' type (SD; HD) and the compression (MPEG-2; MPEG-4).

The coherent methods for synchronizations prevent interference in the channel (n) when a change of the digital carrier in the channel (n-1) and (n+1) occurs - an overlap is absent.

## REFERENCES

- [1] Ciciora, W., J. Farmer, D. Large and M. Adams, *Modern Cable Television Technology*, 2nd ed., Morgan Kaufmann, Elsevier Inc., 2004.
- [2] Large, D., J. Farmer, *Broadband Cable Access Networks: the HFC plant*, 3th ed. Burlington, MA: Morgan Kaufmann/Elsevier, 2009. ISBN 9780123744012.
- [3] Panagiev, O. *Analysis and reduction on nonlinear distortions for signals in the broadband cable communication systems*, Dissertation, Technical university of Sofia, 2006.
- [4] Panagiev, O. B. Intermodulation composite distortions theoretically research in HFC networks. ICEST, Proc. of Papers, vol.1, Ohrid, 24-27 June 2007, pp. 325-328
- [5] LN26A450C1D/LN32A450C1D/LN37A450C1D/LN40A450C1 D, Samsung Electronics America, Inc.105 Challenger Road Ridgefield Park, NJ 07660-0511, [www.samsung.com](http://www.samsung.com).
- [6] SHARP LC-37D6U, [www.sharp.ca](http://www.sharp.ca).
- [7] LT-42E488, LT-37E478, LT-32EX38, JVC company of Amerika, <http://www.eiae.org>, <http://www.jvc.com>.
- [8] HD Encoder / Modulator, Model 5415HD, Channel Plus, (760) 438-7000 USA & Canada (800) 421-1587, Toll Free FAX (800) 468-1340, Nov. 2011, [www.linearcorp.com](http://www.linearcorp.com).
- [9] eSPINNER® indoor cycling bike, Star Trac, USA, Feb. 2009, <http://www.startrac.com>.
- [10] Philips DCD778, Docking Entertainment System DVD Kitchen Radio with dock for iPod, <http://www.philips.com/usasupport>.
- [11] Angelov K., P. Kogias, S. Sadinov, Analysis of the mechanisms in the DOCSIS 3.0 standard to unite the frequency channels in the return channel of cable television networks, UNITECH 2015, Gabrovo, Bulgaria, Proceedings - Volume 2, pp. II-147-II-150, ISSN 1313-230X.



# Control of photon energy sharing with enhanced efficiency for photovoltaic power supply in optical fiber networks

Jovan Shikoski<sup>1</sup>, Juli Zlatev<sup>2</sup>, Tinko Eftimov<sup>3</sup>, Georgi Stanchev<sup>4</sup>

**Abstract** – In the following article we have reviewed two methods of photon energy charge with: an optical splitter and an optical switch for the purpose of smart Sensors charging. Energy efficiency was analysed and compared.

**Keywords** – Smart sensor, Optical splitter, Optical switch, Optical power, Optical fibre, Power IR laser, Photovoltaic power converter PPC-4E.

## I. INTRODUCTION

In the case of high level and high-frequency electric interferences under extreme meteorological conditions, explosive environment etc., it is not appropriate for the sensors to get charged with copper conductors or standard photovoltaics [1,2]. Lately there has been a trend for optical fibers to be used for photon energy transmission in the range of 100mW~2W, generated by a laser diode for the charge of remote smart sensors [3,4]. For that purpose specialized photovoltaic power converters are being offered with a convertor for coupling to optical fiber [5,6] (PPC-E – for 12V, 6V and 4V).

The article reviews the possibility of optical power sharing using two photovoltaic convertors PPC-4E with the purpose of autonomous charging of two separated smart sensors, positioned at a distance from each other.

The sharing of the optical power, coming from the IR laser, to the relevant photovoltaic inputs have been made by a passive optical splitter [7]. The main task of the following release is to present alternative schematic for sharing of optical power, which would have better energy characteristics by specific consumption in the grid circle and conditions of exploitation. There are also algorithms for controlling both types of grid charging.

In addition in the following article both circles of optic power sharing have been analyzed. The main purpose is to compare their energy efficiency as well the conditions in which the suitable method could be used.

<sup>1</sup>Jovan Shikoski is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria.

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

<sup>3</sup>Tinko Eftimov is with University of Telecommunications and Post, 1 Stefan Mladenov street, Sofia 1700, Bulgaria.

<sup>4</sup>Georgi Stanchev is with the Faculty of Mechanical Engineering at Technical University of Sofia. 8 Kl. Ohridski Blvd. Sofia 1000.

## II. PRINCIPLE OF OPERATION

### A. Optical system for distribution of photon power with an optical splitter

A splitter with two outputs has been chosen for the purpose of receiving more energy at the outputs, which has to be delivered to the corresponding photovoltaic by a 62.5  $\mu\text{m}$  core diameter multi-mode optic fiber. The sharing energy at the splitter 1x2 is equal at every output (each 50%) with a loss of 3.83dB.

In the Figure 1 was represented a block diagram of an optical system for optical energy sharing with optical splitter. Microcontroller in the smart sensor SS, parallel with his main functions, observes the condition of the Li- ion battery. When the voltage of the battery reduces to a level of 3V, there is a request for recharging to the central controller transmitted through the optical transmitter on the channel for data CF (optical communication fiber). The central controller switches on and drives the powerful IR laser diode PLD through the PLDD driver. The photon energy from the laser, at the input of the optical splitter OSP, splits equal to both exits. The light energy, through optical fibers PF (electric cable), distributes simultaneously to both photovoltaic converter PPC-4E, then transforms in electric energy and both charger devices BBC in the sensors start charging the batteries. Until the charging process lasts, DC-DC converters supply the sensors with electrical energy, that comes from the input 1 (the voltage of photovoltaic). DC-DC converters have been designed to automatically switch and use the voltage of inputs 2 (voltage of the batteries), when the voltage of inputs 1 are zero (voltage of photovoltaic). That happens, when the charging of the batteries is over. As opposed to the sensor, which requests starting of charging process, the battery of the other sensor is not empty, but after the powerful laser switches on, the photon energy, delivered to the optical splitter, enters equally to both photovoltaic cells and both batteries are starting to recharge simultaneously. The battery with the higher voltage at the beginning of the process will charge to 4.2 V earlier and the sensor SS2 indicates to the central, that the battery is fully charged. The central controller CC checks the ID address of the relevant sensor and confirms that this sensor has not given any request for charging and continues with the charging. When the battery is charged to the level of active mode of sensor SS2, it receives energy trough the DC- DC converter of the photovoltaic, but if the sensor is in sleep mode, then the delivered energy of the photovoltaic would be spend inefficiently. In the moment when the central controller

receives the output signal from the relevant sensor, that the battery is fully charged, the charging process disconnects.

In Figure 2 has been described the specified algorithm of operating of the grid.

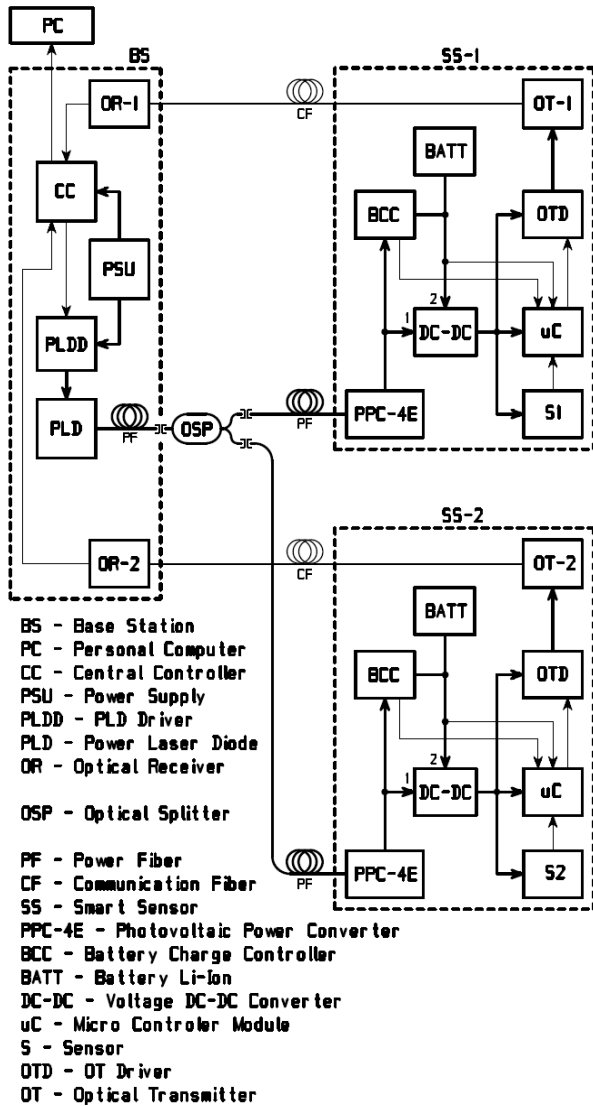


Fig. 1. Optical network with a splitter

### B. Optical system for distribution of photon power with an optical switch

The difference between optical splitter compared with optical switches is, that the photon energy is sharing simultaneously and equal to all of the exits by the splitters and the full energy is distributed only in one output channel, which is the focus of the mirror, by the switches.

The experiments in the article are made for optical switch with 1 and 2 outputs with 62,5  $\mu\text{m}$  core diameter multi-mode optical fiber. The losses on the first exit are 0,58 dB and for the second 0,35 dB. The maximal optic power, which is the limit of working conditions for the switch is 500mW at a

wavelength of 850nm. The switch time from one position to another is 8 ms. The control system of the switch is digital through I2C or TTL interface.

The block diagram of distribution of photon energy with an optic switch is described in Figure 3. When the smart sensor SS1 sends request for charging to the base station, the main controller CC manages the optical switch through I2C interface and aims the mirror to the outputs of the optic fibre to the direction of the relevant sensor. With the help of the PLDD driver the main controller switches on and controls the powerful IR laser diode PLD with a maximum power of 500mW. Compared to the splitter, the optic switch delivers the whole photon energy through the optical fibre to the photovoltaic of the relevant smart sensor. When the battery is charged to 4,2 V, the smart sensor sends a signal to the main station through the communication optical fibre CF, the central controller stops the power laser and the process of charging interrupts. The same process replays by the charging of the other sensor. The frequency of charging of the smart sensor in the grid depends on the time of charging of the batteries. Because the charging of the sensors is successive, it is necessary that the sensors be chosen with different

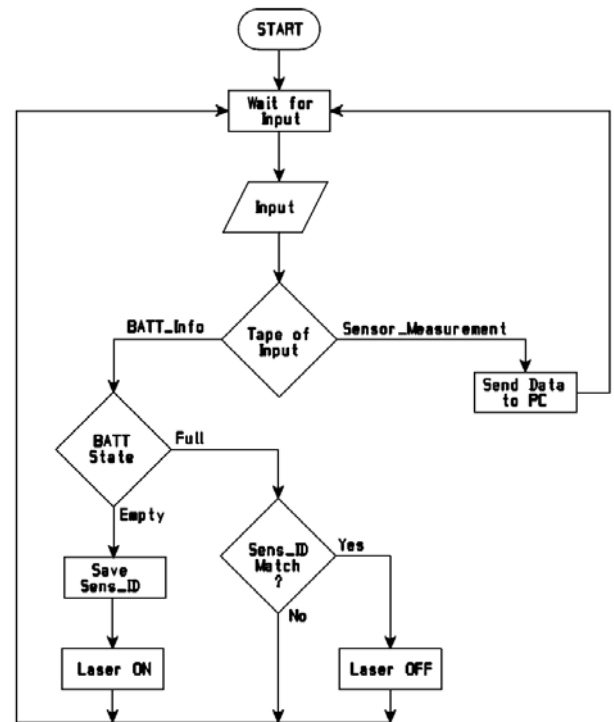


Fig. 2. Algorithm for optical network with a splitter

consumption from the batteries or with phase displacement by the charge process, in order for the system to charge them just in them with energy. The main target is to avoid the situation of charging both of them simultaneously. For the optimisation of the charging time the battery must start to be charged from the moment of voltage under 3,5 V (0,2 V under the nominal value). The consumption and the time of active regime of the sensors to be distributed without critical situations is of great

importance in the design of this type of sensor system, where the both sensors must be charged simultaneously.

The block algorithm which describes the work principles of the optic system for sharing of photon power with optical switch is shown in the Figure 4.

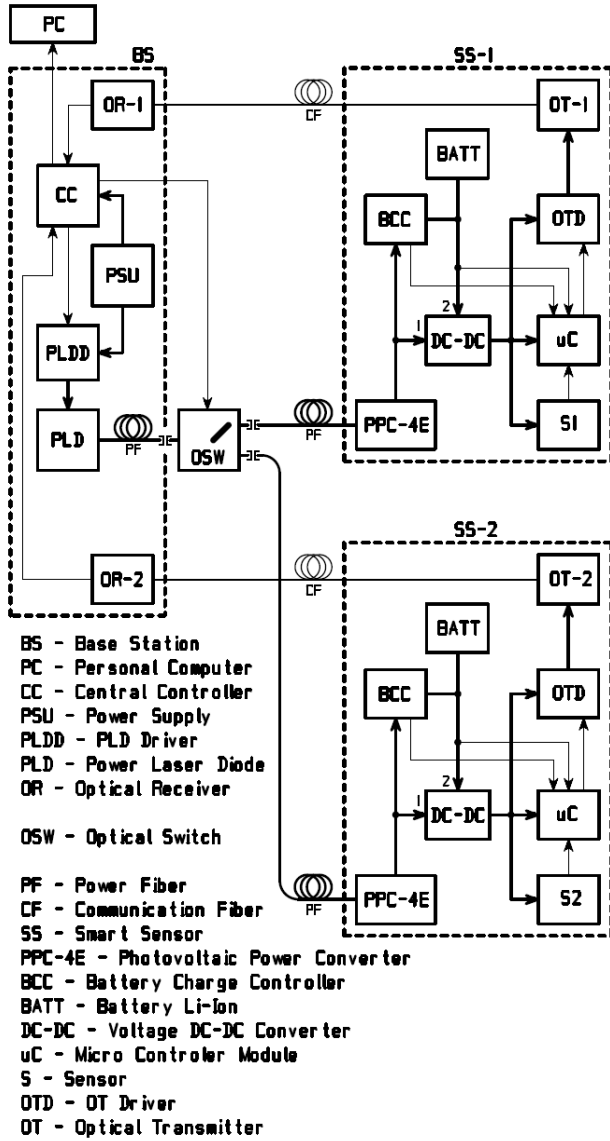


Fig. 3. Optical network with a switch

### III. DESIGN INSTRUCTION

At the beginning of the design of the optic sensor grid must be evaluated the main energy losses as a sum of the losses of every single component from the pick tail of the powerful laser to the photovoltaic device [8]:

$$B = n_{connector} B_{connector} + n_{splice} B_{splice} + z.b_f + B_s \text{ [dB]} \quad (1)$$

as they follow:  $n_{connector}=2$  is the number of connectors from the sensor to the base station,  $B_{connector}=0,5\text{dB}$  are the losses in the connectors,  $n_{splice} = 2$ - number of the weld seams in the optic fiber  $B_{splice}= 0,1 \text{ dB}$ - those are the losses in the welds of the optic fiber,  $b_f = 2,7 \text{ dB/km}$  are the losses for a kilometer through multimode  $62,5 \mu\text{m}$  optic fiber with light with length

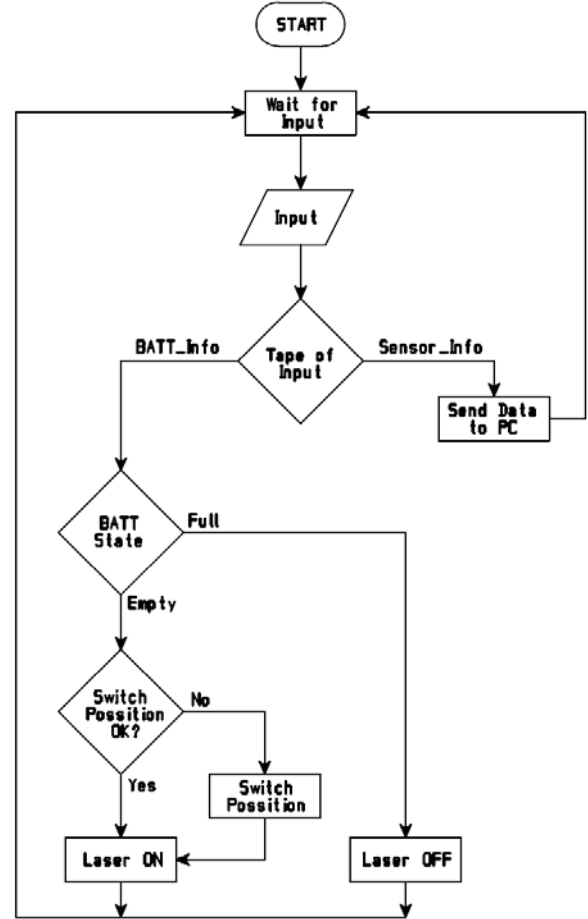


Fig. 4. Algorithm for optical network with a switch

of the wave of  $808\text{nm}$  [9]. The length of the optic fiber in centimeters in “z”, measured in kilometers.  $B_s$  are the losses in the optic distributor as it follows: optic splitter  $B_s=3,83\text{dB}$  and optic switch  $B_s=0,58\text{dB}$  for the first one and  $B_s= 0,35 \text{ dB}$  for the second output. From the equation (1) the main energy loss for the optic system with splitter is  $B=6,38 \text{ dB}$ , for optic system with optic switch is  $B_s=3,13\text{dB}$  for the first and  $B_s=2,9\text{dB}$  for the second output.

The energy design of the optic system is a response of the following equation:

$$P_s = B + P_{pv} \text{ [dBm]} \quad (2)$$

here  $P_s$  is the optical power of the source of light, in our experiment IR laser,  $B$ - the main energy losses in the optic system and  $P_{pv}$  -The optic power on the input of the devices of photovoltaic converter. From equation (2) for the calculation of the provided optic power in the system we obtain the following condition:

$$P_{PV} = P_S - B \quad [\text{dBm}] \quad (3)$$

The IR laser with an optical power of 1W (30 dBm) has been chosen for the optic system of photon energy sharing of with passive splitter, where every photovoltaic have inputs power of  $P_{PV} = 23,6$  [dBm]. In the optic system with optic switch, the power of the IR laser has been reduced to 500mW (27dBm). When we add the losses in (3), we obtain the optic power in the first output of the photovoltaic  $P_{PV} = 23,9$  [dBm] and on the input of the second  $P_{PV} = 24,1$  [dBm].

We convert the power from dBm to W with the following equation:

$$P_{(mW)} = 1mW \cdot 10^{\frac{P_{(dBm)}}{10}} \quad (4)$$

After the convert from (4), by the system with splitter with input optic power of 1W, every photovoltaic of the equipment receives  $P_{(mW)} = 229$  mW. For a system with switch and adjusted power of 500mW input optic power, the hardware of the first photovoltaic receives  $P_{(mW)} = 251$  mW and at the input of the second  $P_{(mW)} = 257$  mW optic power.

In Table 1 we can see the dependency of the maximum output electric power with adjusted different levels of experimental power on the input of the photovoltaic [6].

TABLE I

Electrical power by different levels of optic power for photovoltaic PPC-4E

Optical Power (mW)	100	250	500	750	1000	1500	2000
$P_{\max}$ (mW)	34.8	86	168	240	304	432	512
$V_{mp}$ (V)	4.4	4.3	4.2	4	3.8	3.6	3.2
$I_{mp}$ (mA)	8	20	40	60	80	120	160

#### IV. CONCLUSION

When the times of voltage drop of the batteries are almost the same, the use of an optical splitter sharing of photon energy is the better solution, because it gives the opportunity for simultaneously recharging of all batteries in the optic grid. A larger number of sensors will need a splitter with a larger number of outputs, which comes with higher optic power of the supply laser. The use of a more powerful laser is considered as a disadvantage, because they are more expensive and the increase of the optic power leads to the destruction of the structure of the optic fiber [10]. That's why there is a limit of the optic power through the fibers and it will limit the number of the sensors. A disadvantage of this method of sharing of energy could be the ineffective usage of the energy from the sensors, which need to charge their battery, but they are in sleep mode.

The usage of optic switch as sharing of optic power is the better decision, when the times of discharge are different and

it is possible every single battery to be charged separately. Compared to the splitters, the optical switch sharing of optical power is discrete to every fiber, which allows the usage of laser lower power supply. This makes this type of system more energy effective. Because of the limitation of the power to 500mW, the working power of the switch, the usage of microcontroller and sensors is necessary. The biggest effort comes with the management of the time for charging of the separate sensors and the avoidance of the situation, when two or more sensors must be charged simultaneously. With bigger number of sensor this issue definitively grows. The possible solution could be made with the software with the building of command structure between the sensors in the optic system. That means, that the sensors which have more time in active mode must be with higher priority for recharging in compare with others, which are higher time in sleep mode.

From the analysis and the comparison of the both models for sharing of photon energy can be noted, that the system with optical splitter is technically easier for construction and maintenance and also cheaper.

The method of sharing of optic power with optic switch has better energy effectiveness in the following aspect: the whole optic energy is used for charge, the supply laser is two times less powerful, the effective moment power by the separate sensors is a little bit higher, because of the lower losses in the switch.

#### V. REFERENCES

- [1] JDSU, Photonic power solutions for sensor applications, December, 2006.
- [2] J.G. Werthen, M.J. Cohen, T.C. Wu, and S. Widjaja, ELECTRICALLY ISOLATED POWER DELIVERY FOR MRI APPLICATIONS, Photonic Power Business Unit, JDSU, Milpitas, CA, United States, Proc. Intl. Soc. Mag.
- [3] Furey J., Anaheim, CA (US), Power over optical fiber systems, Ullited States Patent Application Publication, No.: US 2009/0016715 A1, Jan. 15, 2009.
- [4] Wilson C., Kawasaki (JP); Chee S.S., Kokubunji (JP); Nutt L., Houston, TX (US); Yamate T., Yokohama (JP); Kamata M., Kawasaki (JP); Methods and apparatus for photonic power conversion downhole, Ullited States Patent, No.: US 7,696,901 B2, Apr. 13, 2010.
- [5] JDSU, Photovoltaic power converter, 12 V (PPC-12E), datasheet, December 2006.
- [6] JDSU, Power Over Fiber Kit PPM-500-K, April 2014.
- [7] JDSU, Power Over Fiber, March 2014.
- [8] Mitzev C., Dimitrov K., Optical Communications seminar exercises guide, Technical University, Sofia, ISBN:978-954-438-77-8, p.p. 10-27, 2013.
- [9] Corning Incorporated, Corning Clear Curve Multimode Optical Fiber Product Information, , p.p.2-4, 2011.
- [10] Seo K., Nishimura N., Shiino M., Yuguchi R. and Sasaki H., Evaluation of High-power Endurance in Optical Fiber Links, Furukawa Review, No. 24, 2003.



# Polynomial-Based Extraction Procedure for Determination of HEMT Noise Wave Temperatures

Vladica Đorđević<sup>1</sup>, Zlatica Marinković<sup>2</sup>, Olivera Pronić-Rančić<sup>2</sup> and Vera Marković<sup>2</sup>

**Abstract** – The noise wave model defines relationships between the noise wave parameters and the noise parameters. As the noise wave model is related to device intrinsic circuit and available measured transistor noise parameters are related to the whole device, the noise wave parameters are usually extracted using time-consuming optimization procedures in circuit simulators. In this paper, a new, faster and more efficient extraction procedure based on using polynomial functions is presented. The detailed validation of the proposed procedure is done by comparison of the transistor noise parameters of the entire circuit, obtained by using the noise wave parameters extracted by the proposed approach, with the measured transistor noise parameters.

**Keywords** –HEMT, noise parameters, noise wave model, noise wave parameters, polynomials.

## I. INTRODUCTION

The need for precise noise modeling of microwave transistors (MESFETs/HEMTs) that are used in modern microwave communication systems, has led to development of different transistor noise models [1-10]. These models enable implementation of transistors within microwave circuit simulators, which results in the efficient noise analysis. The main classification of transistor noise models is into physical and empirical noise models. Most microwave circuit designers use the empirical noise models since the parameters related to device physical characteristics are often unavailable. The parameters of the empirical noise models are usually extracted from the measured transistor noise parameters ( $F_{min}$  – minimum noise figure,  $\Gamma_{opt}$  – optimum source reflection coefficient and  $r_n$  – normalized noise resistance).

In recent years, the noise wave model treating the noise in terms of waves has appeared as a very appropriate noise model at the microwave frequencies [2], [7-9], [11-19]. The wave representation of noise provides a suitable alternative to the most commonly used representations of noise generated in two-port network based on the equivalent voltage and/or current sources [6]. The noise wave parameters are the noise wave temperatures and it is shown that these temperatures are frequency dependent [14]. The extraction of the frequency dependent noise wave temperatures is usually done using the

optimization procedures in circuit simulators. However, the fact that the optimization procedures are time-consuming makes them relatively inefficient extraction tool.

In this paper, a new, faster and more efficient extraction procedure based on using polynomial functions is proposed. In addition to saving time, the proposed extraction procedure also enables a very accurate transistor noise modeling using the noise wave model.

The paper is organized as follows: after Introduction, Section II contains a short description of the noise wave model. Polynomial-based extraction procedure is presented in Section III. Section IV contains the most illustrative numerical results and obtained observations. Concluding remarks are given in Section V.

## II. NOISE WAVE MODEL OF MICROWAVE FETs

Noise in linear two-port networks can be characterized in many different ways. In the noise wave representation, a noisy two-port network is described by using a noiseless linear equivalent circuit and the waves that emanate from its ports [2].

It is very convenient to use the noise wave temperatures as empirical model parameters, as in that way the noise performance of any two-port network can be completely characterized by the two real temperatures,  $T_a$  and  $T_b$ , and the complex correlation temperature,  $T_c = |T_c| e^{j\omega\tau_c}$ . These temperatures can be expressed in terms of the noise parameters of transistor intrinsic circuit - minimum noise figure,  $F_{mini}$ , optimum source reflection coefficient,  $\Gamma_{opti} = |\Gamma_{opti}| e^{j\varphi_{opti}}$ , and noise resistance,  $R_{ni}$ , as [2]:

$$T_a = T_0(F_{mini} - 1) + \frac{4R_{ni}T_0|\Gamma_{opti}|^2}{Z_0|1 + \Gamma_{opti}|^2}, \quad (1)$$

$$T_b = \frac{4R_{ni}T_0}{Z_0|1 + \Gamma_{opti}|^2} - T_0(F_{mini} - 1), \quad (2)$$

$$T_c = \frac{4R_{ni}T_0\Gamma_{opti}}{Z_0|1 + \Gamma_{opti}|^2}, \quad (3)$$

where  $Z_0$  - the normalization impedance ( $50\Omega$ ) and  $T_0$  - the standard reference temperature (290K).

<sup>1</sup>Vladica Đorđević is with the Innovation Center of Advanced Technologies, Bulevar Nikole Tesle 61, lokal 5, 18000 Niš, Serbia, E-mail: vladica.djordjevic@icnt.rs.

<sup>2</sup>Zlatica Marinković, Olivera Pronić-Rančić and Vera Marković are with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: zlatica.marinkovic@elfak.ni.ac.rs, olivera.pronic@elfak.ni.ac.rs, vera.markovic@elfak.ni.ac.rs.



### III. PROPOSED POLYNOMIAL-BASED EXTRACTION PROCEDURE

As already mentioned, the optimization procedures in circuit simulators that are usually used for extraction of the noise wave temperatures are time-consuming. Instead, the extraction of the noise wave temperatures can be done by using efficient polynomial-based procedure illustrated in Fig. 1, which is detailed as follows:

1. Design the small-signal equivalent circuit schematic of the considered transistor within the standard circuit simulator and implement the noise wave model expressions,
2. Generate  $n$  random samples of the noise parameters of transistor intrinsic circuit ( $F_{min}$ ,  $R_n$ ,  $|\Gamma_{opt}|$ , and  $\varphi_{opt}$ ),
3. For each of  $n$  samples of the noise parameters generated in a previous step, calculate the noise wave temperatures ( $T_a$ ,  $T_b$ ,  $|T_c|$ , and  $\tau_c$ ) using Eqs. (1-3),
4. Assign the calculated noise wave temperatures to the noise wave model implemented within the standard circuit simulator in step 1,
5. Simulate the noise parameters of entire transistor ( $F_{min}$ ,  $R_n$ ,  $|\Gamma_{opt}|$ , and  $\varphi_{opt}$ ) at the considered frequency range, ambient temperature and operating conditions,
6. Apply the Eqs. (1-3) to the noise parameters of entire circuit obtained by simulations, and calculate the fictive noise wave temperatures ( $T_{af}$ ,  $T_{bf}$ ,  $|T_{cf}|$ , and  $\tau_{cf}$ ) referring to the whole circuit,
7. Express the correlations between the noise wave temperatures and the fictive noise wave temperatures for the considered frequency range by using first degree ( $m = 1$ ) polynomials,
8. Based on the measured noise parameters of the considered transistor, calculate the fictive noise wave temperatures ( $T_{af}$ ,  $T_{bf}$ ,  $|T_{cf}|$ , and  $\tau_{cf}$ ) using Eqs. (1-3),
9. Use the calculated fictive noise wave temperatures and obtained correlations expressed by polynomials to determine the noise wave temperatures ( $T_a$ ,  $T_b$ ,  $|T_c|$ , and  $\tau_c$ ),
10. In order to validate the proposed extraction procedure, assign the extracted noise wave temperatures to the noise wave model implemented within the standard circuit simulator in step 1,
11. Simulate the noise parameters of entire transistor ( $F_{min}$ ,  $R_n$ ,  $|\Gamma_{opt}|$ , and  $\varphi_{opt}$ ) at the same frequency range, ambient temperature and operating conditions as in step 5,
12. Compare the obtained noise parameters with the measured ones,
13. If the obtained results do not have satisfactory accuracy, increase a polynomial degree ( $m$ ) by 1, and repeat steps from 7-12. Otherwise, use the noise wave temperatures extracted in step 9.

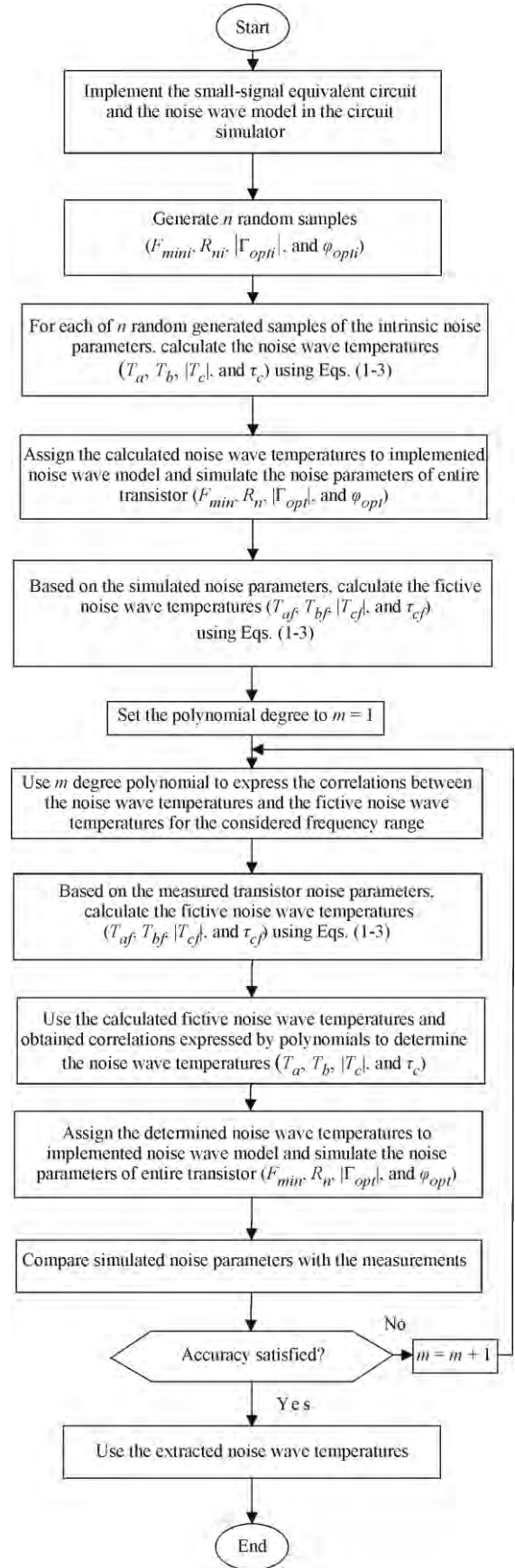


Fig. 1. Proposed polynomial-based extraction procedure flowchart.

#### IV. NUMERICAL RESULTS

The proposed polynomial-based extraction procedure was applied to a packaged HEMT, type NE20283A by NEC, and some of the obtained results are presented in this paper. All simulations were performed using microwave circuit simulator ADS (Advanced Design System) [20]. Measured  $S$  and noise parameters have been available in the frequency range 6-18 GHz over the temperature range 233-333 K, step 20 K [21].

The equivalent circuit of a packaged HEMT used in this research is shown in Fig. 2 [21]. The intrinsic circuit is denoted by the dashed line, and it is common to the most of microwave FET models. The remaining elements embedded in the extrinsic circuit represent parasitic effects and the package.

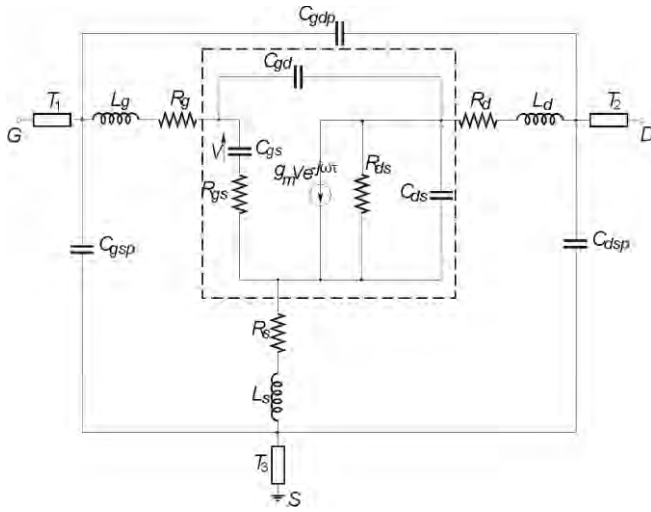


Fig. 2. Equivalent circuit of HEMT in packaged form.

As can be seen in Fig. 2, there are 19 equivalent circuit elements (ECPs). The values of small-signal ECPs of the considered transistor were taken from [21].

First, the values of the noise parameters of transistor intrinsic circuit were generated randomly. Then, the noise wave temperatures determined from these random generated intrinsic noise parameters were used to obtain the fictive noise wave temperatures within ADS [20] circuit simulator. In this case, for the considered frequency range, the correlations between the noise wave temperatures and the fictive noise wave temperatures were expressed with high accuracy by the first degree polynomials:

$$T_a[\text{K}] = x_a + y_a T_{af}[\text{K}] + z_a f[\text{GHz}], \quad (4)$$

$$T_b[\text{K}] = x_b + y_b T_{bf}[\text{K}] + z_b f[\text{GHz}], \quad (5)$$

$$|T_c|[\text{K}] = x_c + y_c |T_{cf}|[\text{K}] + z_c f[\text{GHz}], \quad (6)$$

$$\tau_c[\text{ps}] = x_d + y_d \tau_{cf}[\text{ps}] + z_d f[\text{GHz}], \quad (7)$$

where  $x_{a-d}$ ,  $y_{a-d}$  and  $z_{a-d}$  are the polynomial coefficients given in Table I for different ambient temperatures, and  $f$  is the frequency. The frequency is included in Eqs. (4-7) because the noise parameters of transistor extrinsic circuit are frequency dependent. Also, all the units of variables in Eqs. (4-7) were marked within square brackets.

TABLE I  
POLYNOMIAL COEFFICIENTS FOR DIFFERENT AMBIENT TEMPERATURES

		233 K	253 K	273 K	293 K	313 K	333 K
$T_a$	$x_a$	-82.09	-99.12	-109.3	-82.07	-48.95	-54.6
	$y_a$	1.507	1.584	1.567	1.257	0.97	0.976
	$z_a$	4.92	4.973	4.836	5.9	7.662	8.364
$T_b$	$x_b$	-98.4	-77.26	-149.1	-164.2	-100.1	-110.7
	$y_b$	1.54	1.218	1.702	1.72	1.233	1.244
	$z_b$	7.226	8.385	9.251	9.852	10.48	11.63
$ T_c $	$x_c$	-70.38	-67.85	-107.3	-171.4	-102	-105.1
	$y_c$	1.178	1.055	1.248	1.618	1.136	1.096
	$z_c$	8.696	9.901	11.7	13.62	13.47	14.85
$\tau_c$	$x_d$	25.3	24.23	23.49	23.37	24.24	24.75
	$y_d$	-0.098	-0.065	-0.046	-0.044	-0.046	-0.058
	$z_d$	-0.332	-0.345	-0.348	-0.346	-0.372	-0.372

The polynomial coefficients from Table I and the fictive noise wave temperatures calculated from the measured transistor noise parameters were used for determination of the noise wave temperatures. In order to validate the proposed polynomial-based extraction procedure, the determined noise wave temperatures were assigned to the noise wave model implemented within ADS [20], and the noise parameters of entire transistor circuit were simulated. The simulated noise parameters were then compared with the corresponding measured data.

As an illustration, Fig. 3 presents the simulated  $F_{min}$ ,  $r_n$ , and  $\Gamma_{opt}$  and the corresponding measured data. The results shown in Fig. 3 were obtained for the ambient temperature of 313 K in the frequency range from 6 to 18 GHz. It can be seen that the simulated values of noise parameters are very close to the measured ones, confirming the accuracy of the proposed extraction procedure. The results for the other available temperatures show the same level of the noise modeling accuracy.

#### V. CONCLUSION

Because the noise wave temperatures are frequency dependent, the optimization procedures in circuit simulators usually used for their extraction become time-consuming. For this reason, a new efficient extraction procedure was proposed in this paper. The presented procedure is based on the polynomial correlations between the noise wave temperatures and the fictive noise wave temperatures of entire transistor.

The proposed procedure was applied to a specific HEMT device in a packaged form. Extraction of the noise wave temperatures was carried out based on the polynomial expressed correlations between them and the fictive noise wave temperatures calculated from the available measured transistor noise parameters. Based on the obtained noise wave temperatures, the corresponding noise parameters of entire transistor circuit were calculated in the circuit simulator. A good agreement between simulated and measured transistor noise parameters in a wide range of ambient temperatures proves validity of the proposed procedure.

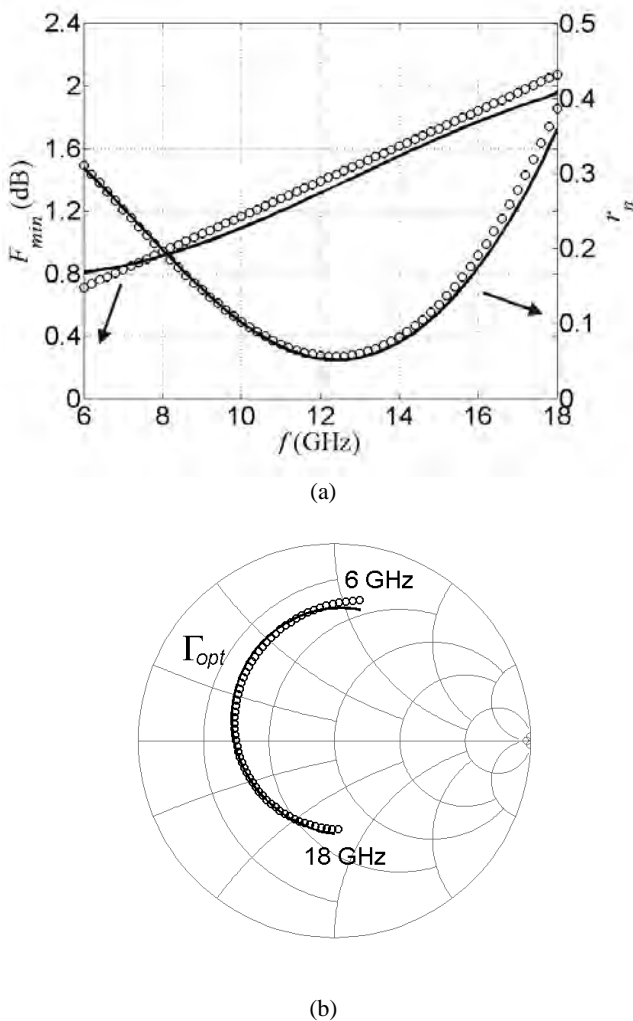


Fig. 3. Measured (symbols) and simulated (lines) values of: (a)  $F_{min}$  and  $r_n$ , (b)  $\Gamma_{opt}$ , depending on the frequency at 313K.

### ACKNOWLEDGEMENT

The work was supported by the TR-32052 project of the Serbian Ministry of Education, Science and Technological Development. The authors would like to thank prof. Alina Caddemi, University of Messina, Italy, for providing the measured data.

### REFERENCES

- [1] R. A. Pucel, H. A. Haus, H. Statz, "Signal and Noise Properties of Gallium Arsenide Microwave Field-Effect Transistors", *Advances in Electronics and Electron Physics*, vol. 38, pp. 195-265, 1975.
- [2] R. P. Meys, "A Wave Approach to the Noise Properties of Linear Microwave Devices", *IEEE Trans Microw Theory Tech*, vol. 26, no. 1, pp. 34-37, 1978.
- [3] H. Fukui, "Design of Microwave GaAs MESFET's for Broad-Band Low-Noise Amplifiers", *IEEE Trans Microw Theory Tech*, vol. 27, no. 7, pp. 643-650, 1979.
- [4] A. Cappy, A. Vanoverschelde, A. Schortgen, C. Versnaeyen, G. Salmer, "Noise Modeling in Submicrometer-Gate Two-Dimensional Electron-Gas Field-Effect Transistors", *IEEE Trans Electron Dev*, vol. 32, no. 12, pp. 2787-2795, 1985.
- [5] M. S. Gupta, O. Pitzalis, S. E. Rosenbaum, P. T. Greiling, "Microwave Noise Characterization of GaAs MESFETs: Evaluation by On-Wafer Low-Frequency Output Noise Current Measurement", *IEEE Trans. Microwave Theory Tech*, vol. 35, no. 12, pp. 1208-1218, 1987.
- [6] M. W. Pospieszalski, "Modeling of Noise Parameters of MESFET's and MODFET's and Their Frequency and Temperature Dependence", *IEEE Trans Microw Theory Tech*, vol. 37, no. 9, pp. 1340-1350, 1989.
- [7] S. W. Wedge, D. B. Rutledge, "Wave Techniques for Noise Modeling and Measurement", *IEEE Trans Microw Theory Tech*, vol. 40, no. 11, pp. 2004-2012, 1992.
- [8] O. Pronić, V. Marković, N. Maleš-Ilić, "The Wave Approach to Noise Modeling of Microwave Transistors by Including the Correlation Effect", *Microw Opt Technol Lett*, vol. 28, no. 6, pp. 426-430, 2001.
- [9] O. Pronić, V. Marković, "A Wave Approach to Signal and Noise Modeling of Dual-Gate MESFET", *AEU-Int J Electron C*, vol. 56, no. 1, pp. 61-64, 2002.
- [10] G. Crupi, A. Caddemi, A. Raffo, G. Salvo, A. Nalli, G. Vannini, "GaN HEMT Noise Modeling Based on 50-ohm Noise Factor", *Microw Opt Technol Lett*, vol. 57, no. 4, pp. 937-942, 2015.
- [11] R. P. Hecken, "Analysis of Linear Noisy Two-Ports Using Scattering Waves", *IEEE Trans Microw Theory Tech*, vol. 29, no. 10, pp. 997-1004, 1981.
- [12] J. A. Dobrowolski, *Computer-Aided Analysis, Modeling and Design of Microwave Networks-The Wave Approach*, Norwood, Artech House, 1996.
- [13] O. Pronić, V. Marković, N. Maleš-Ilić, "MESFET Noise Modeling Based on Noise Wave Temperatures", *TELSIKS'99, Conference Proceedings*, pp. 407-410, Niš, Yugoslavia, 1999.
- [14] O. Pronić-Rančić, V. Marković, "Microwave Transistors Noise Modeling by Using Variable Noise Wave Temperatures", *TELSIKS'01, Conference Proceedings*, pp. 313-316, Niš, Yugoslavia, 2001.
- [15] V. Marković, O. Pronić-Rančić, Z. Marinković, "Noise Wave Modeling of Microwave Transistors Based on Neural Networks", *Microw Opt Technol Lett*, vol. 41, no. 4, pp. 294-297, 2004.
- [16] D. Pasquet, E. Bourdel, S. Quintanel, T. Ravalet, P. Houssin, "New Method for Noise-Parameter Measurement of a Mismatched Linear Two-Port Using Noise Power Wave Formalism", *IEEE Trans Microw Theory Techn*, vol. 56, no. 9, pp. 2136-2142, 2008.
- [17] A. Colliander, T. Narhi, P. de Maagt, "Modeling and Analysis of Polarimetric Synthetic Aperture Interferometric Radiometers Using Noise Waves", *IEEE Trans Geosci Remote Sens*, vol. 48, no. 9, pp. 3560-3570, 2010.
- [18] J. A. Dobrowolski, "Noise Characterization of Differential Multi-Element Multiport Networks - the Wave Approach", *Int J Electron Telecommun*, vol. 61, no. 4, pp. 395-401, 2015.
- [19] V. Đorđević, Z. Marinković, G. Crupi, O. Pronić-Rančić, V. Marković, A. Caddemi, "Wave Approach for Noise Modeling of Gallium Nitride High Electron-Mobility Transistors", *Int J Numer Model Electron Network Dev Field*, 2015, DOI: 10.1002/jnm.2138.
- [20] Advanced Design System, Agilent Eesof EDA, 2009.
- [21] A. Caddemi, A. Di Paola, M. Sannino, "Microwave Noise Parameters of HEMTs vs. Temperature by a Simplified Measurement Procedure", *EDMO'96, Conference Proceedings*, pp. 153-157, Leeds, UK, 1996.



# Study the impact of the instability of oscillators in the Head End modules

Oleg Borisov Panagiev<sup>1</sup>

**Abstract** – Here are presented the results of the researches about the influence of changes in the oscillation frequencies in the modulators and channel converters in a Head End for cable TV. The test channel (n) is the standard channel 46 from the UHF range of B/G standard.

The studies of this article are made for modules forming both analog and digital signals with a one-circular chain PLL.

**Keywords** – CATV/HFC, BER, PLL, transmodulator, upconverter.

## I. INTRODUCTION

The forming of the group signal in the Head End is done in few ways – in serial, parallel and mixed connection of the channel converters' outputs (UpC – in the common case for Downstream) by using splitters and taps. The necessary level of group signal in the whole operating frequency range of the CATV/HFC systems is achieved by:

- Connecting of an electronic amplifier between the output of the last summator and the coaxial (trunk) cable in the CATV system;
- Connecting an optical transmitter (with built in amplifier) between the output of the last summator and the optical fiber/splitter in dependence of the architecture of the cable distribution system.

Regardless which of the two methods is used (CATV or CATV/HFC) the stability of the carrier frequencies ( $f_{pc}$ ,  $f_{sc}$  for analogue AM-VSB signals and  $f_c$  for the digital M-QAM signals), as well as of the intermediate frequencies ( $IF_{pc}$ ,  $IF_{sc}$  and  $IF_c$ ), is from an essential importance for a reliable, qualitatively and seamlessly transmission of information to the subscribers.

There are several methods for supporting the stability of the carrier (channel and intermediate) frequencies of the analogue and digital signals (PLL, DLL, DCM, HRC/IRC), but in the basis of most of them is the PLL (Fig.1), where the stability of the output frequencies is defined mainly by the stability of the reference oscillator (RO). The signals from RO and VCO (voltage controlled oscillator) are feed to a phase detector (PD), who is determining the phase relationships between them. If the signals are in phase, the variation of the output voltage is equal to zero. If there is a phase difference between the two signals, alterations in the output voltage ( $\pm\Delta U$ ) occur, whereat the variations are proportional to the phase difference ( $\pm\Delta\phi$ ).

With the most complicated phase detectors can be also

determined the sign of mismatch of the phases. By a low pass filter (LPF) that eliminates the high frequency composites of the voltage from the PD, ( $\pm\Delta U$ ) go on the VCO. If the LPF is missing, we observe the "Jitter" effect – a sharp variation of the output voltage's fronts. The VCO produces phase corrected sync signals, one of which is used as an input signal for the PD. Because the frequency values of the VCO and RO are different ( $f_{VCO} \gg f_{RO}$ ), it is necessary to include frequency dividers (integer and fractional), by which  $f_R$  and  $f_{VCO}$  become comparable, which itself allows their comparison in the PD. The change of the output frequency of the VCO is performed step-by step programmatically, controlled via microprocessor  $\mu P$ , as a being set different values of the dividers by I<sup>2</sup>C bus or SPI bus [1], [2], [3].

The control can be accomplished in any module or for many modules from an outside/standalone block, in which the  $\mu P$  [4], [5], [6].is situated. In the first case the  $\mu P$  needs to be built in every module [4], [7], [8].

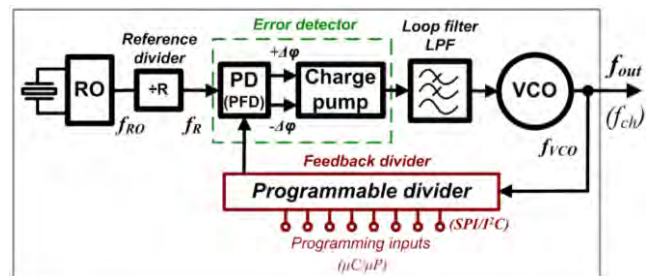


Fig.1. PLL block diagram

## II. PROBLEMS LEADING TO INSTABILITY OF THE OSCILLATOR FREQUENCIES

As it is of significant importance to maintain the carrier frequencies constant in time (lack of detuning until the operator decided to change them), it is required to show, analyze and research the problems, which can lead to their instability.

The causes, which can lead to detuning of the carrier frequency/-ies and are not dependent on the kind of the signals (analogue or digital), which a certain module processes, modulates and converts by frequency. They can be connected with different factors, but their influence is only in the frequency determining, and converting elements: integral circuits, capacitors, diodes, inductances, resistors, quartz resonators and etc., which build VCO; Mixer; divider; PD; RO, and even  $\mu P/\mu C$  as well as SMPS.

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

One of the main reasons, which can cause the unwanted alternation of the carrier frequencies is the so called “cold solder”, since all elements are being soldered to a printed circuit board (PCB). Furthermore, in the multilayer PCD with bad metallization of the openings or breaking of a track on the PCB again some frequency determining circuits will not work at all or partially. Other reasons are the manufacturing tolerances of the parameters of the elements, which appear after a certain period of operation of the modules, temperature dependencies, and damage in the very frequency determining elements, change of power voltages and etc.

The above mentioned reasons affect, however, differently the modules for analog and digital signals. In the present paper are researched the influences of the adjacent lower (n-1) and upper (n+1) channels in the testing channel (n). In the most channel converters (UpC) the receipt of the channel frequency is done by the upper setting of the oscillation frequency:

$$f_{out} = f_{ch} = f_{osc} - IF, [MHz]. \quad (1)$$

This way, for example, for AM-VSB modulator (Fig.2) a synchronization with PLL occurs only with the second sound intermediate frequency ( $IF_{sc,2}$ ), which for B/G standard is 5,5MHz, while for D/K it is 6,5MHz. However, the oscillator frequency  $IF_{pc} = 38,9MHz$ , through which is obtained the first sound intermediate frequency  $IF_{sc}$  (B/G→33,4MHz and D/K→32,4MHz), does not synchronize. By change of the oscillator frequency with  $\pm \Delta IF_{pc}$ , also  $IF_{sc}$  changes and after a frequency conversion can be caused an unfavorable influence (disturbance) in the adjacent upper channel (n), (Fig.3).

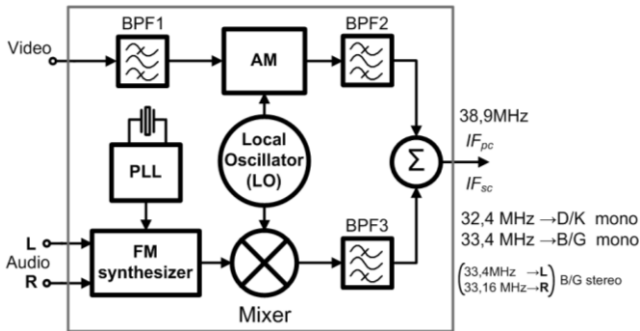


Fig.2. Simplified AM-VSB modulator block diagram

$$IF_{sc} = f_{osc} - IF_{sc,2}, [MHz] \quad (2)$$

$$IF_{sc} = (f_{osc} \pm \Delta f_{osc}) - (IF_{sc,2} \pm \Delta IF_{sc,2}), [MHz] \quad (3)$$

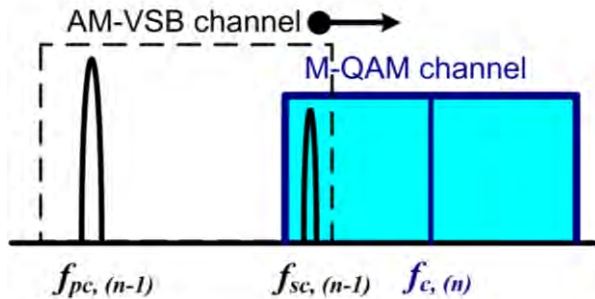


Fig.3. Unfavorable influence in the adjacent upper channel (n)

The synchronization of  $IF_{pc}$ , respectively  $f_{pc}$ , is performed with bi-circular chain PLL in UpC (Fig.4), as through first circle of the PLL,  $IF_{pc}$  and the programmable synthesizer stabilize the frequency of VCO. Through the second circle, the programmable synthesizer, the VCO, Mixer 1 and BPF form  $f_{pc}$  for the corresponding channel, but for  $f_{sc}$  its stability is not support in UpC. In a given moment  $f_{sc}$  increases (i.e.  $f_{sc} > 5,5MHz$  or  $f_{sc} > 6,5MHz$ ) or decreases (i.e.  $f_{sc} < 5,5MHz$  or  $f_{sc} < 6,5MHz$ ). In the first case, the distortion influences negatively the upper (n) channel (Fig.3) and in the second case – the effect is reflected in the channel itself (n-1), as the sound does not reproduce itself or reproduces with distortions.

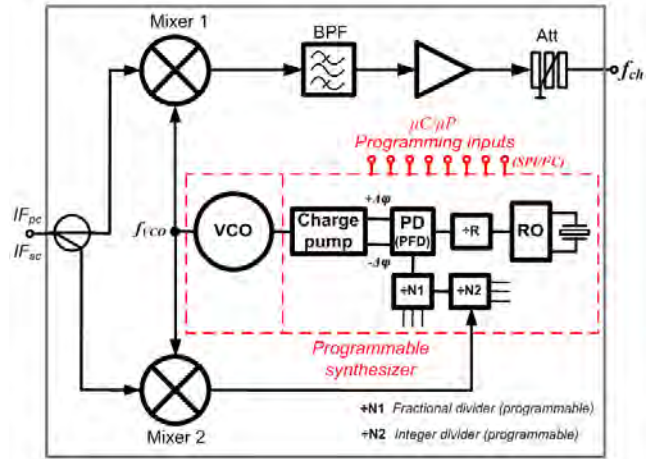


Fig.4. Up converter block diagram

$$f_{ch} = f_{vco} - (IF \pm \Delta IF), [MHz] \quad (4)$$

$$f_{ch} = (f_{vco} \pm \Delta f_{vco}) - (IF \pm \Delta IF), [MHz] \quad (5)$$

By stable  $IF_{pc}$  and  $IF_{sc}$  in the modulator, but unstable  $f_{pc}$  (i.e. decrease  $f_{pc}$  of with  $-\Delta f_{pc}$  or increase of  $f_{pc}$  with  $+\Delta f_{pc}$ ) is negatively influenced over the lower (Fig.5) or upper (Fig.3) channel (n).

The synchronization of  $IF_c=36MHz$  in QAM modulators (Fig. 6) and  $f_c$  in the channel converters (UpC), respectively – transmodulators, is being performed mainly with the one-circular chain PLL. Some manufacturers [8] use infradyne conversion (with two frequency converters), as the second intermediate frequency is out (upper) of the operating frequency range (Fig.7). In such case are used two of each: VCO, Mixer, converter (UpC, DwC), PLL (not tuning, tuning). VCO1 works with one frequency  $\gg 862MHz$  (in this

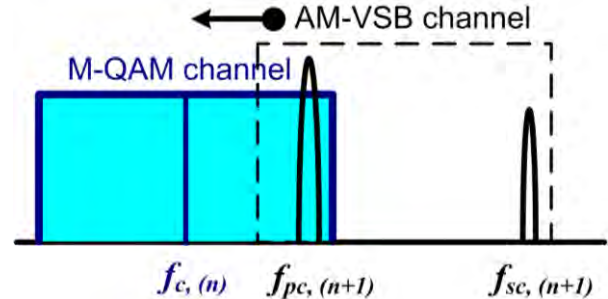


Fig.5. Unfavorable influence in the adjacent lower channel (n)



case  $f_{vco1}=1244\text{MHz}$ ). The management and the settings are performed by  $\mu C$  (MICOM). The stability of the frequencies depends on the foregoing factors, while the influence of the detuning  $\pm\Delta f_c$  and  $\pm\Delta f_c$  is illustrated on Fig.8 and Fig.9 with adjacent to the testing channel (n) disturbing channels (n-1) and (n+1), which are also digital with QAM modulation.

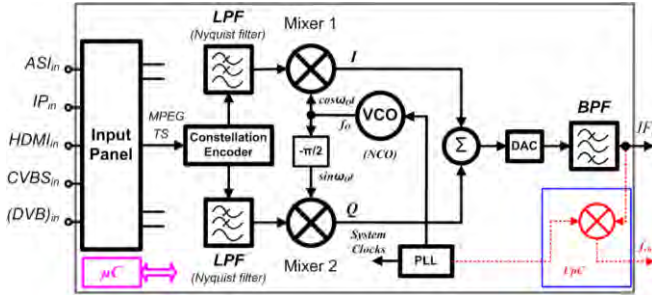


Fig.6. Simplified QAM modulator block diagram

$$IF_c = f_o \pm \Delta f_o, [MHz] \quad (6)$$

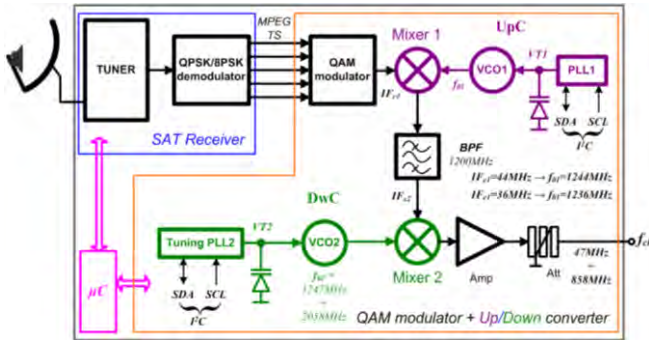


Fig.7. QPSK/8PSK-QAM transmodulator block diagram

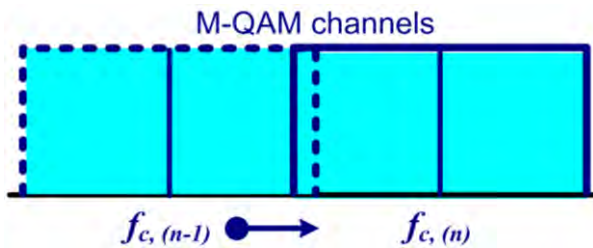


Fig.8. Unfavorable influence in the adjacent upper channel (n)

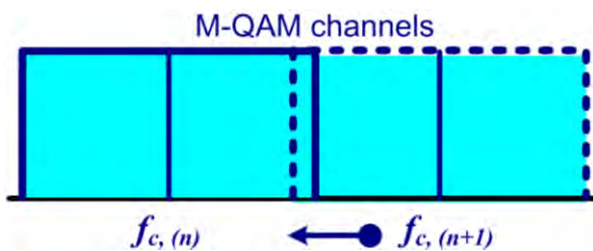


Fig.9. Unfavorable influence in the adjacent lower channel (n)

By all detuning  $\pm\Delta f$  ( $\pm\Delta f$ ) no matter if the signals are analog or digital is produced an overlap of the channels, which leads to worsening of signal parameters and channels.

For the analog, CSO, CTB, C/N, respectively S/N worsens, while for the digital: BER, MER, C/N and etc.

### III. EXPERIMENTAL RESULTS

Here are presented the results of the researches about the influence of changes in the oscillation frequencies in the modulators and channel converters in a Head End for cable TV. The test channel (n) is the standard channel 46 from the UHF range of B/G standard. The signals, which are transmitted in it are digital with 64-QAM, symbol rate 6900ks/s,  $f_{ch}=f_c=674\text{MHz}$ , channel level  $U_{46}=80\text{dB}\mu\text{V}$ . Lower disturbing channel (n-1) is the standard channel 45: analog with AM-VSB modulation;  $f_{pc}=663,25\text{MHz}$ ;  $f_{sc}=668,75\text{MHz}$ ; level of the channel with sound carrier  $f_{sc}$  is  $U_{45,sc}=60\text{dB}\mu\text{V}$ . The upper disturbing channel (n+1) is the standard channel 47: analog with AM-VSB modulation;  $f_{pc}=679,25\text{MHz}$ ,  $f_{sc}=684,75\text{MHz}$ ; channel level with picture carrier  $f_{pc}$  is  $U_{47,pc}=70\text{dB}\mu\text{V}$ .

The change of the corresponding carrier frequency is with a step  $\pm\Delta f=250\text{kHz}$ , as for channel (n-1) sound carrier increases ( $f_{sc}+k\cdot\Delta f$ ) and for channel (n+1) picture carrier decreases ( $f_{pc}-k\cdot\Delta f$ ). Here  $k$  is the serial number of the step, such as its maximal value in the researches is determined by the value of postBER. When postBER reaches values  $\geq 10^{-4}$ , the research is terminated.

The results presented, in Table 1, are for the influence of channel (n-1), and in Table 2, are for the influence of channel (n+1). Fig.10 is constellation diagram of the signal with absence of disturbance, and Fig.11a and Fig.11b are constellation diagrams of the signal with existence of disturbance (with maximal number of the step, respectively  $k=8$  and  $k=7$ ). The levels of the signals are in accordance to the nominal output levels of the modulators, channel converters and transmodulators, as for the disturbing channels are also accounted the influences of varicaps tunable bandpass (BPF) filters over the levels of the signals, whose frequency is outside of passband  $B_{0,7}$ .

TABLE 1  
VALUES OF BER, MER, C/N AT THE INFLUENCE OF  
THE CHANNEL (n-1)

k	$f_{sc}+k\cdot\Delta f$ MHz	► BER (preBER)	BER ► (postBER)	MER dB	C/N dB
0	668,75	4,9E-4	1,2E-9	25	31
1	669	1,0E-3	1,2E-9	25	31
2	669,25	9,8E-4	1,2E-9	25	31
3	669,5	1,0E-3	1,2E-9	25	31
4	669,75	1,0E-3	2,4E-9	25	31
5	670	1,0E-3	1,2E-9	25	31
6	670,25	1,5E-3	4,5E-8	24	31
7	670,5	3,7E-3	1,8E-4	23	31
8	670,75	4,0E-3	4,0E-3	23	31

TABLE 2  
VALUES OF BER, MER, C/N AT THE INFLUENCE OF  
THE CHANNEL (n+1)

k	$f_{pc-k}, \Delta f$ MHz	BER (preBER)	BER (postBER)	MER dB	C/N dB
0	679.25	1.1E-3	6E-9	25	31
1	679	1.3E-3	2E-8	24	31
2	678.75	1.2E-3	1E-8	24	31
3	678.5	1.2E-3	1.6E-8	24	29
4	678.25	1.2E-3	1.2E-8	24	21
5	678	1.2E-3	1.4E-8	24	21
6	677.75	3E-3	3.2E-4	23	25
7	677.5	>4E-3	>4E-3	24	28



Fig.10. Constellation diagram for ch.46 with absence of disturbance



a) from ch.45



b) from ch.47

Fig.11. Constellation diagram for ch.46 with disturbance

## IV. CONCLUSION

The results for the standart B/G can successfully refer also to the 46th channel of the D/K standard [9]. The picture carrier for both standards is the same, so the influence of the (n+1) channel on the channel (n) by decrease of its value (respectively the oscillator frequency) is the same as the present case. Differently stays the case of the influence of the (n-1) channel onto the (n) channel, because the sound carrier at the D/K is closer with 1MHz to the next upper channel. In this case, smaller alterations in the sound carrier lead to overlapping with the (n) channel and deterioration of BER, MER and C/N. The applied approach for channel 46 can be applied for every digital (QAM) DVB-C channel, which is adjacent to an analogue (AM-VSB) channel, where the

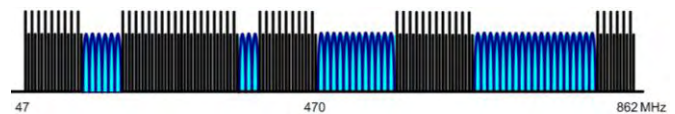


Fig.12. Non standard (mixed) frequency plan  
of CATV/HFC system

number of the channel depends on the frequency plan [10], [11] of the corresponding cable provider. By a standard (classical) frequency plan an overlapping of the examined type: channel (n-1) with the channel (n) is possible only between the last analogue channel and the first digital channel, whereas when other conditions are equal, the non standard (mixed) frequency plan (Fig.12) ensures lower level of nonlinear products from the crossmodulation.

## REFERENCES

- [1] MT-086 Tutorial, Fundamentals of Phase Locked Loops (PLLs), Analog Devices, Inc., 2009.
- [2] W. Kester, "Converting Oscillator Phase Noise to Time Jitter," Tutorial MT-008, Analog Devices, Inc., 2009.
- [3] E. P. Ugryumov, *Cifrovaya shemotekhnika*, 3th ed., BHV-Peterburg, ISBN 978-5-9775-0162-0, 2010.
- [4] <http://wisi.de/en/wisi-group>
- [5] <https://www.kathrein.com/>
- [6] <http://www.televes.com/en/eng/home>
- [7] <http://www.blankom.de/>
- [8] Digital SMATV Transmodulator, Handan Broad Info.Co.,Ltd., <http://handan.en.ec21.com/>
- [9] European standard CENELEC "EN 50083:1-10", 2012.
- [10] O. B. Panagiev, Investigation of Second and Third order distortions Influence in the CATV/HFC networks. ICEST, Proc. of Papers, vol.1, Ohrid, 26-29 June 2013, pp. 33-36
- [11] K. Koitchev, K. Angelov and S. Sadinov, *Design of interactive cable television networks*, Ex-Press, Gabrovo, 2010.



# Mixer Linearization in Direct Conversion Receiver

Aleksandar Atanasković<sup>1</sup>, Aleksandra Đorić<sup>2</sup> and Nataša Maleš-Ilić<sup>1</sup>

**Abstract** – In this paper, the linearization of the mixer in direct conversion receiver is performed by the technique that exploits the baseband signals. The signals for linearization are formed and processed in digital domain, set on the appropriate amplitude and polarity and inserted at the mixer. The linearization effects of the applied linearization method on the third- and fifth-order nonlinearities are observed for the case when the signals for linearization are driven at the transistors' drain of the RF stage differential pair in the Gilbert mixer cell. Additionally, the effects of I/Q signal imbalances on the linearization of the mixer are examined. Analysis are performed for two types of the signal – ideal I/Q signal without imbalances and I/Q signal with imbalance effect (up to 30% amplitude imbalance and 50 degrees phase imbalance). Tests were performed for two different input signal power levels and for two cases of frequency spacing between signals.

**Keywords** – Direct Conversion, Mixer, Linearization method, I/Q imbalances.

## I. INTRODUCTION

The direct-conversion receivers (DCRs), also known as zero-IF receivers, over the last decade have become popular alternative approach to the classical heterodyne architecture in the development of RF integrated circuits (ICs) in modern wireless communication systems. The DCR architecture has become an attractive solution for the commercial applications due to its exquisite characteristics, such as low-cost, low-power, wide bandwidth, and highly integration with RF circuitry. On the other hand, linearity of the receiver become necessary feature and mixer is one of the influential components which can determine system performances. The mixers have frequency-conversion/demodulation function in RF and microwave receivers. The major goals of the mixer design are to minimize conversion loss, noise figure and intermodulation distortion.

Different techniques for the mixer linearization have been deployed, such as predistortion, feedforward, a technique based on transconductance cancelation of the third-order, techniques based on the insertion of the second harmonic and/or the difference frequency signal in the analogue domain [1-5].

The technique applied in this paper for the mixer linearization uses the modified signal in the baseband which is a low-frequency product of the second-order nonlinearity

of a nonlinear system induced by the useful baseband signal, [6], [7]. The in-phase I and quadrature-phase Q components of the signal are digitally processed in order to create adequate signals for linearization, which are tuned in amplitude and polarity and injected at the mixer cell.

The effects of the proposed linearization method are examined through the simulation process for QAM signal at two input power levels, where I and Q components are single tones with frequency interval between spectral components of 0.2 MHz and 2 MHz. Additionally, the impact of the imbalances of the I and Q signals on the intermodulation products is investigated. Output power levels of the fundamental signal, as well as levels of the third- and fifth-order intermodulation products, are observed in terms of the amplitude and phase mismatch of the I and Q signals.

## II. THEORETICAL APPROACH

The direct-conversion receivers translate the desired RF spectrum directly to DC using a local oscillator (LO) which frequency is equal to the RF-carrier frequency of the desired signal. The mixed output is the signal that is downconverted directly to the baseband, so that the IF stage is not required. Figure 1 shows the schematic diagram of the direct-conversion receiver including the mixer linearization circuit.

The theoretical approach of the proposed linearization technique is based on the nonlinearity of the transistor output current [7-9]. The in-phase, I and quadrature phase, Q components are extracted at the demodulator output in the receiver to be adequately processed in the baseband to create signals for linearization:

$$BB_{\text{mod}} = f(I, Q) = I^2 + Q^2 \quad (1)$$

The formed linearization signals are separately adjusted in amplitude and polarity  $a_{\{e/o\}}$  across two branches, as indicated in Figure 1. Indexes,  $e$  and  $o$  in subscript are related to the signals prepared for the insertion in the mixer cell through the serial LC circuit.

According to the analysis performed in [6-9], the second order nonlinearity of the transistor in the mixer cell leads to the interference of the injected baseband signal for linearization and fundamental signal, which generates additional third-order nonlinear products that may suppress the original intermodulation products distorted by the transistor nonlinear characteristic.

<sup>1</sup>Aleksandar Atanasković and Nataša Maleš-Ilić are with the Faculty of Electronic Engineering, University of Niš, Serbia, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: [aleksandar.atanaskovic, natasa.males.ilic]@elfak.ni.ac.rs

<sup>2</sup>Aleksandra Đorić is with the Innovation Centre of Advanced Technology, Niš, Serbia, Bulevar Nikole Tesle 61, 18000 Niš, Serbia, E-mail: alexdjoric@yahoo.com

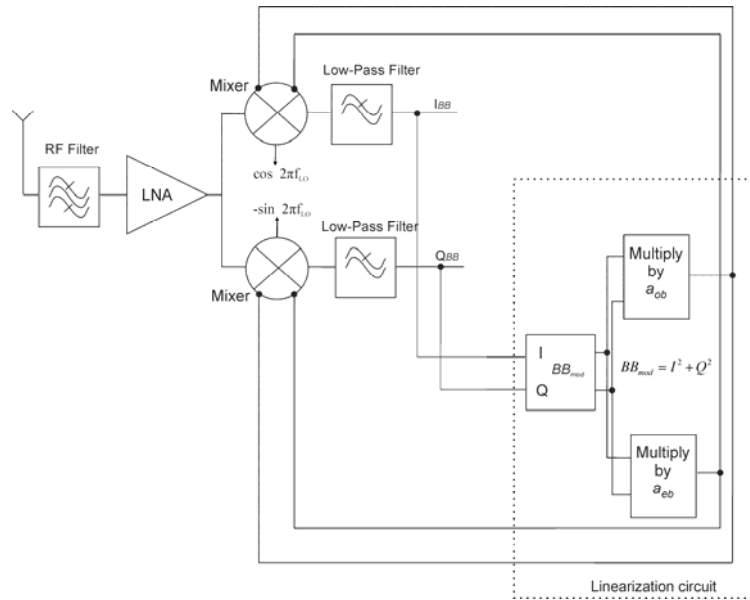


Fig.1. Schematic diagram of the DCR with the mixer linearization circuit

### III. LINEARIZATION RESULTS

The linearization was applied to the Gilbert mixer that is used in the direct conversion receiver (Figure 1). The impact of the performed linearization method on the intermodulation products reduction was analysed through the simulation process in ADS for the mixer cell that uses transistor MOSFET model. The linearization was carried out for the ideal case where I and Q components have equal amplitudes and phase difference of 90 degrees.

The mixer cell was tested for QAM modulated signals that comprise the I and Q single tone baseband components. The frequency spectrum of such a signal contains two spectral components and we considered two cases, when the spectral components are separated by 0.2 MHz and 2 MHz.

The carrier frequency of the input signal is 1 GHz as well as the frequency of the local oscillator. Linearization of the mixer was performed for the cases when the input power of the RF carrier is  $P_{inRF} = -20$  dBm and  $-30$  dBm, while the power of the signal from the local oscillator is  $P_{inLO} = -3$  dBm.

The optimization process of the adjustable parameters of the linearization signals was performed to reduce the third-order intermodulation products, IM3 and to restrain the fifth-order intermodulation products, IM5 at the levels below the suppressed IM3 products.

Figures 2 and 3 show the intermodulation products, IM3 and IM5, before and after the applied linearization method. After applied linearization, suppression of the IM3 products is around 12 dB for higher power level and both frequency spacing. For lower power, the IM3 products are improved about 22 dB for 0.2 MHz frequency spacing and 8 dB for 2 MHz signal separation. On the other hand, the IM5 products are aggravated, but they are still below linearized IM3 products.

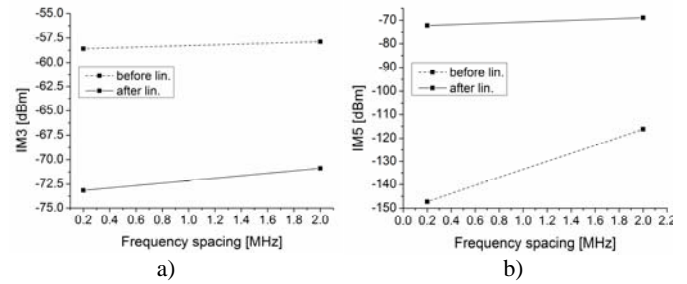


Fig 2. Intermodulation products before and after the linearization for  $P_{inRF} = -20$  dBm,  $P_{inLO} = -3$  dBm: a) IM3 i b) IM5

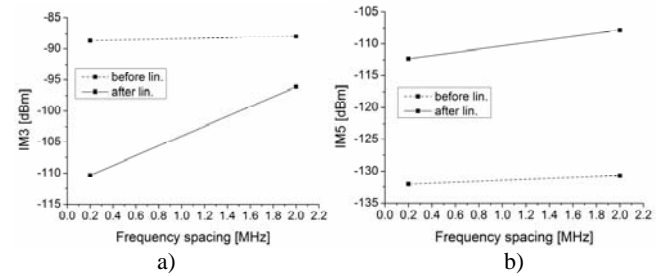


Fig 3. Intermodulation products before and after the linearization for  $P_{inRF} = -30$  dBm,  $P_{inLO} = -3$  dBm: a) IM3 i b) IM5

### IV. EFFECTS OF I/Q IMBALANCES

In ideal case, the signal from the local oscillators in the I and Q channels have equal amplitude and phase difference of  $-90$  degrees, as depicted in Figure 1. When the asymmetry occurs, the amplitudes and phases of the LO signals in the channels deviate from the values in the ideal case. In practice, I channel is defined as a reference (0 degrees phase, amplitude value 1).

The signal at the mixer input is in the form:

$$X_{RF}(t) = I(t)\cos(\omega_c t) - Q(t)\sin(\omega_c t) \quad (2)$$

where  $\omega_c$  is the carrier frequency.

Imbalance is characterized by amplitude ( $\alpha$ ) and phase shift ( $\theta$ ) of the signal from the local oscillator  $X_{LO}$  in Q branch as:

$$X_{LO} = -\alpha \sin(\omega_{LO} t + \theta) \quad (3)$$

Then, the IQ imbalanced signal at the mixer output can be written as follows:

$$\begin{aligned} I_{BB}(t) &= I(t) \\ Q_{BB}(t) &= \alpha[Q(t)\cos(\theta) - I(t)\sin(\theta)] \end{aligned} \quad (4)$$

In 3D figures, 4 and 5, the output power of the fundamental signal for both, input power levels and signal spacing, in terms of amplitudes and phases misalignment of the I and Q components is presented. Figures clearly indicate that output power levels stay almost unchanged with the increase of the parameters  $\alpha$  and  $\theta$  for the considered signal separation and input signal levels.

Figures 6 to 9, represent the IM3 and IM5 products after the linearization when IQ imbalances are considered. For low level of IQ imbalances ( $\alpha < 5\%$ ,  $\theta < 3^\circ$ ) the IM3 products after the linearization retain almost unaltered in case of 0.2 MHz signal spacing. When signal spacing is 2 MHz, the IM3 products are less susceptible to the amplitude and phase changing, especially for lower considered power. In the cases of greater IQ imbalances, values of the IM3 products after the linearization are approaching the levels of the IM3 products before the linearization. As far as the IM5 products are concerned they slightly increase with the rise of the IQ imbalances, but they still stay below the linearized IM3 products considered under the same imbalance conditions.

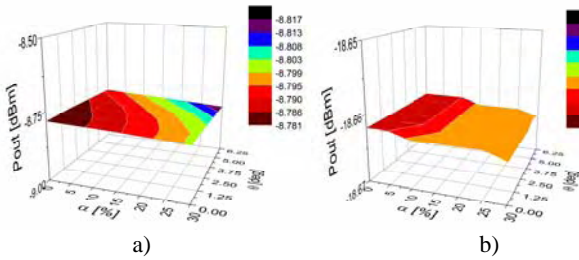


Fig. 4. Output power of the fundamental signal for signal spacing 0.2 MHz in terms of I/Q imbalances : a)  $\text{Pin}_{RF} = -20$  dBm,  $\text{Pin}_{LO} = -3$  dBm; b)  $\text{Pin}_{RF} = -30$  dBm,  $\text{Pin}_{LO} = -3$  dBm

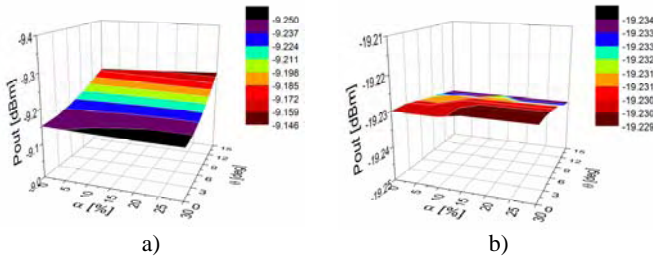


Fig. 5. Output power of the fundamental signal for signal spacing 2 MHz in terms of I/Q imbalances: a)  $\text{Pin}_{RF} = -20$  dBm,  $\text{Pin}_{LO} = -3$  dBm; b)  $\text{Pin}_{RF} = -30$  dBm,  $\text{Pin}_{LO} = -3$  dBm

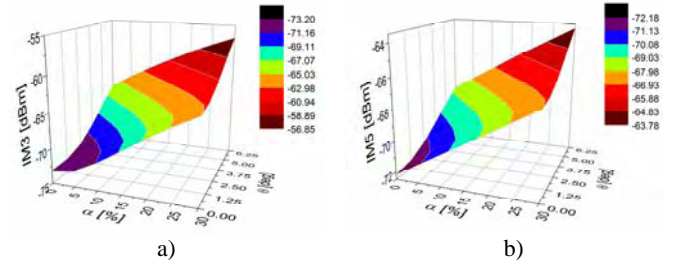


Fig. 6. Intermodulation products of the direct converted mixer for signal spacing 0.2 MHz,  $\text{Pin}_{RF} = -20$  dBm,  $\text{Pin}_{LO} = -3$  dBm: a) IM3; b) IM5

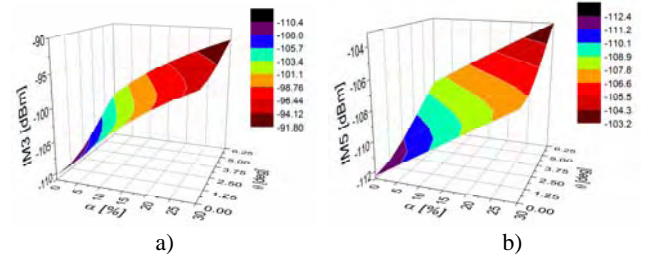


Fig. 7. Intermodulation products of the direct converted mixer for signal spacing 0.2 MHz,  $\text{Pin}_{RF} = -30$  dBm,  $\text{Pin}_{LO} = -3$  dBm: a) IM3; b) IM5

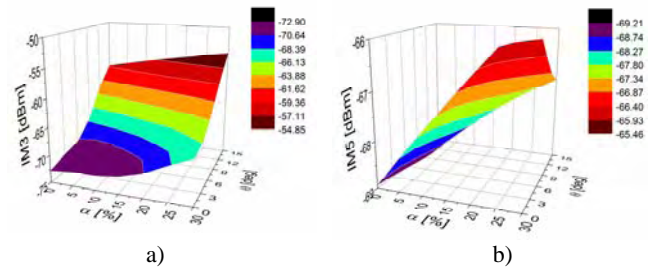


Fig. 8. Intermodulation products of the direct converted mixer for signal spacing 2 MHz,  $\text{Pin}_{RF} = -20$  dBm,  $\text{Pin}_{LO} = -3$  dBm: a) IM3; b) IM5

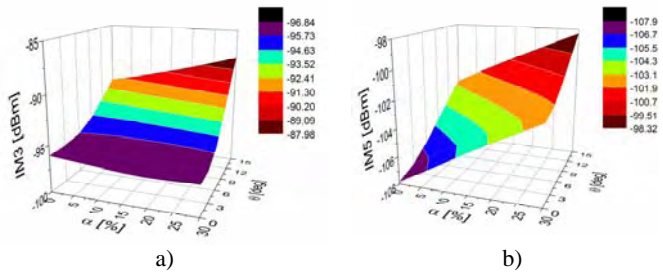


Fig. 9. Intermodulation products of the direct converted mixer for signal spacing 2 MHz,  $\text{Pin}_{RF} = -30$  dBm,  $\text{Pin}_{LO} = -3$  dBm: a) IM3; b) IM5

## V. CONCLUSION

This paper describes the linearization method that uses the modified baseband signals for the Gilbert mixer linearization in direct conversion receiver. The main role of this mixer is direct conversion of the input signal carrier frequency to the baseband. The test was performed for the QAM signal whose I and Q components are sinusoidal signals and the spectrum



contains two frequency components separated for 0.2 MHz and 2 MHz. The proposed linearization method utilizes the I and Q signals that are adequately processed in the digital domain at the receiver with the aim to form the signals for linearization. Linearization effects are examined for different input power levels and different frequency spacing between the signal spectral components. The signals for linearization are fed at the transistors' drain of the RF stage differential pair in the Gilbert cell. It should indicate that very good results are achieved in the reduction of the third-order mixer nonlinearity. The fifth-order intermodulation products are deteriorated, but they are still kept at the levels below the linearized IM3 products. Additionally, it is shown that the low-levels of IQ misalignment have almost negligible effect on the linearization results, especially in case of 2 MHz spacing between signals. Also, we analyse the grade in which the linearization effects deteriorate with the increasing imbalance.

#### ACKNOWLEDGEMENT

This work was supported by the Ministry of Education, Science and Technological Development of Republic of Serbia, the projects number TR-32052.

#### REFERENCES

- [1] Y. Kim, Y. Kim, and S. Lee, "Linearized mixer using predistortion technique", *IEEE Microw. Wireless Compon. Lett.*, vol. 12, no. 6, pp.204–205, 2002.
- [2] T. J. Ellis, "A modified feed-forward technique for mixer linearization", *IEEE MTT-S Int. Dig.*, pp. 1423–1426, 1998.
- [3] K-H Liang, C-H Lin, H-Y Chang, and Y-J Chan, "A New Linearization Technique for CMOS RF Mixer Using Third-Order transconductance Cancellation", *IEEE Microwave and Wireless Components Letters*, vol. 18, no. 5, pp.350-352, 2008.
- [4] S. Ock, Y. Yang and B. Kim, "New Linearization Method for Mixer", *Journal of the Korean Physical Society*, vol. 39, no. 1, pp. 1-3, 2001.
- [5] S. Lou, H. C. Luong, "A Linearization Technique for RF Receiver Front-End Using Second-Order-Intermodulation Injection", *IEEE Journal of Solid-state circuits*, vol. 43, no. 11, pp.2404-2412, 2008.
- [6] A. Đorić, N. Maleš-Ilić, A. Atanasković, B. Milovanović, "Mixer Linearization by Modified Baseband Signals", *Sinteza 2016, Conference Proceedings*, Belgrade, Serbia, 2016 (accepted for publication).
- [7] J. C. Pedro and J. Perez, "Accurate simulation of GaAs MESFET's intermodulation distortion using a new drain-source current model," *IEEE Trans. Microwave Theory Tech.*, vol. 42, pp. 25–33, January 1994.
- [8] J. P. Aikio and T. Rahkonen, "Detailed distortion analysis technique based on simulated large-signal voltage and current spectra", *IEEE MTT Trans Microwave Theory Tech.*, vol. 53, pp. 3057–3065, 2005.
- [10] A. Heiskanen, J. Aikio, and T. Rahkonen, "A 5-th order Volterra study of a 30W LDMOS power amplifier", *ISCAS'03-International Symposium on Circuits and Systems, Conference Proceedings*, Vol. 4, pp. 616–619, Bangkok, Thailand, 2003.

# **Telecommunication Systems and Technology**



# Novel Transport protocol for 5G Networks

Ivan Petrov<sup>1</sup> and Toni Janevski<sup>2</sup>

**Abstract** – In this paper we present novel transport protocol, designed to assure efficient data usage in future 5G mobile networks. Speeds up to 5Gbps are common in case of heavy mobile environment and high data rates up to 400Gbps are achieved in case of static or low speed terminal scenario.

This unique algorithm is designed to prevent congestion collapses of the backhaul network traffic making this protocol superior when compared with common high speed transport protocols.

**Keywords** – 5G mobile networks; congestion control; TCP; throughput.

## I. INTRODUCTION

From the very beginning until today TCP is evolving and the document roadmap describing its development is frequently updated [1]. One of the key triggers in the development process was its poor performance in high speed networks.

Congestion avoidance phase was studied and identified as generator of this unpleasant behavior. Numbers of TCP versions were created with primary aim to improve the protocol utilization in high speed environment and most of them were addressing and redesigning the congestion control mechanism. It is important to notice that during the design phase variety of constraints must be carefully addressed especially when creating protocol for usage in future 5G mobile networks [2-5].

Today we distinguish three main groups of high speed protocols: loss based [6-12], delay based [13-16] and loss based with bandwidth estimation [17-22], each of them improves the protocol performances.

High speed TCP proposals use direct and indirect approach of more aggressive network probing with aim to create congestion control mechanism that will be aggressive enough in case of underutilized network and will remain gentle in case of utilized resources.

There are two main reasons for poor performance of standard TCP protocols in high speed networks:

- Cwnd linear increase by one packet per round trip time in congestion avoidance phase is too slow and multiplicative decrease in case of loss is too drastic.
- Maintaining large average congestion windows at flow level require extremely small equilibrium loss probability and a stable flow dynamic design.

In this paper we present novel high speed transport

protocol (TCP Ohrid) capable to achieve speeds up to 400Gbps in future 5G networks.

In section II we discuss future 5G network design scenarios and we express the need for development of TCP Ohrid. TCP Ohrid protocol is defined in section III. In Section IV we present the simulation environment and discuss the results. Section V concludes the paper.

## II. FUTURE NETWORKS

Next Generation Networks (NGN) and Future Networks [2] are all-IP networks, meaning that all data, control and signaling will be carried through IP-based communication and based on Internet technologies form network protocol layer up to the application layer, with heterogeneous access technologies (fixed and mobile) on the lower layers of the protocol stack. TCP Ohrid meets 5G characteristics which are set prior to the standardization, which is expected to be finalized around 2020.

Most researchers are focused on development of 5G network that will provide improved future communication. Fundamental design, requests and drivers are defined and argued in [3]. 5G Network design scenarios and evaluations are conducted in [4].

Ideas for mobile cloud computing in order to offload most energy consuming tasks to nearby fixed servers and to develop mobile cloud computing are reality and are subject of future development [5].

As we know, small cells and their ultra dense deployment could become main feature of 5G networks indicating that the number of small cells will increase in unit area so the corresponding backhaul gateway traffic will increase exponentially if the conventional centralized control model is adopted as 5G backhaul network architecture. If this network design is chosen during the selection process we have to notice that massive backhaul traffic will create congestion which might result in a collapse of the backhaul network. Therefore, there is a need for development of ultra fast protocol similar to TCP Ohrid able to assure data transfer rates up to 400 Gbps or lower rates in case of heavy mobility of the terminals, making the protocol more efficient than the traditional one. This protocol is able to reduce and dynamically manage the backhaul congestion.

Hybrid backhaul networks using wireless and fiber access technologies can be used to increase the energy efficiency of the backhaul network.

We identify several requests that should be satisfied by future networks. 5G networks will have to employ energy harvesting mechanism at core base station network. For example, "harvest-then-transmit" mechanism or similar can be used in order to save energy [23].

Optimization at terminal application level with offload of most energy consuming tasks to nearby fixed servers presented

<sup>1</sup>Ivan Petrov is with Macedonian Telecom, Sales Excellence Unit, Kej 13 Noemvri, No.6. 1000, Skopje, Macedonia E-mail: ivan.petrov@telekom.mk

<sup>2</sup>Toni Janevski is Professor at the Faculty of Electrical Engineering and Information Technologies, Rugjer Boshkovik bb, 1000 Skopje, Macedonia. tonij@feit.ukim.edu.mk

as mobile cloud computing is additional value that can be added and employed at application layer in order to save terminal battery. Both centralized and decentralized approaches can be used in order to provide this network function.

5G Networks must represent an all IP mobile platform that will provide infrastructure as a service, platform as a service and software as a service. (Unified access of the platform will be provided regardless the medium access control, e.g., WiFi, 2G, 3G, LTE, Copper, Fiber, RF, etc.). The network must offer higher data transfer rates in static and mobile (nomadic) mode of operation.

### III. TCP OHRID PROTOCOL

TCP Ohrid protocol development is inspired by HSTCP protocol. We have noticed that HSTCP response function can be modified by preserving the main request to remain convex. It is possible to define different increase and decrease parameter functions by usage of different interpolation mechanisms. Equation (1) given below describes all loss based TCP protocols and in combination with (2) it defines HSTCP.

$$\begin{aligned} w &\leftarrow w + \frac{\alpha}{w} && \text{CA phase, when ACK received} \\ w &\leftarrow w - \beta w && \text{in case of drop} \\ w &\leftarrow w + \gamma && \text{Slow Start} \end{aligned} \quad (1)$$

$$\alpha(w) = \frac{w^2 2 \beta(w) p(w)}{2 - \beta(w)} \quad (2)$$

$\alpha$  is window increase parameter with default value of 1 Maximal Segment Size,  $\beta$  is decrease parameter with default value of 0.5 MSS,  $w$  denotes congestion window size,  $p$  is packet loss probability and  $\gamma$  is slow start parameter with default value of 1 MSS. When default values of  $\alpha$ ,  $\beta$  and  $\gamma$  are used equation (1) describes the standard TCP (Reno, Newreno) protocol versions.

TCP Ohrid protocol is defined with several response functions where linear interpolation is used for six known points. Usage of several response functions for different speed improves protocol friendliness at lower rates. The algorithm combines response functions in order to provide improved protocol friendliness at low and high rates. We know that fiber optic has error rate of 1 error in  $10^{10}$  bits meaning that the probability of bit error in fiber is  $10^{-10}$ . In our scenario we use packets with size of 1500 bytes that limits packet loss probability in fiber networks at  $P=12000 \times 10^{-10}=1.2 \times 10^{-6}$ . TCP Ohrid protocol preserves usage of switch point defined with  $w=38$  packets,  $p=0.0015$  and  $\beta$  parameter set at 0.5 in order to represent fair protocol that will be friendly with competing TCP. If we want to create a protocol that will achieve Gbps throughput we have to modify the basic definition (1) in order to get faster window increase. However, this implies that we cannot use existing congestion mechanism. We decided to use modified  $\beta$  parameter equation in predefined range of known constraint values. TCP Ohrid protocol is defined with several response functions in dependence of the sending rate. The main idea is to use one primary switch point and five end/switch point's instead one switch and one endpoint at log-log scale. It is defined with points that correspond at speed of 5Gbps, 10Gbps, 40Gbps, 100Gbps and 400Gbps throughput that is supported by the transport medium. Knowing that packet loss probability of fiber networks is in rang of  $10^{-7}$

(when 1500 byte packets are used) we decided to observe real packet loss probability values for optical networks and future packet loss probability of 5G packet transmission networks. TCP Ohrid protocol is 5G protocol that can provide data rates higher than 5Gbps under heavy user mobility and 400 Gbps backhaul data rates in the core network. Today there are patented solutions for providing ultra high speed transmission with capacity of  $10^{15}$  bits/s. In 2005, 100Gbps Ethernet was tested in laboratory and in 2010 commercial active and passive equipment was available. IEEE Taskforce has completed standardization of local area network interfaces for 100Gbps ethernet. We can recall that IEEE P802.3bj supports BER better or equal to  $10^{-12}$  and 100 Gbps (over copper), IEEE P802.3bs supports BER better or equal to  $10^{-13}$  and MAC data rate of 400Gbps. IEEE 802.3ba supports BER better than or equal to  $10^{-12}$  and MAC data rates of 40Gbps and 100 Gbps. Considering all above stated we can decide which packet loss probability values and which data rates in the switch points can be used in future 5G network. Note that the suggested bit error probabilities are real if we use 1500 byte packets. TCP Ohrid is defined with maximal speed of 400Gbps,  $p=10^{-12}$  and  $\beta=0.1$  or 0.4 in dependence of the used decrease parameter variation, the primary switch point preserve its parameter values. Interpolation can be conducted with help of following defined points 5Gbps,  $p=10^{-7}$ ; 10Gbps,  $p=10^{-8}$ ; 40Gbps,  $p=10^{-9}$  and 100Gbps,  $p=10^{-10}$ . It is expected in the future networks BER to be over  $10^{-15}$  which represents the main reason why we have decided to use it as a limit. Line equations are defined among following points  $w-w_1$ ,  $w_1-w_2$ ,  $w_2-w_3$ ,  $w_3-w_4$ ,  $w_4-w_5$  and  $w-w_2$ ,  $w-w_3$ ,  $w-w_4$ ,  $w-w_5$  as presented at Fig. 1.

Five basic response functions denoted as TCPO1, TCPO2, TCPO3, TCPO4, TCPO5 and additional four TCPO6, TCPO7, TCPO8 and TCPO9 are defined.

All of the response functions are defined with line equation with different slope. Several algorithms can be defined. The protocol will start with standard TCP Reno/NewReno mechanism, afterwards it will use TCPO1 and in case of lossless environment after passing  $w_1$  value it will use TCPO2 function, if loss is detected than TCPO6 is used as response function (even if the calculated window value after loss is below the switch point value after passing  $w=38$ , TCPO6 will be used because it will provide improved window growth). After reaching  $w_2$  value TCPO3 will be used and if loss is detected TCP07 will be used. When  $w_3$  value is reached TCPO4 is used and if loss is detected TCPO8 will be used or TCP05 will be used after  $w_4$  value is reached. When loss is detected it switches to use of TCPO9 and after passing  $w_5$  value current response function will be preserved. In case of three consecutive timeouts the protocol counters are restarted and it restarts the initial usage of the response functions according the actual congestion window value (cwnd). It is interesting to notice that this protocol modifies the initial behavior of Reno/Newreno protocol used, during slow mode of operation. In order to calculate the window value after the loss we use:

$$w_{new} = \frac{w_{max} + w_{min}}{2} \quad (3)$$

where  $w_{max}$  denotes highest window value achieved prior the loss occurs and  $w_{min}$  is the decreased window value after the loss that corresponds with the second line of equation (1). Obtained value places the cwnd in the middle between highest achieved value before the loss and the calculated value that should be obtained after the loss.



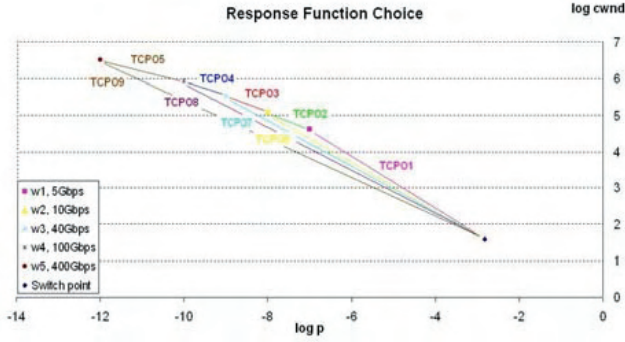


Fig. 1. Response function choice of TCPOhrid

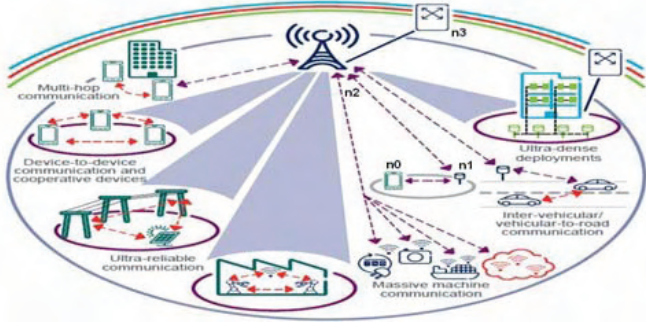


Fig. 2. Simulation scenario. High speed backhaul link is used between MBS and core network gateway. We define wireless medium between n0 and n1, n1 is wirelessly connected with macro base station n2 and optical link n2-n3 is defined between the macro base station n2 and the network gateway n3.

Similar window calculation algorithm is used in BIC protocol [8]. We find usage of (3) more interesting because it impacts  $\beta$  calculation. Till now we have defined the protocol response function variation and its loss window calculation. Packet loss probability  $p$  can be calculated for a given value of  $cwnd$  and  $\alpha$  parameter is defined with equation (2). The  $\beta$  parameter is:

$$\beta_{calculated}(w) = (0.1 - 0.5) \frac{\log(w) - \log(38)}{\log(w_5) - \log(38)} + 0.5 \quad (4)$$

Where  $w$  represents the present value of the congestion window of interest and  $w_5$  is the congestion window size that defines the fifth switch point. Calculated value of  $\beta$  should be calibrated having in mind that the window calculation after the loss is modified, when equation (3) is used to calculate the window value after loss in slow mode of operation. If we substitute all known values in (3) we obtain

$$w_{new} = (1 - \frac{\beta}{2}) w_{max} \quad (5)$$

Equation that indicates new  $\beta$  value is:

$$\beta_{new} = \frac{\beta_{calculated}}{2} \quad (6)$$

This equation is used only in case of time out, afterwards the protocol proceeds with standard  $\beta$  calculation. Since we have defined all important protocol parameters we can implement them in Network Simulator 2 (ns2). In the simulation  $\beta$  calculation is done by using:

$$\beta_{new} = 0.5 - \beta_{calculated} \quad (7)$$

With this equation we achieve increased concavity of  $\beta$  function that provides low  $cwnd$  decrement for low values and high  $cwnd$  decrement when timeout occurs for large window values. Usage of (7) provides improved protocol performance

for large  $cwnd$  values when congestion occurs because the window will be decreased for 40% of the previous value. Lowest  $\beta_{new}$  value is bounded by 0.1 and the highest with 0.4. If the result of equation (7) is lower than 0.1 than  $\beta_{new}$  value is rounded at 0.1.

#### IV. SIMULATION SCENARIO

We study simple simulation scenario presented at Fig. 2 implemented with help of Network Simulator 2 (ns2). We simulate only one mobile user (n0) communication with micro base station (n1) and macro base station (n2) through backhaul high speed link n2-n3 connected with the adequate network gateway (n3).

Parameter of interest is congestion window size variation; we know that it is directly related with throughput values. Packet size is set at 1500 bytes. Simulation lasts 230 sec. The both cases, when maximum bound of window size is 20 packets and 80000 packets result in large trace files. Buffers are Drop Tail; buffer size is set at 100% of the product of bottleneck capacity by the largest RTT divided by packet size. Link speed can vary and high speed link has RTT value of 100 ms. We have conducted simulations for 5, 10, 40 and 100 Gbps and adequate buffer size limitations. Since the max window size is predefined, similar results are obtained regardless the speed and buffer size constraints.

Obtained results are presented at Fig.3 and Fig. 4.

On Fig. 3 we have presented  $cwnd$  change when link capacity is set at 5, 10, 40 and 100 Gbps. Simulation last 230 sec and the window size is set to ns2 default value of 20 packets. This figure justifies previously stated design parameters of Ohrid protocol. It can be noticed that Reno protocol has smallest window growth, directly impacting the throughput that can be achieved when this protocol is used. Ohrid has shown larger window growth compared to Reno but smaller than HSTCP. This proves that Ohrid protocol is more friendly with existing protocols than HSTCP and still it sustains its high speed performances. This protocol behavior assures that can be used to buffer the negative impact of wireless medium especially in case of heavy mobile users. At On Fig. 4 we have obtained results when the window parameter is set at 80000. At the beginning of the simulation Reno protocol achieves lowest  $cwnd$  values while HSTCP  $cwnd$  change is better than the one achieved by Ohrid protocol mainly because of the rapid growth of the HSTCP response function. The Ohrid protocol has better performance which justifies the novel protocol design. The protocol uses more efficient set of response functions in case when the window size has larger value than  $w_1$  packets. It is obvious that the Ohrid protocol provides higher data rates than HSTCP protocol. Fig. 4 assures us that the protocol is robust in case of increased terminal mobility,  $cwnd$  growth will be lower, hence making the protocol capable to cope wireless medium impact at the given throughput when the user is mobile and away from the base station. When the terminal is static or near the base station and the number of active users of the network is low than higher data rates can be achieved. Therefore, this protocol is appropriate for use in cell communication and for backhaul data transfer.

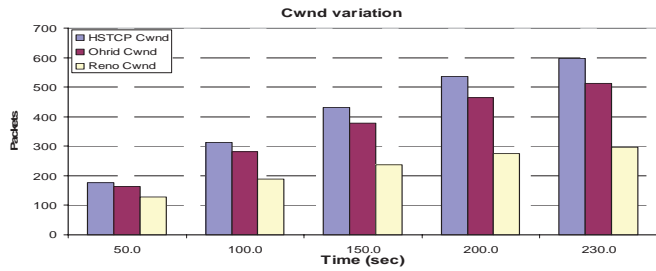


Fig. 3. cwnd change when 5, 10, 40 and 100 Gbps links are used with adequate buffer sizes.

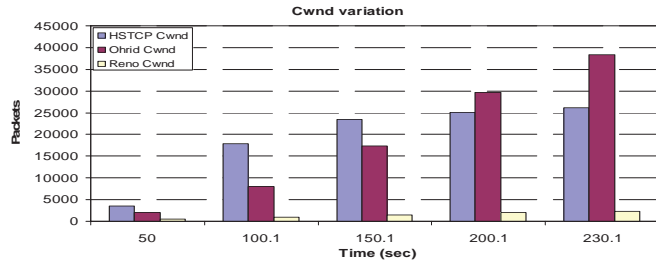


Fig. 4. cwnd change when 5, 10, 40 and 100 Gbps links are used with adequate buffer sizes (max window bound is set at 80000)

## V. CONCLUSION

In this paper we have presented and evaluated the TCP Ohrid protocol that has shown improved cwnd behavior. Increased friendliness and higher data rate performance of TCP Ohrid protocol are justified. The protocol is defined to be used in the future 5G data networks regardless of the engaged MAC Layer. It is defined with several response functions for speeds of 5, 10, 40, 100 and 400 Gbps. The used  $\beta$  parameter expressed with equation (7) guarantees improved behavior especially in high congested networks. Equation (3) is used to improve Reno/Newreno protocol performance till reaching cwnd value of 38 packets which provides faster increase when loss occurs during slow mode of operation. All used parameters can be fine tuned in order to properly design the protocol. Development of high speed protocols that will sustain 5G network capacity is essential hence user can benefit the network capacity. TCP Ohrid is a good candidate since it can support high transmission speeds and provide efficient usage of network capacity up to speeds of 400 Gbps.

## REFERENCES

- [1] M.Duke, R.Braden,W.Eddy, E.Blanton, A.Zimmermann, "A Roadmap for Transmission Control Protocol (TCP)". Internet Engineering task Force (IETF),RFC7414, February 2015.
- [2] T. Janevski, "NGN Architectures, Protocols, and Services", John Wiley and Sons, UK, 2014.
- [3] J. Rodriguez, "Fundamentals of 5G Mobile Networks", John Wiley and Sons, UK, 2015.
- [4] X. Ge, H. Cheng, M. Guizani, T. Han, "5G wireless backhaul networks: challenges and research advances", Network IEEE, Nov.-Dec. 2014
- [5] S. Barbarossa, S. Sardellitti, P. Di. Lorenzo, " Communicationg While Computing", IEEE Signal Processing magazine, Oct. 2014
- [6] T. Kelly, "Scalable TCP: improving performance in highspeed wide area networks," Computer Communications Review, vol. 32, no. 2, April 2003.

- [7] S. Floyd, "HighSpeed TCP for large congestion windows," RFC 3649, 2003.
- [8] L. Xu, K. Harfoush, and I. Rhee, "Binary increase congestion control for fast, long distance networks," in Proc. IEEE INFOCOM, vol. 4, March 2004, pp. 2514–2524.
- [9] D.Leith, "H-TCP:TCP congestion control for high bandwidth-delay product paths," IETF Internet Draft, <http://tools.ietf.org/html/draftleith-tcp-htcp-06>, 2008.
- [10] I. Rhee and L. Xu, "CUBIC: a new TCP-friendly high-speed TCP variant," SIGOPS Operating Systems Review, vol. 42, no. 5, pp. 64–74, July 2008.
- [11] G. Marfia, C. Palazzi, G. Pau, M. Gerla, M. Sanadidi, and M. Roccetti, "TCP Libra: Exploring RTT-Fairness for TCP," UCLA Computer Science Department, Tech. Rep. UCLA-CSD TR-050037, 2005.
- [12] C. Caini and R. Firrincieli, "TCP Hybla: a TCP enhancement for heterogeneous networks," International J. Satellite Communications and Networking, vol. 22, pp. 547–566, 2004.
- [13] A. Baiocchi, A. P. Castellani, and F. Vacirca, "YeAH-TCP: yet another highspeed TCP," in Proc. PFLDnet, ISI, Marina Del Rey (Los Angeles), California, February 2007.
- [14] R. King, R. Baraniuk, and R. Riedi, "TCP-Africa: an adaptive and fair rapid increase rule for scalable TCP," in Proc. IEEE INFOCOM, vol. 3, March 2005, pp. 1838–1848.
- [15] K. Tan, J. Song, Q. Zhang, and M. Sridharan, "A compound TCP approach for high-speed and long distance networks," July 2005.
- [16] S. Liu, T. Basar, and R. Srikant, "TCP-Illinois: A loss and delay-based congestion control algorithm for high-speed networks," in Proc. First International Conference on Performance Evaluation Methodologies and Tools (VALUETOOLS), 2006.
- [17] D. X. Wei, C. Jin, S. H. Low, and S. Hegde, "FAST TCP: motivation, architecture, algorithms, performance," IEEE/ACM Trans. Netw., vol. 14, no. 6, pp. 1246–1259, 2006.
- [18] J. Sing and B. Soh, "TCP New Vegas: Improving the Performance of TCP Vegas Over High Latency Links," in Proc. 4th IEEE International Symposium on Network Computing and Applications (IEEE NCA05), 2005, pp. 73–80.
- [19] K. Yamada, R. Wang, M.Y. Sanadidi, and M. Gerla " TCP Westwood with Agile Probing:Dealing with Dynamic, Large, Leaky Pipes" IEEE Communication Society, 2004.
- [20] D. Kliazovich, F. Granelli, and D. Miorandi, "Logarithmic window increase for TCP Westwood+ for improvement in high speed, long distance networks," Computer Networks, vol. 52, no. 12, pp. 2395–2410, August 2008.
- [21] H. Shimonishi, T. Hama, and T. Murase "TCP-Adaptive Reno for Improving Efficiency-Friendliness Tradeoffs of TCP Congestion Control Algorithm", in Proc. 4th International Workshop on Protocols for Fast Long Distance Networks, 2006
- [22] K. Kaneko, T. Fujikawa, Z. Su, and J. Katto, "TCP-Fusion: a hybrid congestion control algorithm for high-speed networks," in Proc. PFLDnet, ISI, Marina Del Rey (Los Angeles), California, February 2007.
- [23] D. Gunduz, K. Stamatiou, N. Michelusi, and M. Zorzi, "Designingintelligent energy harvesting communication systems," IEEE Commun.Magazine, vol. 52, no. 1, pp. 210 - 216, Jan. 2014

# OP of SC Diversity for OFDM-based WLAN System Operating in Microcell and Picocell Environment

Aleksandra Panajotović<sup>1</sup>, Dragan Drača<sup>1</sup> and Nikola Sekulović<sup>2</sup>

**Abstract** – This paper explores, through the simulation of system outage probability, selection combining (SC) diversity algorithms for orthogonal frequency division multiplexing (OFDM)-based wireless local area network (WLAN) system complaint to IEEE 802.11n standard. Presented results show advantage and disadvantage of an applied SC diversity algorithm depending on a channel profile.

**Keywords** – Diversity System, OFDM, Outage probability, Simulation, Space-time-frequency-selective fading.

## I. INTRODUCTION

In last two decades, remarkable demand for reliable, high-speed wireless communications is evident. The new techniques which can answer to this demand are multiantenna and orthogonal frequency division multiplexing (OFDM) techniques. Multiantenna techniques can be divided into two categories: diversity techniques and spatial-multiplexing techniques. The first ones are based on receiving same information in the multiple antennas aiming to convert an unstable time-varying wireless fading channel into a stable additive white Gaussian noise (AWGN)-like channel [1]. That provides upgrading transmission reliability of wireless system without increasing transmission power. The traditional diversity techniques are maximal ratio combining (MRC), equal gain combining (EGC), and selection combining (SC). Among them, the last one has the least implementation complexity since it processes signal only from one of diversity antennas, which is selectively chosen, and consequently manifests the worst performance. Normally, a SC receiver selects the antenna with the highest signal-to-noise ratio (SNR) [2]. In OFDM-based systems, a selection process can be based on an entire OFDM symbol (wideband SC) or on an individual subcarrier (narrowband SC) [3].

In order to study influence of system and channel parameters on performance of wireless system, two groups of performance measures, first-order and second-order, should be considered. First-order performance measures are outage probability (OP), error probability, system capacity, average output signal, etc. Second-order performance measures, average level crossing rate (LCR) and average fade duration (AFD), are important for an adaptive system in which first-order measures do not provide enough information for the

overall system design and configuration [4]. The OP has been traditionally the most common used performance measure and, moreover, it is necessary for evaluation of the second-order performance measure, i.e. AFD.

To describe fading behaviour in a channel, few statistical models can be found in open technical literature. The most frequently used models are Rice, Rayleigh, Nakagami- $m$ , and Weibull. The performance of SC receivers operating in previous mentioned channels are studied through simulation and numerical analysis in [5]-[8]. In [3] and [9] the performance of OFDM-based systems with a space diversity applied on transmit and receive side are studied, while the performance of SC receiver operating in stochastic Multiple-Input Multiple-Output (MIMO) channel suitable to IEEE 802.11n is investigated in [10]. This paper analyzes the performance of selection diversity applied in Single-Input Multiple Output (SIMO)-OFDM system complaint to IEEE 802.11n standard.

## II. OUTAGE PROBABILITY

The OP is defined as probability that instantaneous error probability exceeds a specified value or, equivalently, probability that output SNR,  $\mu$ , falls below a certain specified threshold,  $\mu_{th}$ . Mathematically speaking, the OP is defined as [2]

$$P_{out} = \int_0^{\mu_{th}} p_{\mu}(\mu) d\mu, \quad (1)$$

where  $p_{\mu}(\mu)$  represents an probability density function of output SNR,  $SNR_{out}$ , which form depends on wireless environment.

In OFDM-based system, an antenna selection process can be done on the individual subcarrier level or on the entire OFDM symbol level. The OFDM receiver with  $L$  diversity antennas is shown in Fig. 1.

### A. Wideband SC

The wideband SC system selects the antenna with the best OFDM symbol. Namely, selected antenna  $j$  is one which satisfying following condition

$$j = \arg \max_{i=\{1, \dots, L\}} \left\{ \sum_{k=1}^{N_c} |h_{i,k}|^2 [f_k]^2 \right\}, \quad (2)$$

<sup>1</sup>Aleksandra Panajotović and Dragan Drača are with the Faculty of Electronic Engineering at University of Niš, 14 Aleksandra Medvedeva, Niš 18000, Serbia, E-mail: {aleksandra.panajotovic and dragan.draca}@elfak.ni.ac.rs.

<sup>2</sup>Nikola Sekulović is with the College of Applied Technical Sciences, 20 Aleksandra Medvedeva, Niš 18000, Serbia, E-mail: nikola.sekulovic@vtsnis.edu.rs.



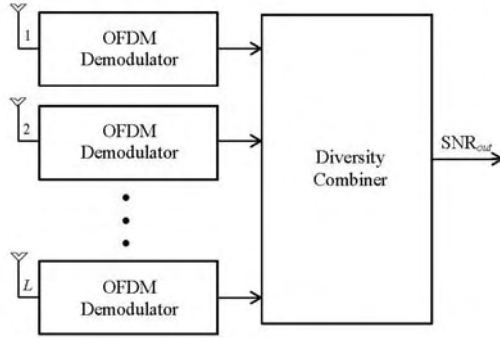


Fig. 1. OFDM receiver with selection diversity

The parameter  $f_k$  is  $k$ th subcarrier,  $N_c$  is number of subcarriers in the OFDM symbol, and  $h_{i,1}[f_k]$  is the channel gain between transmit and  $i$ th receive antenna on  $k$ th subcarrier.

### B. Narrowband SC

The narrowband SC system makes a decision on each individual subcarrier. In this system, the selected antenna satisfies

$$j = \arg \max_{i=1, \dots, L} \{ |h_{i,1}[f_k]|^2 \}. \quad (3)$$

Note that different antennas are selected on different subcarriers.

Numerical analysis is not only easy and time-unconsumed, but accurate way to estimate the system performance. Otherwise, a simulation is used with scientific modeling of system to gain insight into its functioning and also presents good way to confirm the accuracy of already obtained numerical results.

In this paper, the simple stochastic MIMO channel model [11] is applied for a link simulation of SIMO channel. The main strength of this MIMO model is that it relies on small set of parameters to fully characterize communication scenario and all those parameters are extracted from measurement results.

The algorithm presented in Fig. 2 describes the process of modeling of  $L$ -branch OFDM receiver operating in environments compliant to IEEE 802.11n and the simulation of its outage probability.

## III. SIMULATION RESULTS

This section presents the simulation results for OP of SIMO-OFDM system with two receive antennas in which one of proposed selection diversity algorithms is applied. It is supposed that system operates in two different WLAN environments: microcell (Fig. 3) and picocell (Fig. 4). The

term microcell refers to indoor-to-outdoor (B channel profile) propagation, characteristic for environment consists of small offices where distance between mobile station (MS) and base station (BS) is from 31 to 36 m with BS located outside. The picocell is term for indoor-to-indoor (E channel profile) propagation. The E channel profile represents very large open area as airports or modern open offices [11].

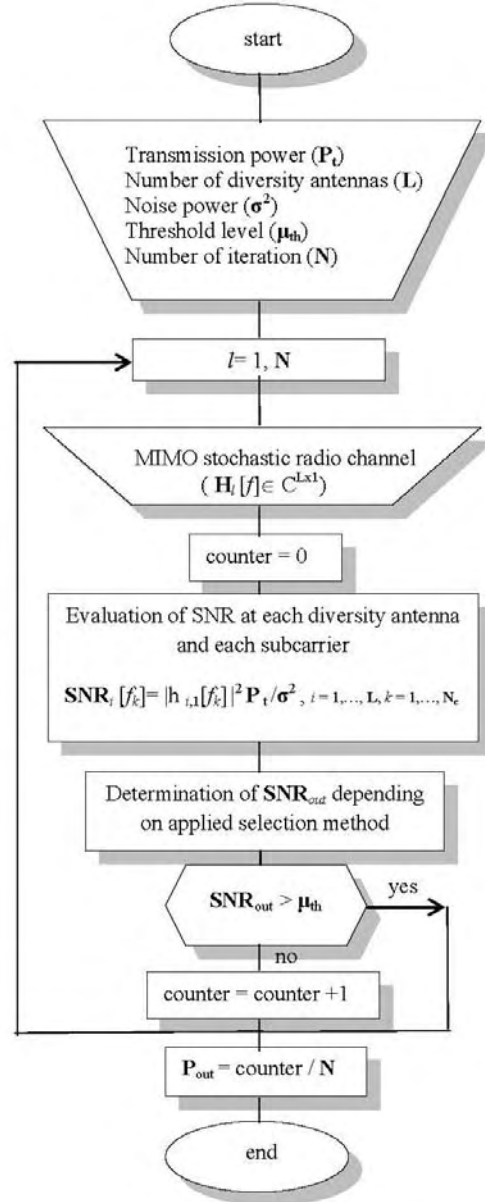


Fig. 2. Modeling and simulation of SIMO-OFDM system with SC diversity

Figures 3 and 4 depict OP versus SNR threshold. In both channel profiles, using of narrowband SC diversity system is more recommended, what is confirmed through Figs. 3 and 4. Namely, the wideband SC selects the best OFDM symbol,

even though some subcarriers on that symbol may be non-optimal compared to the corresponding subcarrier on the other antenna. Greater number of subcarriers on OFDM symbol, more reasons for using the narrowband SC. However, the narrowband SC should realize  $L \cdot N_c$  measurements, instead of  $L$  (wideband SC), what makes it so impractical for realization. Let compare Figs. 3 and 4. Since the fact that E channel profile is more frequently-selective than B, it is expected that performance improvement provided by using narrowband SC is more significant for that channel profile. In addition, outage performance of system with narrowband SC is almost independent on WLAN environment.

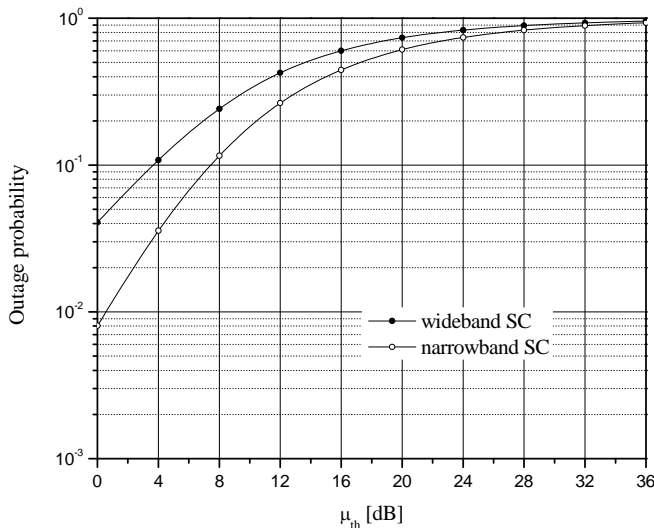


Fig. 3. Outage probability of SIMO-OFDM system operating in microcell

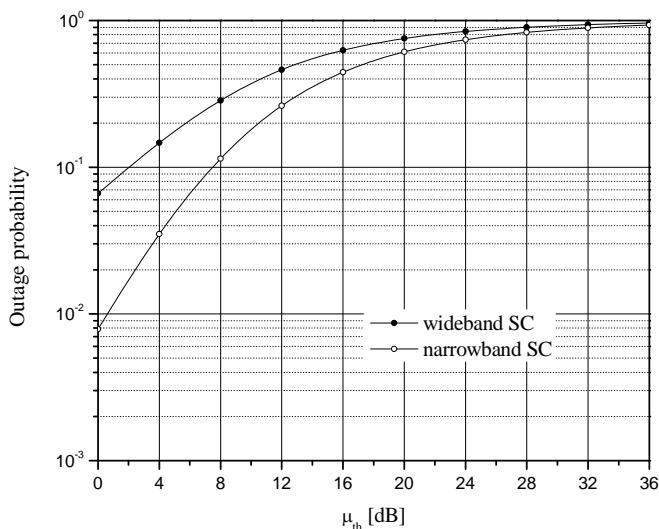


Fig. 4. Outage probability of SIMO-OFDM system operating in picocell

## IV. CONCLUSION

This paper investigates the influence of different selection diversity algorithms on outage performance of OFDM-based WLAN system compliant to IEEE 802.11n standard. It is evident that trading off between system performance and implementation complexity have to be done. This research can be extended to other performance metrics or other diversity techniques.

## ACKNOWLEDGEMENT

This work has been funded by the Serbian Ministry of Education and Science under the projects TR-32052, III-44006 and TR-33035.

## REFERENCES

- [1] Y. S. Cho, J. Kim, W. Y. Yang, and C. -G. Kang, *MIMO-OFDM Wireless Communications with MATLAB*, Wiley, 2010.
- [2] M. K. Simon, M. -S. Alouini, *Digital Communications over Fading Channels*, Wiley, 2005.
- [3] D. Lee, G. J. Saulnier, Z. Ye, and M. J. Medley, "Antenna Diversity for an OFDM System in a Fading Channel", MILCOM'99, Conference Proceedings, pp. 1104-1109, Atlantic City, USA.
- [4] L. Yang, M. -S. Alouini, *Average Outage Duration of Wireless Communication System*, ch. 8, *Wireless Communication Systems and Networks*, Springer, 2004.
- [5] D. A. Zogas, G. K. Karagiannidis, "Infinite-series Representations Associated with Bivariate Rician Distribution and Their Applications", IEEE Trans. Commun., vol. 53, no. 11, pp. 1790-1794, 2005.
- [6] C. C. Tand, N. C. Beaulieu, "Infinite-series Representations of the Bivariate Rayleigh and Nakagami-m Distribution", IEEE Trans. Commun., vol. 45, no. 10, pp. 1159-1161, 1997.
- [7] M. K. Simon, M. -S. Alouini, "A Simple Single Integral Representation of the Bivariate Rayleigh Distribution", IEEE Commun. Lett., vol. 2, no.5, pp. 128-130, 1998.
- [8] N. C. Sagias, G. K. Karagiannidis, D. A. Zogas, P. T. Mathiopoulos, and G. S. Tombras, "Performance Analysis of Dual Selection Diversity in Correlated Weibull Fading Channel", IEEE Trans. Commun., vol. 52, no. 7, pp. 1063-1067, 2004.
- [9] P. Corral, J. L. Corral, and V. Almenar, "Diversity Techniques for OFDM Based WLAN Systems: A Comparison Between Hard, Soft Quantified and Soft No Quantified Decision", PIMRC2002, Conference Proceedings, Lisbon, Portugal, 2002.
- [10] A. Panajotović, Dragan Drača, "Outage Probability of SC Receiver Operating in Both Microcell and Picocell Environment", SSSS2016, Conference Proceedings, pp. 76-78, Niš, Serbia, 2016.
- [11] J. P. Kermoal, L. Schumacher, K. I. Pederson, P. E. Mogensen, and F. Frederiksen, "A Stochastic MIMO Radio Channel Model with Experimental Validation", IEEE J. Select. Areas Commun., vol. 20, no. 6, pp. 1211-1225, 2002.



This Page Intentionally Left Blank

# Chromatic Polynomial Heuristics for Connectivity Prediction in Wireless Sensor Networks

Vladislav Vasilev<sup>1</sup>

**Abstract** – Connectivity related problems in wireless sensor networks are an active area of research. A common requirement in connectivity management is the utilization of a robust and scalable unsupervised learning method. In this paper an isomorphism of a specially defined hypergraph is used to derive a clustering method. The proposed method builds chordal graph and estimates their chromatic polynomial.

**Keywords** – Hypergraph, Isomorphism, Wireless sensor Networks.

## I. INTRODUCTION

One of the simplest and most basic of tasks in a Wireless Sensor Network (WSN) is the procedure of neighbour discovery. It allows a given node to choose upon all of the nodes that it could possibly reach, through its neighbours. Reachability, in turn, is directly related to routing, which is a NP-problem in general. Since, the network's time dynamic state is not changing completely at random, then a node discovery procedure should use an unsupervised clustering method. Currently, there are a lot of available clustering tools in the literature [1]. The most commonly used clustering algorithms are k-Nearest Neighbour and K-Means. Both can scale well with appropriate approximations, but unfortunately, they perform badly in higher dimensions as the distance between data points becomes uninformative. One way to go about this problem is to try to use appropriate projections in lower dimensions and then, use the modes of the projections to make inferences about the clusters [2]. An alternative to this is to represent the data set as vertexes of a hyper-graph and then, to try to find such a hyper-edge that maximizes a criteria for a subset of vertexes to belong to a cluster. An example of this is given in [3]. In general, clustering problems are well defined in the scope of hyper-graph theory [4]. A common problem of most clustering algorithms, which utilize some kind of distance such as an Euclidean norm, is that the number of clusters and the termination conditions are hard to be determined. Hence, it is possible to have empty clusters or clusters that have a number of clusters within themselves. Moreover, these algorithms are very sensitive to the scaling of the data that will artificially make certain clusters more or less separable.

<sup>1</sup>Vladislav Vasilev is a full-time PhD student at the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: v.g.vasilev@tu-sofia.bg Sofia, Bulgaria.

Furthermore, in the case of connection evaluation the clusters are also time dependent. Hence, the system can't spend much resources on cluster estimation, if they are going to loose their relevance in a short period of time. Hence, the model needs some time dependent predicting power of the network structure.

The most common approach to these issues is to use biologically inspired methods like Genetic Algorithms (GA) [5], Evolutionary Algorithms (EA) [6], Ant Colony (AC) [7] or others like in [8]. Probably the most used algorithm is Particle Swarm Optimisation (PSO) [9]–[12] and its many variants. In this paper we take a different approach. We define the problem in a hyper-graph theory framework similar to that in [13], which then after a number of appropriate transformations are made will result in a heuristic cost function for the cluster of close neighbours with respect to a single node.

The rest of the paper is organized as follows: Section II derives the heuristic cost function; Section III explains the data set obtained from the simulation and the respective results; Section IV presents some concluding remarks.

## II. INDEX INVARIANCE

Let us define the hyper graph  $G = (\{V\}, \{E\})$  with vertex set  $\{V\}$  and hyper-edge set  $\{E\}$ , where the degrees of the hyper-edges  $d \leq |\{V\}|$  are arbitrary. An example of such a graph is given in Fig. 1. For each vertex  $V_i$  there exists a node in the network. We assume that two sensors are neighbors and constitute a hyper-edge  $E^2$  of degree two, if they are pairwise in each other's range of coverage.

When the vertex  $V_i$  makes a first time connection to a set of vertexes  $\{V\}$ , it makes no a priori difference between the labels of the vertexes as long as they are different. For that reason, let us define the notion of index invariance.

A hyper-graph  $S_\sigma = (\{V\}, \{E\})$  with vertex set  $\{V_\sigma\} = \{V_i, V_j, V_k \dots\}$  and hyper-edge set  $\{E\}$  is index invariant, if the incidence matrix  $[I]$  of the graph is upper triangular and the elements of the vector of unique indexes  $\sigma = \{i, j, k \dots\}$  can be rearranged in an arbitrary way  $\{\sigma\}$ , without this causing change in the graph  $S_\sigma$  and this is given in Eq. (1).

$$S_\sigma, I_S^\Delta, \vec{\sigma} \equiv S_\sigma, I_S^\Delta, \{\sigma\} \quad (1)$$

An example of such hyper-graph  $S_\sigma$  is given in Fig 1.

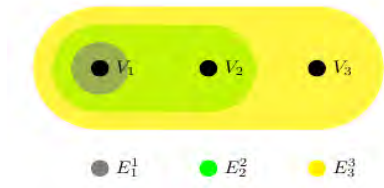


Fig. 1. An example of an index-invariant hypergraph.

Theorem 2.1: An Index Invariant Graph (IIG)  $S_\sigma$  and its dual graph  $S_\sigma^D$  are isomorphic. An alternative statement is that the transpose matrix  $[I]^T$  of the matrix  $[I]$  induces such a graph  $S_\sigma^D$ , that is isomorphic to  $S_\sigma$ .

We prove the theorem by induction on the number of vertices  $n$ . The theorem is true for  $n = 1, 2, 3$ , hence, assume it is true for all graphs with no more than  $n$  vertices. For the induction step, we add one vertex  $V_{n+1}$  and one hyper-edge  $E_{n+1}$  to  $I(n)$  which results in  $I(n+1)$  in Eq. (2).

$$I_S^\Delta(n+1) = \left\| \begin{array}{ccccc|c} & E_1 & \dots & E_n & E_{n+1} \\ V_1 & & & & 1 \\ \vdots & & I_S^\Delta(n) & & 1 \\ V_n & & & & 1 \\ \hline V_{n+1} & 0 & 0 & 0 & 1 \end{array} \right\| \quad (2)$$

We transpose Eq. (2). Since  $I(n)$  is upper triangular, then its transposed matrix  $I^T(n)$  is lower triangular in Eq. (3).

$$[I_S^\Delta(n+1)]^T = \left\| \begin{array}{ccccc|c} & E_1 & \dots & E_{n-1} & E_n & E_{n+1} \\ V_1 & & & & 0 \\ \vdots & & [I_S^\Delta(n)]^T & & 0 \\ V_{n-1} & & & & 0 \\ V_n & & & & 0 \\ \hline V_{n+1} & 1 & 1 & 1 & 1 & 1 \end{array} \right\| \quad (3)$$

We assumed by the induction hypothesis, that all matrices  $I^T(n)$  with a size smaller or equal to  $n$  can be transformed into upper triangular  $I^T(n) = I(n)$ . Next, we note that any reordering in the columns and rows of  $I^T(n+1)$  in Eq. (3), does not change row  $V_{n+1}$  and column  $E_{n+1}$ , neither the structure of the graph. We move the rows of  $I^T(n+1)$  with one position downwards, so that  $V_1$  goes in place of  $V_2$ ,

$V_2$  goes in place of  $V_3$  and likewise for each row  $V_{i-1} = V_i$  other than  $V_{n+1}$ . Next,  $V_n$  goes in place of  $V_1$  which gives Eq. (4).

$$[I_S^\Delta(n+1)]^T = \left\| \begin{array}{ccccc|c} & E_1 & \dots & E_{n-1} & E_n & E_{n+1} \\ V_n & & & & 0 \\ V_1 & & & & 0 \\ \vdots & & I_S^\Delta(n) & & 0 \\ V_{n-1} & & & & 0 \\ \hline V_{n+1} & 1 & 1 & 1 & 1 & 1 \end{array} \right\| \quad (4)$$

In Eq. (4) we swap row  $V_n$  and  $V_{n+1}$ , which yields Eq. (5).

$$[I_S^\Delta(n+1)]^T = \left\| \begin{array}{ccccc|c} & E_1 & \dots & E_{n-1} & E_n & E_{n+1} \\ V_{n+1} & 1 & 1 & 1 & 1 & 1 \\ V_1 & 1 & 1 & 1 & 1 & 0 \\ \vdots & & & & & 0 \\ V_{n-1} & & & & & 0 \\ \hline V_n & 0 & 0 & 0 & 1 & 0 \end{array} \right\| \quad (5)$$

We move columns  $E_i$  with one rightwards, so that  $E_1$  goes in place of  $E_2$ ,  $E_2$  goes in place of  $E_3$  and every column other than  $E_{n+1}$  with one rightwards  $E_{i-1} = E_i$ . Next,  $E_n$  goes in place of  $E_1$ . Finally, we swap columns  $E_n$  and  $E_{n+1}$  which gives Eq. (6).

$$[I_S^\Delta(n+1)]^T = \left\| \begin{array}{ccccc|c} & E_{n+1} & E_1 & \dots & E_{n-1} & E_n \\ V_{n+1} & 1 & 1 & 1 & 1 & 1 \\ V_1 & 0 & 1 & 1 & 1 & 1 \\ \vdots & & & & & 1 \\ V_{n-1} & & & & & 1 \\ \hline V_n & 0 & 0 & 0 & 0 & 1 \end{array} \right\| \quad (6)$$

Eq. (6) proves the theorem, since the matrix is upper triangular.

We can substitute in Eq. (1) to obtain Eq. (7)

$$S_\sigma, I_S^\Delta, \bar{\sigma} \equiv S_\sigma, I_S^\Delta, \{\sigma\} \equiv S_\sigma^D, [I_S^\Delta]^T, \{\sigma\} \quad (7)$$

A direct consequence of Eq. (7) is that for each IIG there exists exactly one dual, while the two of them are isomorphic to one another. Because the ordering of the labels of the vertices of  $S_\sigma$  and the ordering of the labels of the hyper-edges of the dual  $S_\sigma^D$  are the same and are of no consequence, then the two graphs are indistinguishable from one another  $S_\sigma = S_\sigma^D$ .

Let us examine the IIG  $S_\sigma(n)$  and one of its sub hypergraph  $S_\sigma(k), \leq n$ , as it is done in Fig. 2. There exist  $|S_\sigma(k)| = \prod_{i=0}^k (n-i)$  number of sub graphs, that are indistinguishable by Eq. (7). Therefore, we consider them as one and the same graph leading to Eq. (8).

$$\left\{ S_\sigma(k)_i \right\}_{i=1}^{\prod_{i=0}^k (n-i)} \equiv \left\{ S_\sigma(k) \right\} \quad (8)$$

We note at this point that the number of sub IIGs of size  $k$  coincides with the chromatic polynomial  $C_p = |S_\sigma(k)|$  of the complete graph  $G_k$  with  $k$  vertexes and  $n$  colors. Let us observe Fig. 2 further in order to prove that the IIG is a special case of a complete graph.

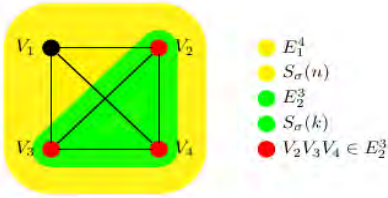


Fig. 2. An example of a sub IIG.

A vertex that is not in  $S_\sigma(k)$  must have a label that is different than the labels of each vertex in  $S_\sigma(k)$ . That is true for all  $S_\sigma(k)$  and all  $k$ . If we build a 2-uniform graph  $K_n$  having the same vertex set as  $S_\sigma(k)$  where two vertexes of  $K_n$  are neighbors if they have different labels, then we can see that there is an edge of degree 2 between each two vertexes of  $K_n$ . This means that  $K_n$  is a complete graph and an IIG is a special case of a complete graph.

Therefore, Eq. (8) allows us to evaluate the index invariance with the chromatic polynomial  $C_p(S_2, \lambda)$  [14] of the 2-section graph  $(S)_2$ .

$$C_p((S)_2, \lambda) = \prod_{i \in \{\sigma((S)_2)\}} [\lambda - N(V_i)]^{|\sigma_i|}, \quad (9)$$

where  $N(V_i)$  is the number of neighbors of the simplicial vertex  $V_i$  in the simplicial decomposition  $\sigma$  and  $|\sigma|$  is the total number of vertexes having that number of neighbors in the decomposition.

Because the metric grows fast, we can normalize each polynomial with its minimum value and further take the logarithm. Let  $n$  be the number of vertexes in the tree, then we write Eq.(10).

$$M_{etric} = \log \left[ \frac{\prod_{i \in \{\sigma((B_g)_2)\}} [\lambda - N(V_i)]^{|\sigma_i|}}{\prod_{i=0}^{n-1} (\lambda - i)} \right] \quad (10)$$

Another consequence of Eq. (8) is that if two IIGs have a common vertex, then they form an entire IIG, meaning that two independent IIGs do not intersect. The inverse is also true. If two IIGs do not intersect, then they are independent.

### III. SIMULATION AND RESULTS

For the purposes of this study we view the network from the perspective of a single initial node. In Fig. 3 it is shown as follows: all nodes are marked as crosses and their coverage range is given in blue; the initial node and its coverage range in green; with red - all neighbouring relations; in black - cover trees rooted at the initial node.

On each iteration, the coverage of each sensor changes, which changes the structure of the graph, which in turn changes the access capabilities and resources. On the other hand, in each time interval requests enter the system. At each time new nodes can become connected to the initial node and in the same time some nodes might get disconnected as the range of each node changes. Naturally, if a node has no resources to accept packets or the path between the nodes gets disconnected packets will be dropped.

After we accumulate data on the network structure as it changes in time, we can then compute each structure's corresponding heuristics. On Fig. 4 we plot the metric in an increasing order and the packet loss rate that is measured for that structure.

The results show, that with the increase of the heuristics, the probability of packet loss also rises. For instance, two hyper-edges with indexes  $e1 = 12$  and  $e1 = 19$ , have heuristics  $H1 = 0.56$  and  $H2 = 1.58$  and their respective packet success probability is  $P1 = 0.43$  and  $P2 = 0.33$ .

It must be also noted, that the metric is time slice specific, while the change in packet losses can be subject to differential processes.

In conclusion, the sensors should negotiate a maximum acceptable value for the heuristics and participate only in such hyper-edges that are below that value in order to minimise their packet loss in the network.

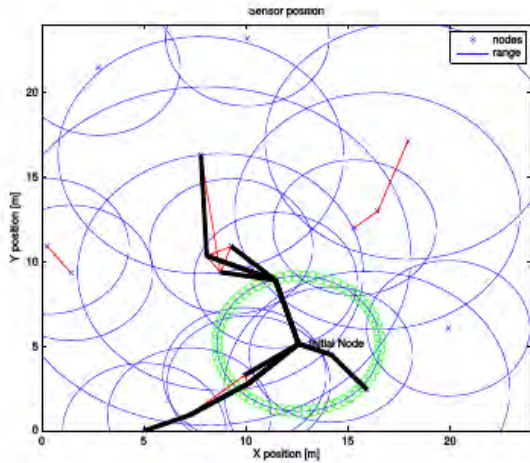


Fig. 3. A State of a WSN.

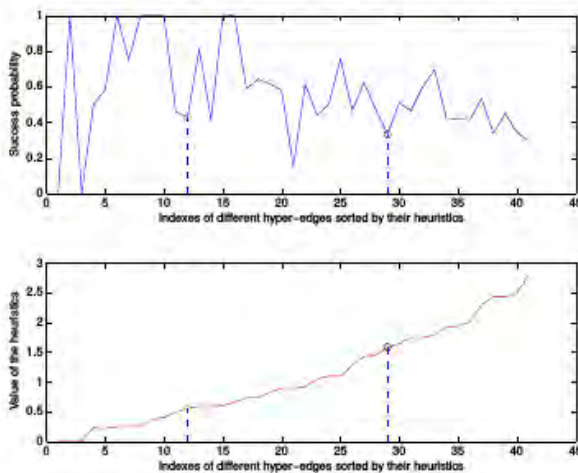


Fig. 4. An example of a sub IIG.

#### IV. CONCLUSION AND FUTURE WORK

In this paper a method for evaluating the connectivity in a wireless sensor network is proposed. In the future we need to compare our results with different algorithms that tackle the same problem and also we need to increase the simulation complexity with respect to the sensor behaviour and environment.

#### REFERENCES

- [1] C. M. Bishop, Pattern recognition and machine learning. Springer, 2006.
- [2] D. Littau and D. Boley, "Clustering very large data sets with principal direction divisive partitioning." in Grouping

- Multidimensional Data, J. Kogan, C. K. Nicholas, and M. Teboulle, Eds. Springer, 2006, pp. 99–126.
- [3] D. Zhou, J. Huang, and B. Scholkopf, "Learning with hypergraphs: Clustering, classification, and embedding," in Advances in Neural Information Processing Systems (NIPS) 19. MIT Press, 2006, p. 2006.
- [4] J. Yang and J. Leskovec, "Overlapping community detection at scale: A nonnegative matrix factorization approach," in Proceedings of the Sixth ACM International Conference on Web Search and Data Mining, ser. WSDM 13. New York, NY, USA: ACM, 2013, pp. 587–596.
- [5] S. Edelkamp and S. Schrödl, Heuristic Search - Theory and Applications. Academic Press, 2012.
- [6] V. Vasilev, V. Poulkov, and G. Iliev, "Uplink power control based on evolutionary algorithm with associative memory," in Proc. 6th Int. Conf. Systems and Networks Communications, 2011, pp. 9–14.
- [7] M. Frey, F. Grose, and M. Gunes, "Energy-aware ant routing in wireless multi-hop networks," in Communications (ICC), 2014 IEEE International Conference on. IEEE, 2014, pp. 190–196.
- [8] S. Kasabova, M. Gechev, V. Vasilev, A. Mihovska, and V. Poulkov, "On modeling the psychology of wireless node interactions in the context of internet of things," in Wireless Personal Communications, 2015.
- [9] X. Cao, H. Zhang, J. Shi, and G. Cui, "Cluster heads election analysis for multi-hop wireless sensor networks based on weighted graph and particle swarm optimization," in Natural Computation, 2008. ICNC'08. Fourth International Conference on, vol. 7. IEEE, 2008, pp. 599–603.
- [10] B. Li, N. Wu, H. Wang, D. Shi, W. Yuan, and J. Kuang, "Particle swarm optimization-based particle filter for cooperative localization in wireless networks," in Wireless Communications & Signal Processing (WCSP), 2013 International Conference on. IEEE, 2013, pp. 1–6.
- [11] C. Cao, Q. Ni, and X. Yin, "Comparison of particle swarm optimization algorithms in wireless sensor network node localization," in Systems, Man and Cybernetics (SMC), 2014 IEEE International Conference on. IEEE, 2014, pp. 252–257.
- [12] J. Yang, H. Zhang, Y. Ling, C. Pan, and W. Sun, "Task allocation for wireless sensor network using modified binary particle swarm optimization," Sensors Journal, IEEE, vol. 14, no. 3, pp. 882–892, 2014.
- [13] V. Vasilev, G. Iliev, V. Poulkov, and A. Mihovska, "Optimization of wireless node discovery in an iot network," in Proceedings of 2015 IEEE GLOBECOM, Workshop on Networking and Collaboration Issues or the Internet of Everything. San Diego, CA, USA: IEEE, December 6-10 2015.
- [14] V. Voloshin, Introduction to Graph and Hypergraph Theory. Nova Science Publishers, Inc., 2009.



# Hybrid Environment Service Orchestration for Fog Computing

Stojan Kitanov<sup>1</sup> and Toni Janevski<sup>2</sup>

**Abstract** – Emerging applications in the context of the Internet of Everything (IoE) introduce real-time and low latency requirements that raise new challenges on the services being provided to the users. These demands can only be partially fulfilled by existing Cloud Computing solutions. A new paradigm called Fog Computing, or briefly Fog has emerged to meet these requirements. It extends Cloud Computing to the network edge, allowing a wider geographic coverage, low latency, load balancing and flexible mobility support, and showing as a promising solution within the IoE. However, several challenges arise when it comes to the provision of Fog services, including the need for supporting the orchestration within large scale and dynamic environments of complex and on-demand services, which should be capable of adapting at runtime in order to ensure resilience and trustworthiness. This paper evaluates a model of Hybrid Environment Service Orchestrator (HESO) for resilient and trustworthy Fog Computing services in terms of network latency. The results demonstrate that HESO model is well positioned for real time big data analytics, 5G network and IoT.

**Keywords** – Cloud, Cloud Computing, Fog, Fog Computing, Mobile Cloud, Mobile Cloud Computing.

## I. INTRODUCTION

The future Internet of Services (IoSs) will become the linkage between extremely complex networked organizations (e.g. telecoms, transportation, financial, health and government services, commodities, etc.), that will provide the basic ICT infrastructure that supports the business processes and the activities of the whole society in general [1]. Frequently, these processes and activities will be supported by orchestrated cloud services, where a number of services work together to achieve a business objective. However, future Internet will exacerbate the need for improved QoS/QoE, supported by services that are orchestrated on-demand and are capable of adapt at runtime, depending on the contextual conditions, to allow reduced latency, high mobility, high scalability, and real time execution. The emerging wave of Internet of Things (IoTs) would require seamless mobility support and geo-distribution in addition to location awareness and low latency. These demands can be only partially fulfilled by existing cloud computing solutions [2].

A new paradigm called Fog Computing, or briefly Fog has emerged to meet these requirements [3]. Fog extends cloud computing and services to the edge of the network. It provides data, computing, storage, and application services to end-users

that can be hosted at the network edge or even end devices such as set-top-boxes or access points. The main features of Fog are its proximity to end-users, its dense geographical distribution, and its support for mobility. Fog will combine the study of mobile communications, micro-clouds, distributed systems, and consumer big data. It is a scenario where a huge number of heterogeneous (wireless and sometimes autonomous) ubiquitous and decentralized devices communicate and potentially cooperate among them and with the network to perform storage and processing tasks without the intervention of third parties [4]. These tasks support basic network functions or new services and applications that run in a sand-boxed environment. Users leasing part of their devices to host these services get incentives for doing so.

By deploying reserved compute and communication resources at the edge, Fog computing absorbs the intensive mobile traffic using local fast-rate connections and relieves the long back and forth data transmissions among cloud and mobile devices [5], [6]. This significantly reduces the service latency and improves the service quality perceived by mobile users and, more importantly, greatly saves both the bandwidth cost and energy consumptions inside the Internet backbone. Fog computing represents a scalable, sustainable and efficient solution to enable the convergence of cloud-based Internet and the mobile computing. Therefore Fog paradigm is well positioned for real time big data analytics, 5G network, and IoT.

The move from Cloud to Fog computing brings out several key challenges, including the need for supporting the on-demand orchestration and runtime adaptation of resilient and trustworthy Fog Services. This is essential for the success of the future Internet of Everything (IoE), which a clear evolution of the IoT [7].

This paper provides a model of Hybrid Environment Service Orchestrator (HESO) for resilient and trustworthy Fog Computing services. It is organized as follows. Section II provides an overview of Fog Computing. Section III proposes the HESO model for Fog Computing. Section IV evaluates the HESO model in terms of Round Trip Time (RTT) latency. Finally, Section V concludes the paper and provides future work research directions.

## II. OVERVIEW OF FOG COMPUTING

An overview of three layered Fog Computing architecture is given in Fig. 1. The intermediate Fog layer consists of geo-distributed intelligent Fog Computing servers which are deployed at the edge of networks, e.g., parks, bus terminals, shopping centers, etc. Each Fog server is a highly virtualized computing system and is equipped with the on-board large

<sup>1</sup>Stojan Kitanov is with the Faculty of Information Systems Multimedia and Animation at University of Information Science and Technology "St. Paul the Apostle," of Ohrid, Macedonia. E-mail: stojan.kitanov@uist.edu.mk.

<sup>2</sup>Toni Janevski is with the Faculty of Electrical Engineering and Information Technologies at Ss Cyril and Methodius University of Skopje, Macedonia. E-mail: tonij@feit.ukim.edu.mk

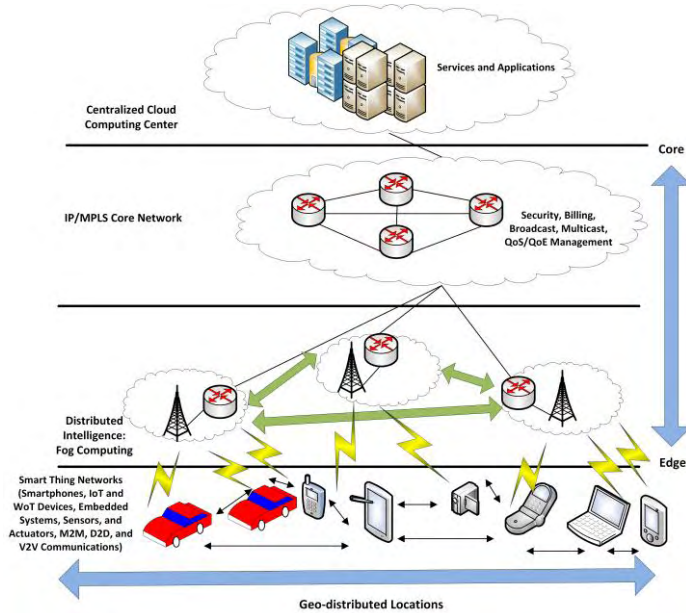


Fig. 1. Fog Computing Architecture

TABLE I  
A COMPARISON BETWEEN FOG AND CLOUD

	<b>Fog Computing</b>	<b>Cloud Computing</b>
Target Type	Mobile Users	General Internet Users
Service Type	Limited localized information services related to specific deployment locations	Global information collected from worldwide
Hardware	Limited storage, compute power and wireless interface	Ample and scalable storage space and compute power
Distance to Users	In the physical proximity and communicate through single-hop wireless connection	Faraway from users and communicate through IP networks
Working Environment	Outdoor (streets, parklands, etc.) or indoor (restaurants, shopping malls, etc.)	Warehouse-size building with air conditioning systems
Deployment	Centralized or distributed in regional areas by local business (local telecommunication vendor, shopping mall retailer, etc.)	Centralized and maintained by Amazon, Google, etc.

volume data storage, compute and wireless communication facility [5].

The role of Fog servers is to bridge the smart mobile device things and the cloud. Each smart thing device is attached to one of Fog servers that could be interconnected and each of them is linked to the cloud [6].

The geo-distributed intelligent Fog servers directly communicate with the mobile users through single-hop wireless connections using the off-the-shelf wireless interfaces, such as, LTE, WiFi, Bluetooth, etc. They can independently provide pre-defined service applications to mobile users without assistances from cloud or Internet. In addition, the Fog servers are connected to the cloud in order to leverage the rich functions and application tools of the cloud.

The existence of Fog will be enabled by the emerging trends on technology usage patterns on the one side, and the advances on enabling technologies on the other side. A comparison between Fog Computing and Cloud Computing is given in [5], and it is summarized in Table I.

The cloud in 5G network and beyond will be diffused among the client devices often with mobility too, i.e. the cloud will become fog. More and more virtual network functionality will be executed in a fog computing environment, and that will provide ubiquitous service to the users. This will enable new services paradigms such as Anything as a Service (AaaS) where devices, terminals, machines, and also smart things and robots will become innovative tools that will produce and use applications, services and data.

### III. SERVICE ORCHESTRATION WITH FOG

The move from cloud to fog brings out several key challenges. This includes the need for supporting the on-demand orchestration and runtime adaptation of resilient and trustworthy fog Services, which is essential for the success of the future IoE, a clear evolution of the IoT.

Traditional service orchestration approaches that have been applied to Cloud services are not adequate to the forthcoming large-scale and dynamic Fog Services, since they cannot effectively cope with reduced latency, high mobility, high scalability, and real time execution. Therefore a new Hybrid Environment Services Orchestrator (HESO) is needed, that will be capable of ensuring the resilience and trustworthiness of open, large scale, dynamic services on the Fog. The Orchestrator will be responsible for the composition of Service Elements available in the Fog environment (e.g. sensing, connectivity, storage, processing, platform services, and software services) into more complex Fog Services (e.g. traffic crowd sensing and trip planning services) to be offered to the users in the Fog environment.

The execution of the Fog Services may involve multiple different components and entities spread in a wide area, increasing the complexity in terms of the decision making process in what regards the resource allocation to achieve acceptable QoS/QoE levels. To coordinate the execution of the Fog services, Orchestration mechanisms need to synchronize and combine the operation of the different service elements in order to meet the specifications of the composed Fog services, including low latency, scalability and resilience.

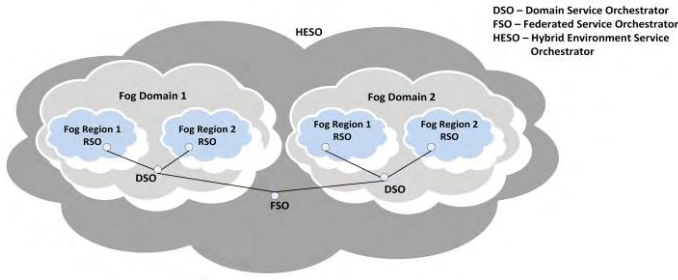


Fig. 2. Hybrid Environment Service Orchestrator Model for Fog Computing

The architectural levels of Fog orchestrated services and mechanisms are given in Fig. 2. The HESO in Fog should operate in a loosely coupled mode, resulting in a solution with several levels: Regional Service Orchestrator (RSO), Domain Service Orchestrator (DSO) and Federated Service Orchestrator (FSO).

The RSOs are located at the edges of the Fog environment and they enable semi-autonomous operation of the different Fog Regions. This allows the distribution of the load which provides scalability and a much higher proximity to the end users. Therefore lower latencies can be achieved.

The DSOs are responsible for the Fog domains and supervises the RSOs below. This level will support federation mechanisms to enable intra-domain cooperation between different regions within one domain.

The FSO allows a fruitful interaction between different Fog domains. It is responsible for the management between different Fog domains and, similarly to the DSOs, it should be properly adapted to operate in a federate Cloud environment. The FSOs will support federation mechanisms to enable cooperation among different Fog Domains (e.g. belonging to different entities or under the administration of different authorities) and the creation of a Multi-Domain Fog Environment able to support service ubiquity.

#### IV. EVALUATION OF THE HESO MODEL

The evaluation of the HESO will be explored in terms of the Round Trip Time (RTT) latency. RTT latency is the time it takes for a single data transaction to occur, meaning the time it takes for the packet of data to travel to and from the source to the destination, and back to the source [10]. In the real networks, latency is measured by performing ping tests, that uses ICMP packets. The total size of each ICMP packet is 74 bytes with the headers.

Let the mobile user be located in the Fog Region 1, which is controlled by the Fog Domain 1. And let Fog Domain 1 through a Federated Service Orchestrator be connected with Fog Domain 2. Fog Region 1 may correspond to an LTE/LTE-Advanced Cloud Radio Access Network (CRAN). Fog Domain 1 may correspond to a Cloud Computing Centre in the same region with the CRAN Network, and Fog Domain 2 may correspond to a Cloud Computing Centre in a different region with the CRAN network.

Let us assume the mobile user wants to upload and download some file for example a map, movie or similar, or

wants to process some data. The RTT latency time required to perform this transaction is equal to:

$$RTT = \frac{T}{R_{UL}} + \frac{T}{R_{DL}} + i_1 t_1 + i_2 t_2 \quad (1)$$

Here,  $T$  is the packet size travel from the source to the destination, and back to the source,  $R_{UL}$  and  $R_{DL}$  are the corresponding uplink and downlink data rates of LTE/LTE-A Network. The values of the uplink and downlink data rates [11] vary from the distance  $d$  between the mobile user and the CRAN network, and are summarized in Table II. The binary information coefficients  $i_1$  and  $i_2$  point the location of the data for which the end user is interested. Table III summarizes the possible location of the file, as well as the possible values of  $i_1$  and  $i_2$ . Finally  $t_1$  and  $t_2$  represent the time for the data file to be received by the Fog Region 1 (LTE Network) from Fog Domain 1 (Cloud Computing Center) or from the Fog Domain 2 (Cloud Computing Center), respectively.

TABLE II  
DATA RATES OF AN LTE-A/LTE NETWORK

Maximum Downlink Data Rate in Mbps	Maximum Uplink Data Rate in Mbps	Distance between the End user and the LTE-A eNodeB in meters
3000	1500	$d < 500$
300	100	$500 \leq d < 1000$
150	50	$1000 \leq d < 1500$
50	25	$1500 \leq d$

TABLE III  
POSSIBLE VALUES FOR THE BINARY INFORMATION COEFFICIENTS

Location of Data File	$i_1$	$i_2$
Fog Region 1	0	0
Fog Domain 1	1	0
Fog Domain 2	0	1

The simulation results are given in Fig. 3. The size of a data file is 74 MB, and the values of  $t_1$  and  $t_2$  correspond to 50 msec and 100 msec, respectively. Here RTT1 represent the network latency when data file requested by the mobile user is located in CRAN network (Fog Region 1). RTT2 represent the network latency when data file requested by the mobile user is located in the cloud computing center (Fog Domain 1), which is in the same region with the CRAN network. RTT3 represent the network latency when data file requested by the mobile user is located in the cloud computing center (Fog Domain 2) which is a different region with CRAN network.

Fig. 3a shows that the RTT latency increases as the mobile user moves away from the eNodeB of the CRAN network. Fig. 3b compares the RTT latencies depending whether the data file requested by the mobile user is located in Fog Region



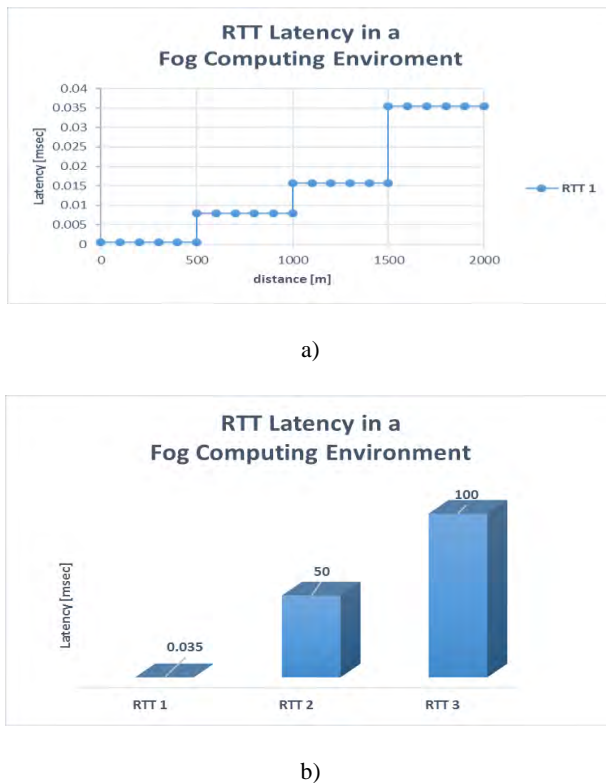


Fig. 3. RTT Latency in the Fog with a HESO Model

1, Fog Domain 1, or Fog Domain 2. The lowest network latency is obtained if the data file is located in the CRAN (Fog Region 1). The highest latency is obtained if the file requested by the user is located in the Fog Domain 2, i.e. in a cloud computing center which is in a different region with the CRAN network.

The results demonstrate that latency is significantly reduced from the order of milliseconds to the order of microseconds, which is very important for the delay sensitive applications.

## V. CONCLUSION AND FUTURE WORK

The purpose of Fog computing is to place a handful of computing, storage and communication resources in the proximity of mobile users, and therefore to serve mobile users with the local short-distance high-rate connections. The move from cloud to fog brings out several key challenges, including the need for supporting the on-demand orchestration and runtime adaptation of resilient and trustworthy fog services, which is essential for the success of the future IoE, a clear evolution of the IoT. This could be solved by the proposed Hybrid Environment Service Orchestrator for resilient and trustworthy Fog Computing services.

The results demonstrate that the latency can be significantly reduced to the order of microseconds by using the HESO model for the Fog Computing, which is very important for the delay sensitive applications. Therefore HESO model with Fog Computing is well positioned for real time big data analytics, 5G network and IoT.

Fog will act as a nervous system of the digital society, economy, and everyday people's life. Fog paradigm is well positioned for real time big data analytics, 5G network, and IoT. The cloud in 5G networks and beyond will be diffused among the client devices often with mobility too, i.e. the cloud will become fog. More and more virtual network functionality will be executed in a fog computing environment, and it will provide ubiquitous service to the users. This will enable new AaaS service paradigms, where devices, terminals, machines, and also smart things and robots will become innovative tools that will produce and use applications, services and data.

However there are also some aspects that should be addressed in order the Fog approach to be successful. This includes defining hybrid and heterogeneous environments, interaction and integration between the execution managements of each domain, and integration between managements inside one domain. In future we plan to work on solving some of these challenges.

## REFERENCES

- [1] Horizon project: A New Horizon to the Internet, 2015 <http://www.gta.ufrj.br/horizon/>
- [2] S. Zhang, S. Zhang, X. Chen, X. Huo, "Cloud Computing Research and Development Trend," in Proceedings of the 2010 Second International Conference on Future Networks (ICFN '10). IEEE Computer Society, Washington, DC, USA, pp. 93-97, 2010.
- [3] F. Bonomi, R. Milito, J. Zhu, S. Addepalli, "Fog Computing and its Role in the Internet of Things," in Proceedings of the first edition of the MCC workshop on Mobile Cloud Computing (MCC 2012), ISBN: 978-1-4503-1519-7, doi:10.1145/2342509.2342513, ACM, New York, NY, USA, pp. 13-16, 2012.
- [4] L. M. Vaquero, L. Roderio-Merino, "Finding your Way in the Fog: Towards a Comprehensive Definition of Fog Computing," ACM SIGCOMM Computer Communication Review, Vol. 44, No. 5, doi:10.1145/2677046.2677052, pp. 27-32, 2014.
- [5] H. T. Luan, L. Gao, Z. Li, L. X. Y. Sun, "Fog Computing: Focusing on Mobile Users at the Edge," arXiv:1502.01815[cs.NI], 2015.
- [6] I. Stojmenovic, S. Wen, "The Fog Computing Paradigm: Scenarios and Security Issues," in Proceedings of the Federated Conference on Computer Science and Information Systems (FedCSIS), ACSIS, Vol. 2, No. 5, doi:10.15439/2014F503, pp. 1-8, 2014.
- [7] B. Brech, J. Jamison, L. Shao, G. Wightwick, "The Interconnecting of Everything. IBM Redbook," 2013 <http://www.redbooks.ibm.com/redpapers/pdfs/redp4975.pdf>
- [8] G. Brown, "Converging Telecom and IT in the LTE RAN," White Paper at Heavy Reading on behalf of Samsung, 2013.
- [9] J. S. Stolfo, B. M. Salem, D. A. Keromytis, "Fog Computing: Mitigating Insider Data Theft Attacks in the Cloud," in Proceeding of IEEE Symposium on Security and Privacy Workshops (SPW), ISBN: 978-0-7695-4740-4, doi: 10.1109/SPW.2012.19. IEEE Computer Society, Washington, DC, USA, pp. 125-128, 2012.
- [10] "Latency Consideration in LTE Implications to Security Gateway," Stoke Inc. White Paper, Literature No. 130-0029-001, 2014.
- [11] A. Roessler, J. Schlien, S. Merkel, M. Kottkamp, "LTE-Advanced (3GPP Rel. 12) Technology Introduction," Rhode and Schwarz White Paper 6.2014 – 1MA252 2E, 2014.

# Transformation of the existing operator to the service provider for business customers

Zlata Radak<sup>1</sup> and Nataša Gospić<sup>2</sup>

**Abstract** – In this paper we describe all the benefits of implementing managed solution, while as an extension of this solution for connecting a large number of locations with a central location through the L3VPN, is presented an innovative solution with virtualized central location placed in the operator's Data Center, in order to achieve additional performance and security improvements.

**Keywords** – *Managed Services, Customer, Connection*

## I. INTRODUCTION

The development of modern information and communication technology represents a major challenge for the operator nowadays, because modern operator must have additional services in their portfolio, beside the basic service for connecting to the Internet, and to be prepare to offer customer complete solution for their business. An increasing number of companies that are highly specialized, like foreign representative offices and large retail chains have specific requirements concerning networking and connecting with their central location and pulling data from a central server, as well as in terms of Internet access which must comply with certain rules and restrictions. In order to meet the company's requirements regarding telecommunications, the company must have highly specialized person and sometimes the whole team that deals with the establishment of services, monitoring and solving any problems. Because branches are often geographically dislocated throughout Serbia, and sometimes abroad, customers increasingly need to shift the responsibility for telecommunications to the service provider - operator.

In order to deal with managed solutions and provide customer services in addition to Internet connectivity, a complete solution, the operator must pass every year a serious revision of the allocated managed certificates. Once awarded the certificate gives the operator the ability to deal with managed solutions with certified processes and professional team that provides support for all processes. However thus obtained certificate is valid for only a year, and each year the committee re-checked all available resources, processes and evaluates. This process monitor all systems for tracking services, application failure, the way and the speed of solving interruption. The technical teams supporting all of these steps in the implementation and utilization of services are certified

and continuously expands with professionals, due to the more specific and more extensive customer requirements.

## II. MANAGED SOLUTIONS FOR BUSINESS

### A. The effectiveness of managed solutions

In some cases business solutions for connectivity have to be predefined so the users who do not have a lot of experience in the IT field could easier identify more suitable solution for their business, and this it also simplifies monitoring and analyzing the processes quality. All solutions with their alternatives offer the possibility to the client to create his own solution that is most appropriate for his business or operator's presales team could do this for him according to the requirements (tailor made solutions).

The leading operator in the field of Managed Solutions has a platform that is designed to provide innovative and high quality ICT services and solutions for business customers. The entire service portfolio is based on a regional multi-service network supported its own optical infrastructure, modern telecommunications equipment and modern data centers. The primary mission for the operator is to provide a comprehensive portfolio of services to all business customers regardless of the type of activity or size of company. Beside all of this, it should take into account the specific requirements of individual companies and to provide flexible services and customized business solutions.

In order to achieve superior performance and continuous operation service, the operator provides fiber optic infrastructure in his network for the customers. These services are designed for organizations that need a reliable, highly available and fast Internet service. Due to the increasing needs of large companies for symmetric internet access download and upload speeds in this type of services are identical. An important feature is the SLA (Service Level Agreement), which guarantee the required level of service availability. The optical infrastructure provides the ability to add other services that the operator has in its services portfolio easily by VLAN separation. Equipment configuration and activation of the service need to be provided by a team of experts on the operator side, on the equipment collocated at the customer premises.

The main advantages created solutions must be defined and represented to the customer. Clearly must be introduced cost savings and improvements in quality of service and all advantages for the customer. The basic advantages of managed solutions are:

1. Savings in procurement and replacement of equipment, because operator implement the needed equipment and replace it when there is a need for that

<sup>1</sup>Zlata Radak is with the Faculty of Transport and Traffic Engineering, University of Belgrade, Vojvode Stepe 305, Belgrade, Serbia, E-mail: zlata.radak@sbb.co.rs.

<sup>2</sup>Nataša Gospić is with the Faculty of Transport and Traffic Engineering, University of Belgrade, Vojvode Stepe 305, Belgrade, Serbia, E-mail: n.gospic@sf.bg.ac.rs



2. Providing service availability and performance quality by monitoring the SLA and QoS (Quality of Services) parameters

3. The specific technical support (NOC - Network Operations Center). The engineers in the NOC perform proactive monitoring managed CPE equipment 24/7/365 in case they notice any inconsistent behavior they proactively open a trouble ticket and notify the customer at the same time they troubleshooting to solve the problem in the shortest possible time and to raise the performance of the service to the appropriate level

4. Customer has available portal for monitoring QoS (quality of service) parameters

5. By choosing managed solutions company receives a customized solution that is integrated with existing equipment, which means that is compatible with the existing infrastructure and can easily be upgraded to it

6. Pay as grow model, means that the costs of creating and maintaining solutions increase proportionally to the growth of client needs

7. Reducing the cost of procurement of infrastructure (Capex) to migrate to the new technology, reduce maintenance costs

8. Users focus on the core business. So, what's the objective of the managed solutions created this way is that the client relieve worries about the Internet and focuses on the development and improvement of their own businesses

As a leading regional provider of information - communication technologies (ICT), the operator offers business customers complete IT solutions and systems for unified communications. An integral part of the offer is technical support 24/7/365 that is making the integration, installation, commissioning and maintenance of equipment and applications from the company's overall portfolio. Operator bases its services on the products and solutions of the world's largest manufacturers of ICT and developing their own solutions, relying on strong local teams, regional cooperation and individual approach to each customer.

#### *B. Customer Care 24/7/365*

Operator has specialized teams of people who provide support to clients, as well as the NOC (network operations center) that is available 24/7/365 and proactively monitor all links and managed services equipment, for all customers to have adequate support for their services. Proactive monitoring is designed to track the logs in which the central router in a very short intervals of time (a few seconds) sends a ping to all managed routers (CPE) and waits for a response from them. If it happens that some of the routers (which are located on the sites the user) does not return the ping in the time frame that is strictly defined, fire alarm logs become red and engineer immediately begins the troubleshooting link to eliminating possible malfunction. Automatically opens trouble ticket and notifies the user about it. Furthermore, the system continues to count the time it takes to solve the problem and that the trouble ticket is closed. If it is determined that there was a deviation from the guaranteed QoS parameters, the operator is required to pay penalties to the customer.

The biggest advantage of managed service is that sometimes an interruption or other problem is recognized and resolved (trouble ticket opened and closed after solving the problem) and a customer receives a notification that the problem existed and that it was resolved, but he do not has any interruption because it did not occur during working hours. This technique significantly shorten interruptions and the time required to solve the problem, because it is performed early detection of problems and immediately NOC is seeking a solution for the problem occurred. The traditional way of functioning technical support means to wait for information about login failure or interference by the customer and only after complaints from the customer moves in troubleshooting and opens trouble ticket.

#### *C. Further development*

IT budgets are increasingly limited, some of them reduced, and the IT department is required to meet the needs of companies that are more specific. Whether it is a small or medium size company or a large company, cloud services are imposed as a logical step forward.

Cloud services include the rental of IT and communications infrastructure and services of specialized IT house, without making capital investments in their own infrastructure, software and human resources, which is becoming more profitable and more common business model of companies and organizations around the world. They are ideal for companies that have a need for effective implementation of new IT services that will help them to improve business processes, communication with customers and stakeholders and to operate effectively. Services are flexible in terms of volume, are formed towards the real needs, and services are charged according to the scope of use.

Managed services provided by the leading operator can be grouped as follows:

1. Managed Connectivity (Internet connection with managed CPE where the operator at the request of the client changes and adjusts configuration)
2. Managed L3VPN (connecting remote locations with the managed equipment)
3. Managed Voice (Centrex service)
4. Managed Cloud (MS Exchange, SharePoint, SQL Server, ERP)

The advantages of using this services can be summarized as:

1. Reduction of costs
2. Flexible scope of services in accordance with user needs
3. Quickly and cost-effectively launch new services
4. Capacity Expandability
5. Predictability of performance and costs
6. Using the modern equipment
7. The quality and reliability of service with guaranteed SLA
8. Fast troubleshooting with technical support – NOC
9. The possibility of rapid and adequate response in case of changes in the internal / external environment
10. The savings in the area of administration, maintenance and human resources

Advantages and savings achieved through the implementation of a managed services are both short-term (reflected in the initial investment) and long term (lower monthly subscription fee in respect of costs before the introduction of managed services). The tendency of savings in terms of telecommunications and the growing demands in terms of quality leads to the development of innovative model for connecting the large number of locations through a Managed L3VPN with a central location. This involves setting up the Internet connection at the central location (usually the locations with the highest number of employees) and create virtual connections - tunnel (L3VPN) through the operator's network to all the other branches. In this way it is achieved that IT administrator at the central location can easily and securely access data stored on remote locations and vice versa. The administrator who usually sits on the central (main location) has on line access to all locations. Administrator can by simply logging check the status of each location and whether it is active or not. It can also access data at a remote location and perform various settings. If there is a problem with the availability of some sites (some of the locations is offline) may contact the duty NOC operators and engineers to solve the new situation. Since the NOC has remote access to the central (managed) router at each location, they could remotely check all the settings and fix the problem bringing the location to on line status. If the problem can't be solved remotely, duty team is immediately sent into the field to replace the device. By providing such a service with the managed routers, as well as the duty to support and the emergency teams everywhere on the field, we have a situation that administrator from a central location can monitor and maintain the operation of all sites, distributed throughout Serbia, with the assistance of an operator.

#### *D. An innovative managed solution for connecting remote locations through L3VPN with virtualized central location in the operator's Data Center*

The specific requirements that is set was an absolute availability of central location for access by the branches to achieve sensitive and constantly necessary information and to be constantly updated so they could have business continuity and earn their income. The specific client is a chain of retail stores with branches distributed throughout Serbia. At the central location the server through a special application keeps track of the state of the warehouses, current prices and updated information as soon as changes occur because the current and accurate information are necessary at each retail location and at any time. Given the fact that this is a store it is a logical conclusion that every minute unavailability of data involves great financial losses because store is not able to deliver the goods, and buyers in this case, go to the first next stores and thus firm losses its income. So, in order to achieve absolute availability central location managed connectivity has emerged as a logical solution. However, although this type of approach guarantees a high level of availability and proactively monitors the link, the request is still on the availability of 100% and it is reasonably possible only with one or more back-up links. With regard to the possibility of

introducing more optical links by different operators (for redundancy) unjustifiably expensive investment, a proposal by the expert presales team is to take a central server applications with all its data and relocate at the operator's Data Center and to ensure Managed Services. This solution achieves absolute uptime and all the necessary data that is on it. All locations, including the main access to the server and this is provided through a L3VPN tunnel. High availability of the central location makes all the data available for the smooth operation of all stores.

In the standard option of placing the equipment in the Data Center, only the business user (or an authorized person from the company) has access to the equipment both physical and remote. Operator physically protects the equipment and provides redundant power supply, protecting against flooding and fire, with modern systems of protection, and connects to the Internet through triple redundant links, maintain an adequate temperature in the room needed to operate the device. What is also important, there is a team of people who physically protects the equipment at the site from unauthorized access. All of these benefits are exactly client's needs on its central location. Since in this case is the central location relocated to the operator's Data Center customer actually realizes all the benefits of placing equipment in the Data Center plus an additional option of getting managed services which means that the operator takes over the servicing of the monitoring and care of the link to the main server. These solution has achieved maximum performance of the master server which was the request.



Fig. 1. Managed L3VPN Solution with central location in the operator's Data Center

In order to meet the client's needs in the best possible way, the provider is making a real innovative managed service solutions and implemented in an unusual manner (Fig. 1). In the data center is placed adequate router with all the necessary data. This establishes a managed connection through L3VPN from a central location to all remote locations (stores) which also have the managed equipment that was configured and monitored by the operator's expert team. Routers communicate over L3VPN tunnel that is established between the central site and all remote locations.

Also, in addition to all the connections a special tunnel to the central location was made in order to update all the data in a timely manner. Through this tunnel IT manager can access software and data stored in the data center (cloud managed solution) as well as other locations in the L3VPN MPLS network.

Thus configured solution ensures maximum performance for business users and absolute availability of servers in the Data Center, with the data being replicated to a central location, for extra protection. Also, it offers various advantages for IT administrators. Automation services (services managed by a professional team of operators) significantly simplifies the work of IT administrators and allows him to devote additional time improving business applications and to introduce a new systems in order to improve sales and facilitate the work for his employees.

### III. CONCLUSION

Relocation of the central location in the operator's Data Center provide the maximum possible availability of the central location and because it is located in modernly equipped Data Center, maximally protection. It also established a special connection from the actual location of the customer to the virtual central location in the operator's Data Center. The cost of such a solution are far lower than that the client has chosen to make himself redundant connections such as Data Centre at his premises, which would mean higher monthly costs and more engaged workforce. In addition to the initial very large costs of building such facilities, it would also be necessary to hire additional human resources to control the operation and the accuracy of the facilities and equipment in it 24/7/365.

In order to maintain its leading position in the market, management believes that in addition to providing pure services, should be going toward the goal of providing a complete solution for each user separately. Taking into account the specificity of each client's business in the global market, many professional teams are engaged to make the best technical solution for each customer individually. The complete solution includes linking all locations and providing the entire spectrum of telecommunications services and Internet telephony service L3VPN and end cloud services and TELEHOUSING.

It was concluded that these solutions require each client and its implementation described in this paper is a contribution to a specific company and further development of this field. Operator has become a provider of complete solutions and partner that supports the client in its development of the market and thus ensures its position in the market.

### ACKNOWLEDGEMENT

For this paper were used information and facts derived from a specific case that was taken into consideration. The solution

described in this paper exist as such and functions and also delivers significant savings to the client that uses it.

Comparison of costs before the introduction and after the introduction of managed solution with virtualized central location are the subject of further research.

### REFERENCES

- [1] B. Javorović, M. Bilandžić, "Poslovne informacije i business intelligence", Golden Marketing – Tehnička knjiga, Zagreb, pp. 195-202, 2007.
- [2] Ž. Panian, G. Klepac, " *Poslovna inteligencija* ", Masmedia, Zagreb, 2003
- [3] Ž. Panian i suradnici, *Poslovna inteligencija – Studije slučajeva iz hrvatske prakse*. Narodne novine, Zagreb, 2007.

# 2D Localization of Source of Stochastic EM Radiation by using Neural Networks

Zoran Stankovic<sup>1</sup>, Ivan Milovanovic<sup>2</sup>, Nebojsa Doncov<sup>1</sup>, Bratislav Milovanovic<sup>2</sup>

**Abstract** – 2D localization of narrow-band electromagnetic source of stochastic radiation nature in far-field is considered in the paper. The model presented is based on artificial neural networks (ANN). This model is trained to perform accurate and efficient 2D direction of arrival (DoA) determination of electromagnetic signals radiated from some stochastic source, That changes its position in a parallel plane, to the level of a 2D antenna array, in the far-field scan area. This antenna array is used for sampling the spatial correlation matrix signal that provides the necessary information for the training model.

**Keywords** – Source localization, Mobile users, Stochastic radiation, Correlation matrix, Neural networks.

## I. INTRODUCTION

Spatial filtering of the antenna array signal and shaping the radiation characteristics, using adaptive antenna arrays, are now current techniques. They unveil the possibility to efficiently decrease the negative impact of interference, on the signal reception site, and therefore the possibility of a significant increase in the number of users of modern wireless communications as well as, the service quality of services that such systems offer to the customers [1,2]. In parallel with the above fact, the techniques relating to the localization position of the source signal by the passive sensor array also attract the attention today because there is an increasing need for their application in geophysics, satellite communications, radio-astronomy, biomedical engineering, radar systems engineering 5G and other forms of wireless communication.

In applying the above techniques of both classes, procedures have a very important role for DoA estimation of the signal. Today the most commonly used super-resolution algorithms for DoA estimation such as MUSIC [1,2] and its modifications have a high accuracy in determining the directions of where the EM signals come from, but because of its complex matrix calculation, it requires powerful hardware resources and are not suitable for operation in real time. In papers [3-11] is shown that the alternative methods DoA estimation algorithms can be super-resolution models based on artificial neural networks [3,12 to 13]. Neural models for DoA estimation avoid complex matrix calculations, and can have an approximate accuracy of the MUSIC algorithm. They are faster than the MUSIC algorithm which makes them more suitable choice for implementation in real time [3 to 4,11].

<sup>1</sup>Zoran Stanković and Nebojsa Dončov are with the Faculty of Electronic Engineering, University of Nis, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: [zoran.stankovic, nebojsa.doncov,]@elfak.ni.ac.rs

<sup>2</sup>Ivan Milovanović and Bratislav Milovanović are with the Singidunum University, centre Nis, Nikole Pašića 28, 18000 Niš E-mail: [imilovaovic, bmilovanovic]@singidunum.ac.rs

In the papers [4-5] su predstavljani efficient neuron models are presented. For 1D DoA [4] and 2D DoA [5] estimation of deterministic radiation source. In papers [6-11] are presented neural models for 1D DoA estimation origin with the stochastic nature of EM radiation wave [14,15] where the stochastic sources moving along one direction in the azimuthal plane and where their positions are characterized by a single angular coordinate (azimuth).

This paper goes a step further in research in relation to the works [6-11] because it now allows the stochastic radiation source moving in 2D space (the plane), and its position is located using two angular spatial coordinates. These coordinates represent the angles of the spherical coordinate system under which the stochastic EM radiation sources comes to a rectangular planar antenna array, or in other words to the angles that are obtained using the method of 2D DoA estimation to the plane of the antenna array. A scenario is considered where a single source of stochastic radiation changes its position in the plane relative to the planar antenna array, which is located in the far zone of radiation and which is parallel to the plane of movement of the relative level of stochastic sources. This scenario, with certain approximations, and ignoring the effect of curvature of the earth's surface, can be present in passive radar and other sensors that are based on antenna arrays and mounted on satellites that are in low-earth orbit, planes or drones in order to make detection and localization of the source of radiation at the earth's surface.

## II. STOCHASTIC SOURCE RADIATION MODEL

In this paper, a model of stochastic radiation source in a far zone which is also applied in the works [6-11]. With this model stochastic radiation sources in the far zone presents a linear uniform radiation of the N antenna array elements that are at the distance d (Figure 1). The degree of correlation between the supply current of the elements of an antenna array that is described buy the vector  $\mathbf{I}=[I_1, I_2, \dots, I_N]$  is defined by the correlation matrix  $\mathbf{c}^I(\omega)$  [14,15]:

$$\mathbf{c}^I(\omega) = \lim_{T \rightarrow \infty} \frac{1}{2T} [\mathbf{I}(\omega) \mathbf{I}(\omega)^H] \quad (1)$$

In the zone of far-field electric field strength at the selected sampling point is calculated in a way

$$\mathbf{E}(\theta, \varphi) = \mathbf{M}(\theta, \varphi) \mathbf{I} \quad (2)$$

where  $\mathbf{M}$  represents the scanning with Green's function

$$\mathbf{M}(\theta, \varphi) = jz_0 \frac{F(\theta, \varphi)}{2\pi r_c} [e^{jkr_1} \quad e^{jkr_2} \quad \dots \quad e^{jkr_N}] \quad (3)$$



In the equations (2) and (3)  $\theta$  and  $\varphi$  represent the spatial position of the corners of the first antenna element of the antenna array that represents the source in relation to the selected sampling point,  $F(\theta, \varphi)$  is the radiation characteristic of the antenna element,  $r_c$  the distance between the selected point from the center of the antenna array,  $z_0$  is the impedance of free space,  $k$  is the phase constant ( $k=2\pi/\lambda$ ) while  $r_1, r_2, \dots, r_N$  represent the distance selected point from the first to the  $N$ -th antenna element respectively (Fig. 1). When at the reception we use a planar antenna array rectangular dimensions  $M \times P$ , sampling points will be in positions of planar antenna elements and a series of them will be  $K=M \cdot P$ . In our scenario (Fig. 1) receiving planar antenna elements that represent a set of sampling points are arranged in  $x$ - $y$  level which is paralelna level where the generation source  $S$  relative moving relative to the receiving antenna array. The distance between the elements of the receiving antenna array along the  $x$  axis is  $s$ , while the distance between the elements along the  $y$  axis  $h$ . The distance between the level of the planar receiving antenna array and the plane in which the mobile generation source  $S$  is  $r_0$ . Introduced the assumption that the linear antenna array, which is modeled stochastically source at the beginning of the movement oriented along the  $x$  axis and in  $d \ll r_0$  ignored changes its orientation when moving sources. When copying (3) is applied to each sampling point individually, the appropriate distance between the  $n$ -th element of the antenna array that represents the stochastic radiation source  $S$  and the sampling point at the position sensor  $(m, p)$  of a planar receiver array is

$$r_{mp}^{(n)} = \frac{1}{\cos \varphi_{mp}} \sqrt{r_0^2 + [r_0 \cdot \tan \theta_{mp} + (n-1) \cdot d]^2} \quad (4)$$

where  $r_0$  is the distance between the planes of the planary antenna array at the receiving end, and the plane in which the stochastic source  $S$  is moving while  $\theta_{mp}$  and  $\varphi_{mp}$  are spatial angles relating to the position of the first antenna element in relation to the source  $(m, p)$  sensor position and they are

$$\theta_{mp} = \arctan \left[ \frac{(m-1) \cdot s}{r_0} + \tan \theta_{11} \right] \quad (5)$$

$$\varphi_{mp} = \arctan \left[ \frac{(p-1) \cdot h}{r_0} + \tan \varphi_{11} \right] \quad (6)$$

whereby  $\theta_{11}$  and  $\varphi_{11}$  spatial angles relating to the position of the first antenna element sources relative to a reference  $(1,1)$  position of the sensor. The angles  $\theta_{11}$  and  $\varphi_{11}$  also represent the angular position  $(\theta, \varphi)$  stochastic source  $S$  in relation to the linear receiving antenna array so that the  $\theta_{11} = \theta$  i  $\varphi_{11} = \varphi$ .

Correlation matrix signal in the sampling points is defined like so

$$\tilde{C}_E[i, j] = \mathbf{M}(\theta_{[i/P]+1}, \varphi_{[i/P]+1}) \mathbf{c}' \mathbf{M}(\theta_{[j/P]+1}, \varphi_{[j/P]+1})^H, \quad (7)$$

$$i = 1, \dots, K \quad j = 1, \dots, K \quad K = M \cdot P$$

On the basis of the equations (4)-(6) is determined by the vector  $\mathbf{M}$  is given in equation (3), then in accordance with the angular position of stochastic energy sources to the sampling point, the elements of the correlation matrix. If the array elements are normalized with respect to the first element of the matrix

$$\mathbf{C}_E = \frac{1}{\tilde{C}_{E11}} \cdot \tilde{\mathbf{C}}_E \quad (8)$$

in our scenario is obtained correlation matrix that does not depend on the value of  $r_0$ ,  $r_c$ ,  $F(\theta, \varphi)$  i  $N$ . For training the neural network of that model made enough to take only the first type of matrix  $\mathbf{C}_E$  ( $[C_{E11}, C_{E12}, \dots, C_{E1K}]$ ) because it turned out that the first type contains sufficient information to determine the angular position of the radiation source [5,6].

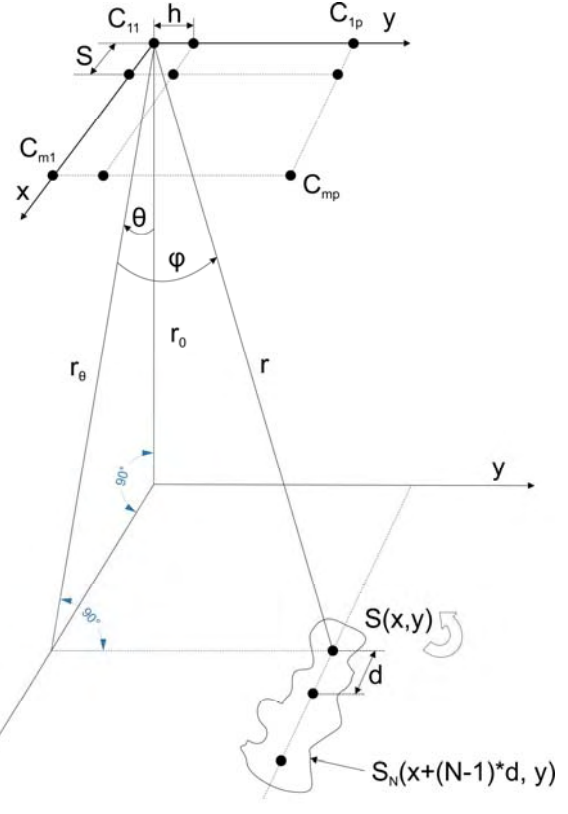


Figure 1. The position of stochastic source in  $x$ - $y$  plane with respect to the location of EM field sampling points in the far-field scan area.

### III. NEURAL NETWORK MODEL

The neural model based on MLP ANN [12,13] is developed with the purpose to perform the mapping from the space of signals described by correlation matrix  $\mathbf{C}_E$  to the 2D DoA space

$$[\theta \ \varphi]^T = f(\mathbf{C}_E) \quad (9)$$

where  $[\theta \ \varphi]^T$  is vector of spatial angles of arrival of the stochastic source radiation. The architecture of developed neural model is shown in Fig.2. Its MLP network can be described by the following function:



$$\mathbf{y}_l = F(\mathbf{w}_l \mathbf{y}_{l-1} + \mathbf{b}_l) \quad l=1,2 \quad (10)$$

where  $\mathbf{y}_{l-1}$  vector represents the output of  $(l-1)$ -th hidden layer,  $\mathbf{w}_l$  is a connection weight matrix among  $(l-1)$ -th and  $l$ -th hidden layer neurons and  $\mathbf{b}_l$  is a vector containing biases of  $l$ -th hidden layer neurons.  $F$  is the activation function of neurons in hidden layers and in this case it is a hyperbolic tangent sigmoid transfer function:

$$F(u) = \frac{e^u - e^{-u}}{e^u + e^{-u}} \quad (11)$$

In order to perform mapping it is sufficient to take only the first column of correlation matrix and therefore  $\mathbf{y}_0 = [\text{Re}\{\mathbf{C}_E[1,1]\}, \dots, \text{Re}\{\mathbf{C}_E[1,K]\}, \text{Im}\{\mathbf{C}_E[1,1]\}, \dots, \text{Im}\{\mathbf{C}_E[1,K]\}]$ . Also, output of the neural network model is given as  $[\theta \ \varphi]^T = \mathbf{w}_3 \mathbf{y}_2$  where  $\mathbf{w}_3$  is a connection weight matrix between neurons of last hidden layer and neurons in output layer. The optimization of weight matrices  $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$  and biases values during the training allows ANN to approximate the mapping with the desired accuracy.

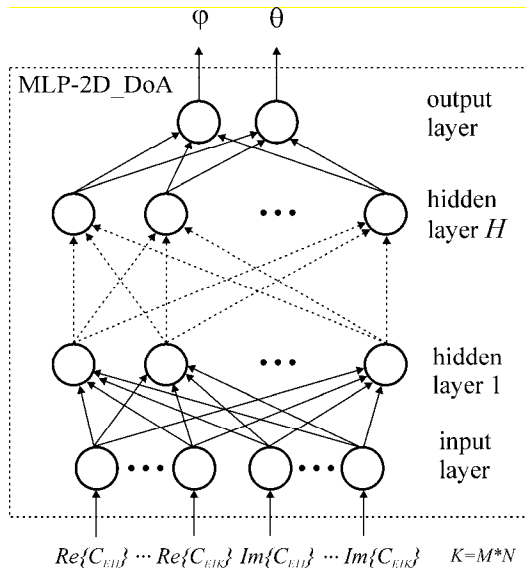


Figure 2. Architecture of MLP neural model for 2D DOA estimation of stochastic EM source signal in  $x$ - $y$  plane plane

The general designation for this defined MLP neural model is  $\text{MLP}H-N_1 \dots N_H$  where  $H$  is the total number of hidden layers used MLP network, while  $N_i$  is the total number of neurons in the  $i$ th hidden layer.

#### IV. MODELING RESULTS

Neural model with the architecture presented in the previous section modeled the hypothetical scenario of stochastic radiation source described in section II, where the stochastic source antenna represents a series of two elements with uncorrelated currents supply a sampling signal in a far zone is carried out in nine points that are distributed the equidistant spacing in the form of a rectangular planar set of dimensions  $3 \times 3$ . For the characteristic radiation element antenna array origin is taken isotropic characteristics. The feed currents of two elements are mutually uncorrelated so that  $\mathbf{c}^l$  is the unit diagonal matrix. Table 1 provides the

values of parameters of the scenarios that were used to generate samples for training the neural models. For the realization of the training model used is the Matlab software development environment. Sets of samples for training and testing MLP models were generated by using equation (3) and (7). Any combination of angles  $\theta$  and  $\varphi$  which is defined by the distribution patterns associated with the vector of 18 elements, represents the first type of signal correlation matrix (9 elements for the real part, and 9 elements in the imaginary part of the complex value of the first type correlation matrix).

TABLE I  
THE VALUES PARAMETERS WHICH USED IN SAMPLING PROCESS

Frequency	$f = 22$ GHz
Number of antenna array elements per one source	$N = 2$
Sampling points distance from source trajectory	$r_0 = 600$ km
Number of sampling points along $x$ axis	$M = 3$
Mutual distance of the sampling points along $x$ axis	$s = \lambda/2$
Number of sampling points along $y$ axis	$P = 3$
Mutual distance of the sampling points along $y$ axis	$h = \lambda/2$

TABLE II  
TESTING RESULTS FOR SIX MLP NEURAL MODELS WITH THE BEST AVERAGE ERRORS STATISTICS

MLP model	WCE [%]	ACE [%]
MLP2-15-11	2.52	0.38
MLP2-12-12	2.74	0.39
MLP2-18-14	2.78	0.38
MLP4-13-13	2.79	0.37
MLP2-20-10	2.80	0.38
MLP2-18-7	2.79	0.38

For the model training, a set of 14641 samples with uniform distribution  $\theta$  and  $\varphi$  angles in the range  $[-30^\circ \ 30^\circ]$  with a  $0.5^\circ$  step. Quazi-Newton method with prescribed accuracy of  $10^{-4}$  is used as a training algorithm. For the testing of the model a set has been generated of 7396 samples with uniform distribution  $\theta$  and  $\varphi$  angles in the range  $[-30^\circ \ 30^\circ]$  with a  $0.7^\circ$  step. The testing results for six MLP models with the lowest average (ACE) and worst case error (WCE) are shown in Table II, and MLP2-15-11 is chosen as representative neural model. The neural model scattering diagram of testing samples set shows a very good agreement between the output values of neural model and referent  $\theta$  and  $\varphi$  values (Fig.3 and Fig.4).

Using the MLP2-15-11 model a simulation has been conducted for tracking the movement of the hypothetical source of stochastic radiation on earth's surface in a square area in size  $800 \times 800$  km. The source has changed its position along the test trajectory, which has been set by the function  $y = 3 \cdot 10^{-6} \cdot (x \cdot 10^5)^2 - 3 \cdot 10^5$  where  $x$  and  $y$  are relative latitude and longitude expressed in meters. Evaluation paths of origin was carried out by sampling time correlation matrix of the 69 points shown in Figure 5. A satisfactory agreement can be observed between the values of the source positions

which was estimated by the neuron model and the referent source trajectory.

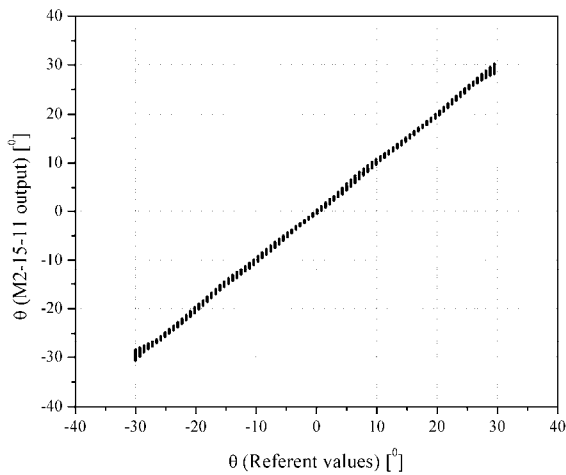


Figure 3. Scattering diagram of MLP2-15-11 model  $\theta$  output

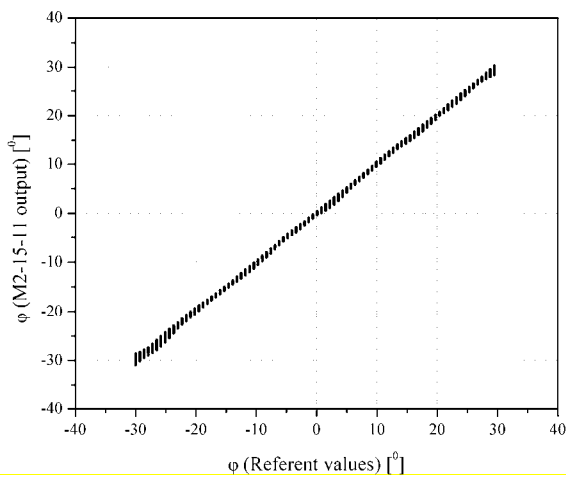


Figure 4. Scattering diagram of MLP2-15-11 model  $\phi$  output

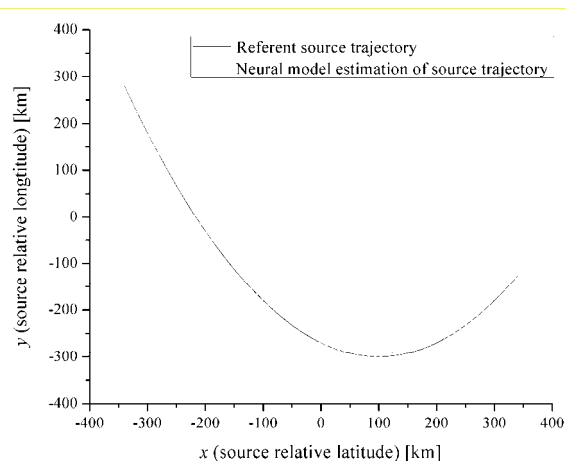


Figure 5. A simulation of localization and tracking the movement of the hypothetical source of stochastic radiation on earth's surface with the MLP2-15-11 model

## V. CONCLUSION

Neural model ability to accurately and efficiently determine the 2D location of the stochastic source is illustrated on one example. As proposed neural model avoids intensive and time-consuming numerical calculations it is more suitable than conventional approaches for real-time applications. At the moment, neural model is capable to determine the 2D location in a plane for one stochastic source. Future research will be focused to the more general 2D DOA estimation of multiple stochastic sources.

## ACKNOWLEDGEMENT

This work has been supported by the Ministry for Education, Science and Technological Development of Serbia, projects number TR32052 and TR32054.

## REFERENCES

- [1] R. Schmidt, "Multiple emitter location and signal parameter estimation", *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276-28, 1986.
- [2] B. Allen, M. Ghavami, *Adaptive Array Systems: fundamentals and applications*, Wiley, 2005.
- [3] C. G. Christodoulou, M. Georgiopoulos, *Application of neural networks in electromagnetics*, Artech House, December 2000.
- [4] A. H. El Zooghby, C. G. Christodoulou, M. Georgiopoulos, "A neural network based smart antenna for multiple source tracking", *IEEE Transactions on Antennas and Propagation*, Vol. 48, no. 5, pp. 768 – 776, 2000.
- [5] M. Agatonović, Z. Stanković, N. Dončov, L. Sit, B. Milovanović, T. Zwick, "Application of artificial neural networks for efficient high-resolution 2D DOA estimation", *Radioengineering*, Vol. 21, No. 4, pp. 1178-1186, 2012.
- [6] Z. Stanković, N. Dončov, J. Russer, T. Asenov and B. Milovanović, "Efficient DOA estimation of impinging stochastic EM signal using neural networks", *Proceedings of the International Conference on Electromagnetics in Advanced Applications* (15th Edition including EMS), ICEAA 2013, Torino, Italy, September 9-13, pp.575-578, 2013.
- [7] Z. Stanković, N. Dončov, J. Russer, I. Milovanović, M. Agatonović, "Localization of Stochastic Electromagnetic Sources by using Correlation Matrix Trained MLP Neural Network", *Microwave Review Serbia*, No.2, Vol. 19, pp. 44-49, December 2013.
- [8] Z. Stanković, N. Dončov, J. Russer, I. Milovanović, B. Milovanović, "Neural Network Approach for Efficient DOA Determination of Multiple Stochastic EM Sources in Far-field", *Proceedings of the 1st IEEE International Conference on Numerical Electromagnetic Modeling and Optimization for RF, Microwave, and Terahertz Applications*, NEMO 2014, Pavia, Italy 14-16 May, 2014, pp.1-4, 2014.
- [9] Z. Stankovic, N. Doncov, I. Milovanovic, B. Milovanovic, M. Stoiljkovic, "Localization of mobile users of stochastic radiation nature by using Neural Networks", *Proceedings of the 49th International Scientific Conference on Information, Communication and Energy Systems and Technologies*, ICEST 2014, Niš, Serbia, June 25 – 27, 2014, Vol.2, pp.347-350, 2014.
- [10] Z. Stankovic, N. Doncov, I. Milovanović, B. Milovanović, "Neural network model for efficient localization of a number of mutually arbitrary positioned stochastic EM sources in far-field", *Proceedings of the 12th Symposium on Neural Network Applications in Electrical Engineering*, NEUREL 2014, Beograd, Serbia, pp. 41-44, 2014.
- [11] Z. Stanković, N. Doncov, B. Milovanovic, I. Milovanovic, "Efficient DoA Tracking of Variable Number of Moving Stochastic EM Sources in Far-Field Using PNN-MLP Model", *International Journal of Antennas and Propagation*, vol. 2015, Article ID 542614, 11 pages, 2015. doi:10.1155/2015/542614
- [12] S. Haykin, *Neural Networks*, New York, IEEE, 1994.
- [13] Q. J. Zhang, K. C. Gupta, *Neural networks for RF and microwave design*, Artech House, Boston, MA, 2000.
- [14] J.A. Russer, T. Asenov and P. Russer, "Sampling of stochastic electromagnetic fields", *IEEE MTT-S International Microwave Symposium Digest*, Montreal, Canada, pp. 1-3, 2012.
- [15] J.A. Russer, P. Russer, "Modeling of Noisy EM Field Propagation Using Correlation Information", *IEEE Transactions on Microwave Theory and Techniques*, Volume 63, Issue 1, pp.76-89, 2015.

# Analysis and evaluation of the human factor in the communication and information systems

Evgeniy Hubenov<sup>1</sup>, Antonio Andonov<sup>2</sup>, Zoya Chiflidzhanova<sup>3</sup>

**Abstract** – The article presents the generation and processing of flows of information objects associated to the status of the networks and the services in TCP/IP environment. It is explained a discrete-event processing of the flows of the information objects for monitoring and control of the networks and the services. It is shown the control of the communication network in order to optimize the transport and the processing of data flow.

**Keywords** – discrete-event communication and information systems, TCP/IP environment, information objects, event processing, services, monitoring, control

## I. INTRODUCTION

The specifics of the modern telecommunication systems is their transition to a global information society and as a consequence to communication and information systems (CIS) and networks. The technologies, using the principle of commutation of channels are used together with the networks with packet commutation, networks based on different protocols, including protocols of the TCP/IP architecture. Moreover, there is a considerable diversity of the used equipment and technologies, even in the systems of one operator. The development and the mobility of the modern network information services is directed towards an expansion, covering more areas in order to increase the subscriber base and the services provided, and as a consequence - leads to increase of the scale and complexity of the network structure. All this implicitly affects the quality of the provided services, as one of the most important components is the reliability of the network.

The work of the operators becomes more complicated and to the quality of the service are placed higher and higher requirements. Therefore, the effectiveness of the customer service and the increasing of the level of the performance of the operators are key for the increasing of the efficiency of the operation of the centers for monitoring and management of the services (Network Operations Center, NOC, Call center) in the conditions of strong competition.

In this regard, in present time there is an actual scientific and technical task for research and development of monitoring, analysis and management of the quality of the work of the operators in the centers for monitoring and management of the services (network operations center -

NOC), the solution of which opens up prospects for the establishing of an uniform methodology for evaluation and rise in the indicators of functioning. The centers for management and monitoring are the most common types of systems, which employ a large number of operators. Over the time they develop as: 1) systems for calls, 2) information centers, 3) call centers, 4) integrated call centers and service management. The mechanisms for the service of the calls in the call centers integrated with Internet are considerably more complex, requiring powerful subsystem of administration management, allowing operative reacting to the changes in the external environment. NOC are technical complexes for reception and processing by operators of a flow of phone calls from subscribers. They accept and serve the incoming and outgoing calls and consist of a certain amount of working places, distributed among several departments. The main functions of these centers are realization of the interaction with the subscribers in the necessary volume and efficient mode.

The maintenance operators of telecommunication systems typically operate in a deficit of time. The quality of their work is determined by various factors, including the training of the operator, the working conditions on the workplace, the duration of the shift, the organization of his activities.

The constantly increasing demand for Call centers in the business, in the banking and in the financial sector, in the insurance, in the telecommunications facilities, etc. requires theoretical and experimental research in order to develop algorithms and tools for monitoring [1,2], mathematical modelling [3], analysis and management of the quality of the work of the operators [4,5].

The purpose of the article is the guideline for the increase of the efficiency of the customer service and reduction of the operational losses in NOC by development of algorithms and tools for monitoring, analysis and management of the quality of the work of the operators.

## II. ANALYSIS OF THE PROBLEM AREA AND FORMULATION OF THE PROBLEM.

According to the available literature and the different expert statements, the work of the operators in the communication network is becoming more skilled, strict and stressful as a result of the increasing complexity of the network systems. Many of the studies [4, 6] are exploratory in nature and do not focus on the requirements and the competences assigned to the operators.

Of interest are the decisions of the following problems:

- how the operators / employees / participants / perceive the network as an object of their activity;

<sup>1</sup>Sofia University St. Kl. Ohridski., 5 James Bourchier Blvd., 1164 Sofia, e-mail: [hubenov@ucc.uni-sofia.bg](mailto:hubenov@ucc.uni-sofia.bg);

<sup>2</sup>University of Transport Todor Kableshkov, 158 Geo Milev str., 1574 Sofia, Bulgaria, E-mail: [andonov@vtu.bg](mailto:andonov@vtu.bg).

<sup>3</sup>Space Research and Technology Institute – BAS, Acad G. Bonchev Str., Bl. 1, 1113 Sofia, Bulgaria, E-mail: [zhubenova@space.bas.bg](mailto:zhubenova@space.bas.bg)

- what skills, knowledge and communication skills and other requirements are laid down to the work of the CIS;
- what is the impact of the human errors on the reliability of the network.

In the most publications, the question of the reporting of the impact of the errors of the human - operator on the operation of the CIS, is rarely touched on, but the practice shows that the effectiveness of the new technique increases, if you take into account the human factor ever since the design stage. When developing the new generation of CIS in connection with the intensive development of the technical and software means, special attention is paid to the improving of the quality and the reliability of the functioning of the communication system, leading to the need for analysis of the reliability of each separate unit, including also that of the human-operator. Telecommunication systems ensure the exchange of information between subscribers, thus the subscriber is not only an user (client), but operator in the process of exchange in the systems and the communication networks. Therefore, CIS can be viewed as a "human-machine" system (HMS), one of which units is the human-operator. In the most cases in practice, the problem of interaction between the human and technical subsystems is ignored, which entails a risk of unforeseen situations. The increasing cost of the mistakes of the operator determine the continued need for seeking of ways and means to ensure the reliable functioning of the human in the communication system [7]. Consequently, one of the new aspects of the problem of the reliability of the CIS became the study of the unpredictable, unforeseen emergency situations caused by errors, inactivity or insufficiently fast - acting of the operator.

The analysis of the nature of the human-machine interaction in the centers for information services allows to increase the level of algorithm development of the activities carried out by the operators and to determine the optimal amount of hierarchical levels in the system. It allows the opportunity to realize a model of group interaction between the operators and to do decomposition of the tasks and of the different types of requests (regulated and not regulated). Generalized structure of a hierarchical human-machine system of this type is shown on Figure 1.

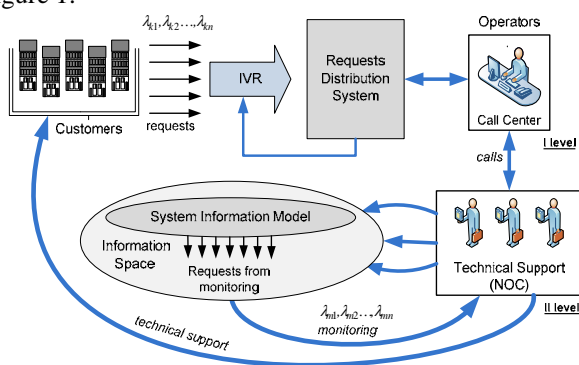


Fig.1. Organization of the system for processing of requests and service in CIS

The output data for the quality of the fulfillment of the individual operations of the activity of the operator at given characteristics of the operators and other factors affecting the environment are considered to be set. The principle of

organization of the system for processing of the requests for technical support includes the incoming requests to the center and the monitoring results. The activities of the operators have a defined algorithmic character and for some of the problems, there are alternative algorithms for their service [8].

When analyzing the reliability of the communication and information networks, as well as in the other structurally complex technical systems, there are a variety of methods and solutions. Tasks of this type traditionally occur at any stage of their operation. During the designing, they should be legally set, which means to be defined and justified the selected indicators of reliability for each item and for the CIS as a whole. At the stage of inspection and operation it is necessary to clarify the factual reached reliability, to be compared with the required one and thus, predictions about the reliability to be made. The analysis of the reliability of the CIS is often complicated by the fact that in many of the cases the CIS itself under certain circumstances is modified during the operation. The reliability of the human- operator is determined by his ability to fully perform his assigned functions under certain conditions. The reliability of his activity is characterized by his faultlessness, willingness, recoverability, timeliness and accuracy [9, 10]. For the selection of optimal information support is necessary to take into account the psycho-functional features of the operator, i.e. the degree of tiredness and the informational loading of the person.

The solution of these tasks is impossible without creating a comprehensive system for analysis of the reliability of the CIS, including the various factors affecting the reliability, their identification and classification. As a first level of the classification of these factors can be used the localization of the corresponding sources of influence (in particular, the reasons for failure):

- systemic conditions - create sources of impact in the object during the stages of operation;
- operating conditions - create sources of failure that occur during the operation of the object;
- maintenance conditions - create sources of failures that occur during maintenance.

For the carrying out of an analysis of the loading at the receiving of the requests, it is necessary to have information on:

- the number of the incoming calls;
- the number of the calls distributed;
- the number of the serviced calls;
- the number of the missed calls;
- the average time of conversation.

The indicators of the quality of the work of the operators in the NOC [11] can be conditionally divided into:

- indicators for the quality of the service of the subscribers by the operator;
- indicators for the quality of the service of the NOC;
- indicators for the subscribers satisfaction.

The indicators for the processing of the requests determine the interaction of the operator with the subscriber and consist of several characteristics: orientation to the subscriber, authority and professionalism, but also effectiveness and efficiency [12].



To increase the quality of the work of the operators in the NOC, it is necessary to implement a monitoring. The monitoring of the quality of the work of the operators is conditionally divided into two types [13].

- regular (planned) control of all operators;
- control of particular operators, performed if necessary.

The decision for the carrying out of a control of individual operators is accepted, as a rule, on the base of statistic data on their productivity. For example, the average time to process a call from one operator is significantly worse than of that of another one. The truth is that in this case, it is necessary to do a differentiation of the operators and to determine the average time for the processing of the calls.

### III. METHODOLOGY FOR EVALUATION OF THE EFFECTIVENESS OF THE SYSTEM FOR THE PROCESSING OF THE REQUESTS

Since the focus of this exploratory study is to obtain information about the factors that influence the human reliability in the operation of CIS, a Bulgarian telecommunication operator and network service provider with 15 - 20 workplaces and continuous working hours was considered.

To analyze the work on the servicing of the requests, they are divided into two flows. The first flow is from the customers to the operators from first level of the servicing of the requests. The distribution of the requests in the form of phone calls is related to the number and the groups of operators in the center for customer service. This flow of requests can be divided into two flows:  $\lambda_{k1}$  - requests type serviced with a certain delay and of second type  $\lambda_{k2}$  - requests of missed calls. The flow of requests to the staff (technical support at second level), can also be divided into two flows -  $\lambda_{p1}$  - requests that are serviced by the support staff within the expected time for restoration of the service (MTTR) and requests that are missed and the technical problems are not resolved within the expected time  $\lambda_{p2}$ . The probability of omission of the service for the both levels of support of requests are defined as a sum of the probabilities of omission in the different phases of the processing of the requests, by the formulas [14, 15]:

$$P_{k,m} = \lim_{t \rightarrow \infty} \frac{A_k(0,t) + A_{k,l}(0,t) + \sum_{j=1}^V A_{k,o}(0,t)}{A_{k,inc.}(0,t)} \quad (1)$$

where:  $P_{k,m}$  – probability of omission of requests;  $A_k$  – number of refused requests;  $A_{k,l}$  – number of not serviced requests of received calls;  $V$  – number of operators;  $A_{k,o}$  – not serviced by the operator requests;  $A_{k,inc.}$  – number of all incoming requests.

$$P_{p,mis} = \lim_{t \rightarrow \infty} \frac{A_p(0,t) + A_{p,d}(0,t) + \sum_{j=1}^W A_{p,op}(0,t)}{A_{p,inc.}(0,t)} \quad (2)$$

where:  $P_{p,mis}$  – probability of omission of information from the monitoring;  $A_p$  – number of not serviced requests because of the monitoring;  $A_{p,d}$  – number of requests not distributed for operators;  $W$  – number of operators;  $A_{p,op}$  – not serviced by the operator requests;  $A_{p,inc.}$  – number of all incoming requests.

The most advantageous in terms of the probability of failure is the use of large groups of operators. Such groups are difficult to manage and are not economically profitable due to the specialization and the profiling of each group. That is why in practice at overload, the requests are transferred to another group. The escalation of the requests is done also at the automated processing of the telephone calls. With voice menu users are self-served at a requirement, at an availability of a problem for which the information is available and is already working over.

The parameter of the quality of the fulfillment of the requests with escalation to the second group is determined by the expression:

$$Q_S = 1 - P_{k,0} P_{k,1} \dots P_{k,S} \quad (3)$$

where  $S$  is the number of the organized escalations.

Figure 2 shows the dependence of the parameter of the quality of the probability of failure in a different number of escalations.

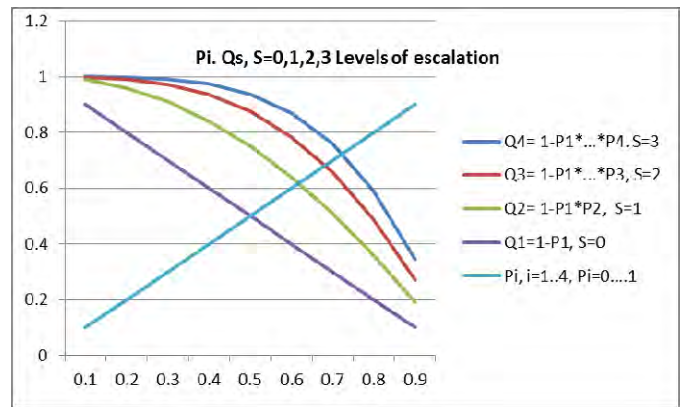


Fig.2. Dependence of the parameter for the quality of the execution of the requests of the probability for omission and the levels of escalation

Such an approach reduces the probability of rejected requests, but at a greater number of levels of escalation, the time for the transfer of the requests to the other level increases. The development of procedures and the automation of the process of escalation reduce this technological time.

The relative share of the probability of escalation of requests (problems) decreases linearly with the time as a result of the experience and the qualification of the operators at all



levels and the improving of the organization of the processing of the requests:

$$\Delta(R)/R = -t/T, \quad R(0) = R_0, \quad (4)$$

where  $R$  is the probability of escalated problems;  $R_0$  - initial value (higher in poorly trained operator).

The role of the human factor for the reducing of the servicing time (training, experience, qualification) is defined by an exponential law:

$$R(t) = R_0 \exp(-t/T), \quad (5)$$

where  $R_0$  is the initial probability for escalation of the problem by the operator (for poorly trained operators is closer to 1)

$T$  is a temporary parameter which is associated with the training of the operators and the lower values correspond to a higher level of qualification and training.

Figure 3 shows dependences of probability of escalation of problems depending on the training of the operators.

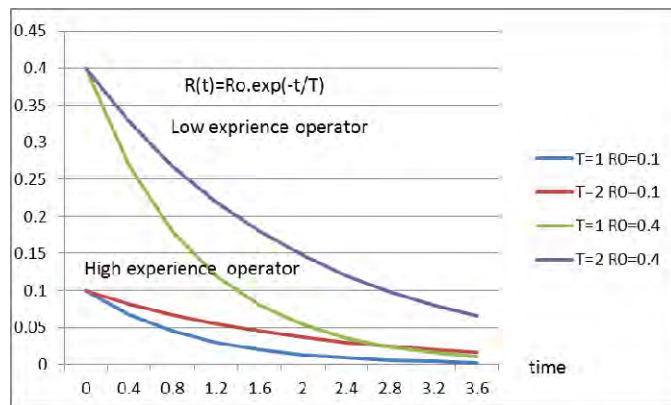


Fig. 3. Probability of escalation of problems depending on the human factor

#### IV. CONCLUSIONS

The use of escalation of the requests in the model of their processing reduces the probability of omission of requests and provides procedures resulting in their final decision. The increasing of the number of the levels of escalation, including also the procedures for automatic escalation increases the factor for the quality of the servicing of the requests. The role of the human factor is associated with decreasing of the initial probability of escalation of the problems and its timing.

The proposed method has the purpose to reduce the operating costs and to increase the efficiency of the customer service by developing of algorithms for distribution of the requests from the users and the monitoring, analysis and management of the quality of the work of the operators in the communication and information systems.

To follow-up this work, fault diagnosis and recovery processes should be observed and analysed in detail as they occur in practice. This would be valuable for the development of good work practices but also for the development of better network management.

#### REFERENCES

- [1]. Gans N., Koole G, Mandelbaum A. Telephone call centers: tutorial, review, and research prospects // *Manuf. & serv. operat. man.* - 2003. - P. 79-141
- [2]. Бельская Н.М. Исследование и разработка алгоритмов мониторинга и анализа качества работы операторов контакт-центра: дисс., ук. – 2012. – 16 с.
- [3]. Виноградов Н.А., Кармазин В.К. Статистические характеристики и математические модели центров обработки вызовов. // *Реєстрація, зберігання і оброб. даних.* 2002, Т. 4.: № 2, с. 40–50
- [4]. Leena Norros, Ilkka Norros, Marja Liinasuo, Kari Seppanen, Impact of human operators on communication network dependability, ECCE '11 Proceedings of the 29th Annual European Conference on Cognitive Ergonomics Pages 35-42
- [5]. Marja Liinasuo, Iina Aaltonen, Hannu Karvonen, Human Operator Perspective to Autonomic Network Management, ACHI 2012 : The Fifth International Conference on Advances in Computer-Human Interactions, 128-134
- [6]. Schulman PR, Roe E, van Eeten MJG, de Bruijne M (2004) High reliability and the management of critical infrastructures. *J Conting Crisis Manage* 12(1):263–280
- [7]. Лавров Е., Паско Н., Подход к формализованному описанию дискретной деятельности в системах "человек-техника-среда" // *Вісник Сумського державного університету. Сер. Технічні науки.* 2012, № 3. с. 55-67
- [8]. Su R, Yurcik W (2005) A survey and comparison of human monitoring of complex networks: University of Illinois Urbana-Champaign, NationalCenter for Computing Applications
- [9]. Boring Ronald L., David I. Gertman, Human Reliability Analysis for Computerized Procedures, Part Two: Applicability of Current Methods, Human Factors and Ergonomics Society, 2006
- [10]. Андонов А., Хубенова З., Функционална устойчивост на информационно-управляващи комплекси в критични приложения, монография, София, 2011
- [11]. Kraemer S., and Carayon, P. (2007). Human errors and violations in computer and information security: The viewpoint of network administrators and security specialists. *Applied Ergonomics*, 143-154.
- [12]. Salvendy G. (Editor), Handbook of Human Factors and Ergonomics, Purdue University, 2006, Canada
- [13]. Yamamura T, Yata K, Yasushi T, Yamaguchi Y (1989) A basic study on human error in communication network operation. *IEEE Xplore*, Downloaded 23 June 2009
- [14]. Крылов В.В., Самохвалова С.С. Теория телетрафика и её приложения. – СПб, 2005.
- [15]. Noah Gans, Ger Koole, and Avishai Mandelbaum. Telephone call centers: Tutorial, 2003, review, and research prospects. *Manufacturing and Service Operations Management*: p79-141.

# Software development of Content Management System in Online-CADCOM platform

Ognyan Chikov<sup>1</sup>

**Abstract** – In this paper the software development of Content Management System in Online-CADCOM platform is described. The two building subsystems – Content Management Application (CMA) and Content Delivery Application (CDA) and the used MVC approach are presented as well. Their role and place in the platform are given. Elements of code and "Dashboard" responsive view from smart device are presented as illustrations.

**Keywords** – Content Management System, MVC approach, Software development, Front-end, Back-end, E-content

## I. INTRODUCTION

Online-CADCOM is an online assisted platform for computer-aided design in communications [1]. It is built of large number of instruments, resources and links to already studied and estimated tools and would be very difficult to manage all these materials without a reliable method. This method is by using separate platform that can manage all of the content which is available and accessible through Online-CADCOM – this platform is called Content Management System. Content Management Systems can be developed as using procedural or object-oriented programming and also using MVC Framework. Online-CADCOM's Content Management System (CMS) is developed using MVC Framework approach also and the advantages that this method has, makes the platform a reliable method for manage the content of large web platforms. Additional plugin – jQuery is used for event handling optimizations and speed up loading of first pages in the management system. The platform allows many users to efficiently share e-content such as audio files, video files, pictures, graphics, knowledge materials and content in the different categories as described in [1]. The Content Management System controls the access to this content in the repository as described in [2]. It is used to dynamically change the content of the Online-CADCOM without using complicated program languages and upload files every-time to the web content server through file transfer protocol (FTP), as well as central platform software for maintenance and user management. Users are authorized through user-level security and redirected to the appropriate "index" page of the platform where it is located the main navigation through different categories and functionalities. The platform uses a relational database to store the page content, metadata and other useful information assets that are

needed for the system. System administration is done through browser-based interface and do not require the use of "fat" client. The browser-based interface is "responsive" and appropriate for using with all Operational Systems (OS) and standard resolutions and provides scalable vision and structure representation of the content. Using Online-CADCOM's Content Management System provide some advantages of implement the platform. One of the advantages that this platform gives is that the cost of developing the content is minimal. This means that future programmers work could only be needed for developing new functionalities and not for uploading content or expanding it. Another advantage of the Content Management System is that it makes Online-CADCOM easy to use and support.

First challenges of using Content Management System in Online-CADCOM are presented, then two separated parts that build up the Content Management System are described and different main building script modules are presented. Also different code controllers are presented.

## II. CONTENT MANAGEMENT IN ONLINE-CADCOM

Using Content Management System (CMS) in Online-CADCOM platform raises a lot of questions about this approach for managing the content. As described in [1], the Online-CADCOM platform will include a large number of user files that must be accessible 24/7. This means that the platform will have a lot of pages with different type of data [3]. The challenges for creating a Content Management Systems are related with the directory structure of the platform because of the large number of files. As described in [4] it is very important to properly scale the directory structure. As showed in [4] the non-CMS platform directory structures are ended with the name and type of file. The reason this is done is because there is no platform for managing the content and the user is directly connected with the desired file. If file is there, it is loaded, if not – there is an error. This approach is easy to implement and simple but not suitable for our case. That why as described in [4] in Online-CADCOM is used a directory structure approach that is implemented with CMS. Because the chosen approach is database-backed, it is not necessary to have a separate physical source file for every page. This approach solves the challenge related with the large number of e-content and materials in Online-CADCOM. Furthermore, the Online-CADCOM Content Management System does not write a physical file when a new page in the administration is done, it uses a "controller" to read from a database, based on what the URL was. This reduces the need for the CMS to have write-permissions for the publicly visible parts of the platform,

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

which increase the security as described in [5] and solve the security challenge.

Database structure is another challenge facing the Content Management System. For optimization purposes, Online-CADCOM should make as few database queries as possible. As described in [5] database of Content Management System is optimized using the following techniques:

- Proper indexing strategy
- Proper database schema
- Using external cache

As described in [2], the Content Management System is built of two separated parts that works interconnected:

1. Online-CADCOM Content Management Application (CMA) - it is a front-end user interface that allows all the users of the system to add, modify and remove content from a web site without the intervention of a programmer. This part of the platform is the part that users can see and operate with it. This is the front-end, the design of the visual part that contain all the categories, menus, buttons and forms for filling. Also this part includes all font, color and other visualization documents and scripts.

2. Online-CADCOM Content Delivery Application (CDA) - it's a powerful program script (backend) that compiles all the information that is provided from the CMA piece of program and use it to update the content of the website platform. Here work all the scripts on low login level. For example, the main CRUD (Create, Read, Update, Delete) methods and etc.

### III. SOFTWARE DEVELOPMENT OF CONTENT MANAGEMENT APPLICATION (CMA)

The Content Management Application (CMA) is the front-end user interface of Content Management System (CMS). The idea of the Content Management Application is to take in hands the user interactions in the platform and later in the Content Delivery Application (CDA) - to perform this interaction. The front-end interface of the Content Management System is created with following rules as described in [6]:

- It is developed for multiple browser access – browsers are supposed to render webpages in the same way, but they don't and that is the reason why the code is developed to look correctly on multiple browsers
- The design is responsive – the CMS platform will be accessed from multiple and variety of devices – from smartphones and smart gadgets to notebook and desktop computers. Using CSS media queries, the management platform is with implemented responsive design.
- Colors are clean and simple – colors chosen for the design are simple and intuitive.
- Navigation is intuitive – Big problem for users operating with new platforms are to navigate for the first couple of times through the content. The CMS platform is made with big, simple and useful navigation menu and elements (buttons).

One of the most important elements of Content Management Application (CMA) is the “Dashboard”. The Dashboard is the first page that the user sees after successful login in the administration platform and gives immediately information about the productivity of Online-CADCOM. It is developed using the methods listed above for multiple browser access, responsiveness, simplicity, intuition navigation and user sight visibility. Figure 1 represents where the dashboard is located in the system architecture.

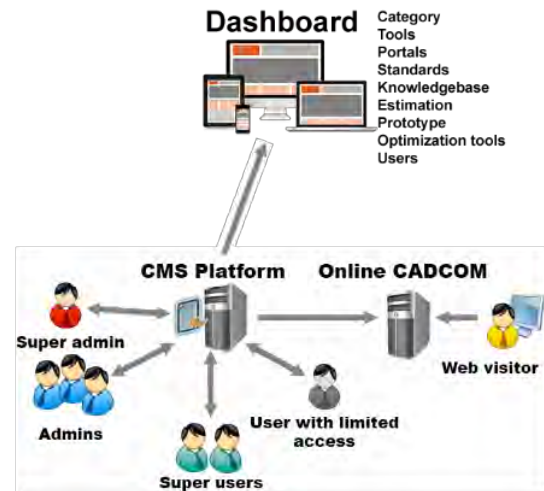


Fig. 1. Overview of development approach

As described in [3], Online-CADCOM has three main panels with content – Panel 1, Panel 2 and Panel 3. As Panel 1 and Panel 2 cover the Daniel-Gajsky Y model, these panels can contain more than one subcategory and in the same time these subcategories can contain more than one subcategories containing CAD tools. This statement makes no sense to hardcode the categories in the Panel 1 and Panel 2 and makes them static and impossible for changing. That's why the approach used in creating the Dashboard is to provide a dynamic method to add unlimited numbers of categories that contain subcategories and also unlimited number of tools that are stored in these different categories or/and subcategories. Dashboard manages all the related content in the platform and has additional menus:

- Portals,
- Standards,
- Knowledge Base,
- Estimation,
- Prototype,
- Optimization Tools,
- Users,
- Current User profile.

Figure 2 represents part of the responsive view of “Dashboard” loaded on a smart device.



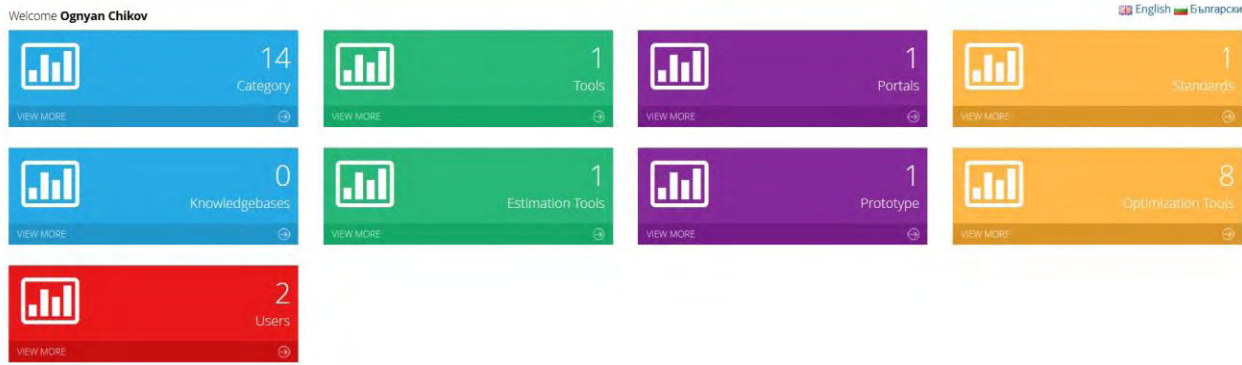


Fig. 2. “Dashboard” responsive view from smart device

#### IV. SOFTWARE DEVELOPMENT OF CONTENT DELIVERY APPLICATION (CDA)

Content Delivery Application is consisting of a large number of script files, compiled together which make possible all user interactions to take in use right after they were accomplished. The program code is called and changes are visible to the person who is editing the content and of course for the end user of Online-CADCOM. Used method and approach for CMS is highly complicated and reliable for future development and support. This is true because as described in [2] the approach that is used to build up both Online-CADCOM and CMS is with MVC Framework. The compiled platform is made of a hundreds of files, but one of the most important and major file that takes care for navigation through the “routes” and the “methods” of the platform is “routes.php” file [7]. Its content can be divided logically on two parts:

- Routes that are related with the CMS platform – these routes take care about the processes that are in the Content Management System (Admin routes)
- Routes that are related with the front-end part of Online-CADCOM – these routes take care about the processes that are in Online-CADCOM (Front-end routes)

Figure 3 represents how “routes.php” looks. File “routes.php” contains all the routes and controllers in the platform and this approach is covered in [7].

First step to reach the CMS platform is to request the URL “cadcom.demo/login”. When a user request the URL for the login screen, the browser send request for “cadcom.demo/login” to the file “routes.php”. Here if the “login” is not described, then the browser will return an error page. This is the reason why all the “URLs” are strictly described in the file showed on Figure 3 and they have a special function – to call a specific controller according to their functions.

```
Route::get('login','AdminController@showLogin');
Route::post('login', array('uses' => 'AdminController@doLogin'));
Route::get('logout', array('uses' => 'AdminController@doLogout'));
Route::get('registration', array('uses' => 'AdminController@showRegistration'));
Route::post('registration', array('uses' => 'AdminController@doRegistration'));
Route::get('forgot-password', array('uses' => 'AdminController@showForgotPassword'));
Route::post('forgot-password', array('uses' => 'AdminController@doForgotPassword'));

/*
|-----
| Front end - routes
|-----
*/

Route::get('/', 'HomeController@index');
Route::get('portals', 'FEPortalsController@index');
Route::get('portals/{id}', 'FEPortalsController@show');
Route::get('standards', 'FEStandardsController@index');

Route::get('knowledgebase', 'FEKnowledgebaseController@index');
Route::get('ajax-subcat', 'FEKnowledgebaseController@getRecords');

Route::get('economical-estimation', 'FEEconomicalEstimationController@index');
Route::get('proto-type', 'FEPrototypeController@index');
Route::get('optimization-tools', 'FEOptimizationToolsController@index');
Route::get('contact', 'FEContactsController@index');
Route::post('contact', 'FEContactsController@store');

/*
|-----
| Admin routes
|-----
*/
Route::group(['before' => 'auth:admin'], function()
{
    Route::group(array('prefix' => 'admin'), function()
    {
        Route::get('dashboard', 'AdminDashboardController@getIndex');
        Route::resource('users', 'AdminUserController');
        Route::resource('category', 'AdminCategoryController');
        Route::resource('tools', 'AdminToolsController');
        Route::resource('portals', 'AdminPortalsController');
        Route::resource('standards', 'AdminStandardsController');
        Route::resource('knowledgebase', 'AdminKnowledgebaseController');
        Route::resource('economical-estimation', 'AdminEconomicalEstimationController');
        Route::resource('proto-type', 'AdminPrototypeController');
        Route::resource('optimization', 'AdminOptimizationToolsController');
    });
});
```

Fig. 3. Overview of development approach

With the given example above, when a user tries to reach the login page of CMS platform, the following route is checked and performed:

*Route::get('login','AdminController@showLogin');*

Route method that is used is “get”. The “get” method send the information together with the URL to the controller “AdminController” and calling function “showLogin”. In this controller is built-in the logic for user’s login. Below is shown how user’s input information is taken and used for the login function:

```
public function doLogin()
{
    $rules = array(
```

```

        'username' => 'required',
        'password' => 'required|min:4'
    );
    $validator = Validator::make(Input::all(),
$rules);
    if ($validator->fails()) {
        return Redirect::to('login')
            ->withErrors($validator)
            ->withInput(Input::except('password'));
    } else {

        $userdata = array(
            'username' => Input::get('username'),
            'password' => Input::get('password')
        );

        if (Auth::attempt($userdata)) {
            return Redirect::to('admin/dashboard');
        } else {
            return Redirect::to('login')-
>with('login_error',
trans('common.incorect_user_pass'))
            ->withInput();
        }
    }
}

```

“AdminController” is the most important controller in the platform. It contains the following methods in it doing these major functions:

- showLogin() –method which loads the “view” of the login form
- doLogin() – method which process the data from the login form
- doLogout() – method which destroys the current session of the user and log him out of the system
- showRegistrationForm() – method which loads user registration form
- doRegistration() – method that process the registration of the new user
- showForgotPassword() – method that loads a form for resetting the password
- doForgotPassword() – method that process the data and if it is valid send an email with a new password.

Second step to reach the CMS platform is to load the “Dashboard”. In the showed code above, if the variable “\$userdata” return true, then the URL is redirected to “admin/dashboard” and again this URL is processed from the file “routes.php”. Again the requested URL is checked in it and it calls the line:

```
Route::get('dashboard','AdminDashboardController@getIn
dex');
```

In this controller the main methods that are used to perform showing and changing of information are:

- Index() - Display a listing of the resource.
- Create() - Show the form for creating a new resource.

- Store() - Store a newly created resource in storage.
- Show() - Display the specified resource.
- Edit() - Show the form for editing the specified resource.
- Update() - Update the specified resource in storage.
- Destroy() - Remove the specified resource from storage.

## V. CONCLUSION

Online-CADCOM together with Content Management System forms a complicated software application that works on the web and that is reachable from everywhere and from any device with Internet connection. This makes the application with middle to high loading. The showed controllers and the methods that work inside are proof of quality and reliability of the code that ensures stable operation of the platform and avoids vulnerability. The used MVC Framework approach ensures easy maintenance of the code, good platform structure, lack of security bugs and easy traceability of platform logic. The described method for developing of scripts is reliable and highly effective for this type of web applications. The Content Management System is implemented for content development and task solution which will be presented in separate paper.

## ACKNOWLEDGEMENT

This research is partly supported by Project №152PD0039-07 in Technical University – Sofia.

## REFERENCES

- [1] G. Marinova, V. Guliashki, O. Chikov, “Concept of Online Assisted Platform for Technologies and Management in Communications – OPTIMEK”, Int.l Conference in Computer Science, Information System and Telecommunication, ICCSIST 2014, 7-9 November 2014, Durres, Albania, 2014, pp.55-62
- [2] O. Chikov, G. Marinova, MVC Framework Approach in the Online assisted platform form Computer-aided Design in Communications – Online-CADCOM, Computer and Communications Engineering, Submitted.
- [3] G. Marinova, O. Chikov, Methodology for tools integration in the Online Assisted Platform for Computer-aided design in communications, Proc. of papers of L International scientific conference on information, communication and energy system and technologies, ICEST’2015, 24-26 June 2015, Sofia, pp.31-36, ISBN:978-619-167-182-3
- [4] CMS Design Using PHP and jQuery, Kae Verens, Packt Publishing, 2010, ISBN: 978-1-849512-52-7
- [5] PHP5 CMS Framework Development, Martin Brampton, Packt Publishing, 2008, ISBN: 978-1-847193-57-7
- [6] V. Nikolov, D. Pilev, An Application of Cocoa Bindings and MVC Technology by Designing of a User Interface with Data Bases, Computer & Communications Engineering, 2/2013, pp.49-54, ISSN: 1314-2291
- [7] Web Content Management: System, Features and Best Practices – 1<sup>st</sup> Edition, Deane Barker, O’Reilly, 2015, ISBN: 978-1491908129



# Investigation of algorithms for VoIP signaling

Todorka Georgieva<sup>1</sup>, Ekaterina Dimitrova<sup>2</sup> and Slava Yordanova<sup>3</sup>

**Abstract** – This paper discusses issues regarding the operation of algorithms for VoIP Signalling. The resulting data includes jitter, packet delays, delays related to signalization, distribution of packets in protocols and relative assessments of QoS and MOS.

**Keywords** – Codec, Packet, Signalization, VoIP.

## I. INTRODUCTION

The main objective of this paper is about solving problems of the implementation of voice applications over wireless networks with packet switching, the ability to overcome them and to optimize this type of communication. The Extended IP structures, including wireless, provide large opportunities for the development of VoIP technologies. The paper concerns the methods to improve the quality of service, which most closely affect the end user. Concrete results are obtained proving the future application of these technologies.

Studies conducted show that the problems associated with the range of wireless networks and their security cannot be a real obstacle to the implementation of Vo802.11.

The thesis which is propounded, demonstrates the broad possibilities for the implementation of VoIP over 802.11 and the provision of new services to existing ones [1]. One of the main challenges in the wireless delivery of data and Internet is the providing of a quality service. QoS is associated also with the ability of service providers to adapt their networks for transmission of voice (Fig. 1).

To properly analyze the possibilities of using Vo802.11, as an alternative voice communication, it is necessary to analyze potential weaknesses of the network to be minimized or eliminated. Largely functions and applications of network depend on signalization. With the use of modern VoIP protocols and good programming of new functions, variety of applications, satisfying the end user are achieved.

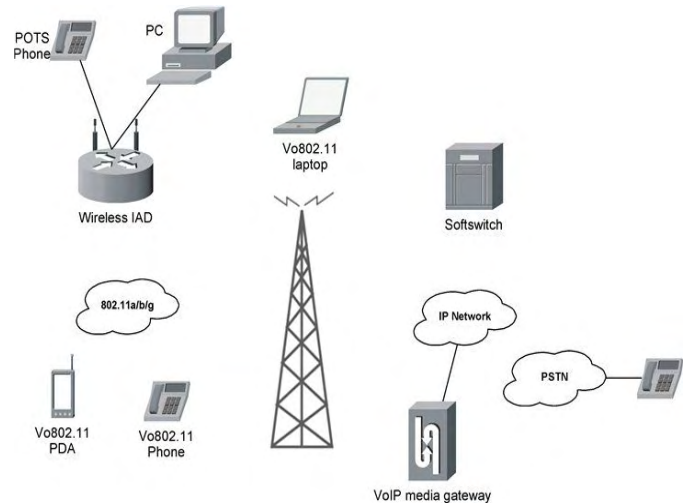


Fig. 1. Broadband network for speech transmission

## II. EXPERIMENTAL RESULTS

To provide quality speech comparable to the PSTN, it is necessary to minimize delays, jitter and packet loss in Vo802.11 networks. Regulating QoS to the wireless broadband network is accompanied by additional network requirements.

To measure the voice quality currently available two tests to ensure metric: MOS (mean opinion score) and PSQM (perceptual speech quality measurement) (Fig. 2).

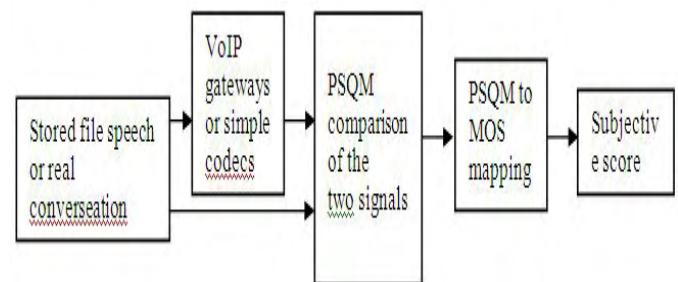


Fig. 2. A Process of PSQM [2]

<sup>1</sup>Assist. Prof. PhD. Eng. Todorka Georgieva, Telecommunication Department, Technical University of Varna, Studentska 1, 9010 Varna, Bulgaria, E-mail: [tedi\\_ng@mail.bg](mailto:tedi_ng@mail.bg)

<sup>2</sup>Assoc. Prof. PhD. Eng. Ekaterina Dimitrova, Department of Electronics and Microelectronics, Technical University of Varna, Studentska 1, Bulgaria, E-mail: [edimitrova55@gmail.com](mailto:edimitrova55@gmail.com)

<sup>3</sup>Assoc. Prof. PhD. Eng. Slava Yordanova, Department of Computer Science and Engineering, Technical University of Varna, Studentska 1, Bulgaria, E-mail: [slava\\_y@abv.com](mailto:slava_y@abv.com)

Investigation on IP traffic and signaling tests which were made, were carried out under the scheme of Fig. 3.

Two simultaneous calls from endpoint 192.168.1.3 to 192.168.1.4 were utilized and the results were seen within two minutes.

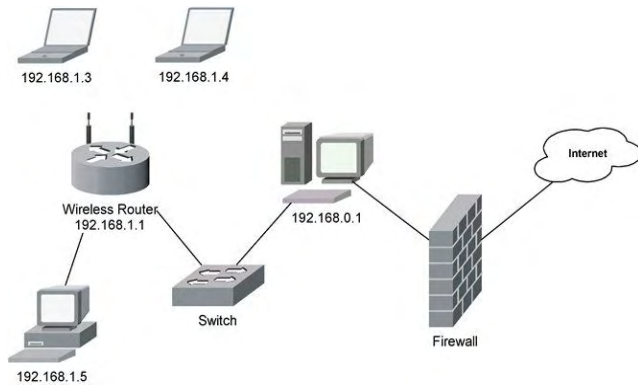


Fig. 3. Network configuration

The minimum, maximum and average values of some of the most important parameters such as jitter, packet intervals and used bandwidth are given in Table I.

TABLE I  
AUDIO DETAILS

Metric	Average	Low	High
Src Jitter (ms)	0,712	0,007	3,273
Dest Jitter (ms)	0,911	0,005	2,735
Src Packet Interval (ms)	2,431	0,009	45,001
Dest Packet Interval (ms)	2,561	0,110	33,792
Src Bandwidth (kb/s)	127,691	127,421	133,333
Dest Bandwidth (kb/s)	127,654	127,405	137,143

The addresses of the source and the recipient along with their ports and the type of used media are given in Table II.

TABLE II  
TYPE OF USED MEDIA

Parameter	Node A	Node B
Address	192.168.1.3	192.168.1.4
Port	40000	40000
Media Type	G.711 Alaw	G.711 Alaw
SSRC	00294823	4AE13D6C
Audio/Package (ms)	3	3
Frames/Package	3	3
Total Packages	79 834	79 814
Packages Lost	0	0
Longest Package Loss Burst	20	13
Total Payload Bytes	1 916 016	1 915 536

The results in Table III refer to indicators of monitored calls [3].

RFactor / MOS Score are characteristics of the speech tract for each individual media stream and are assessed subjective. The software used in the research is processing by models including time-varying characteristics such as: the sequence of lost packets, discarded packets due to high jitter and others.

Burst Packet Loss Rate gives the number of the lost packets in percentages in case of unusual circumstances.

Average Burst Length shows the average duration in milliseconds of the unusual circumstances that would lead to packet loss.

Average Gap Length is the average length of the intervals in milliseconds in the event of unusual circumstances in the respective media streams.

Good Packages is the number of received packets at the two endpoints.

Lost Packages is the number of lost packets in the network for both endpoints.

Discarded Packages are the rejected packages when the delay is exceeded or earlier arrival at the source or recipient is happened.

Media Type shows the type of used media format.

TABLE III  
AUDIO QoS

Metric	Source	Destination
Listening R Factor	93	93
Conversational R Factor	93	93
Listening MOS Score	4,195	4,195
Conversational MOS Score	4,195	4,195
P.862 Raw MOS Score	4,449	4,449
Burst Packet Lost Rate (%)	0,000%	0,000%
Average Burst Length (ms)	0	0
Average Gap Length (ms)	28411	48546
Good Packages	40005	40010
Lost Packages	0	1
Discarded Packages	2	0
Media Type	G.711 Alaw	G.711 Alaw

A graphic expression of the quality of the audio signal is shown in Fig.4. QoS reports are given, as a result of analysis of VoIP calls over 30 seconds.

The parameters of the audio stream in graphical form are presented in Fig.5. These are number of packets per second, jitter in the audio channel, packet interval of audio packages and the speed used in kbps.

The ranges of delays of the individual packets during transmission of G.711 audio stream have network delay below 10ms, and the greatest delays do not exceed 60-70ms with limit values 150ms in one direction (Fig. 6).

The majority of packages have jitter below 3ms, while the largest values do not exceed 8-10ms, which again is in the norm of 20ms (Fig.7).

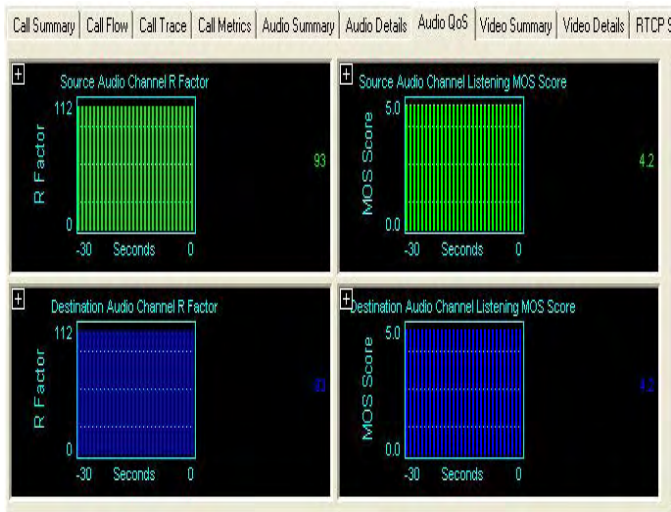


Fig. 4. Audio QoS

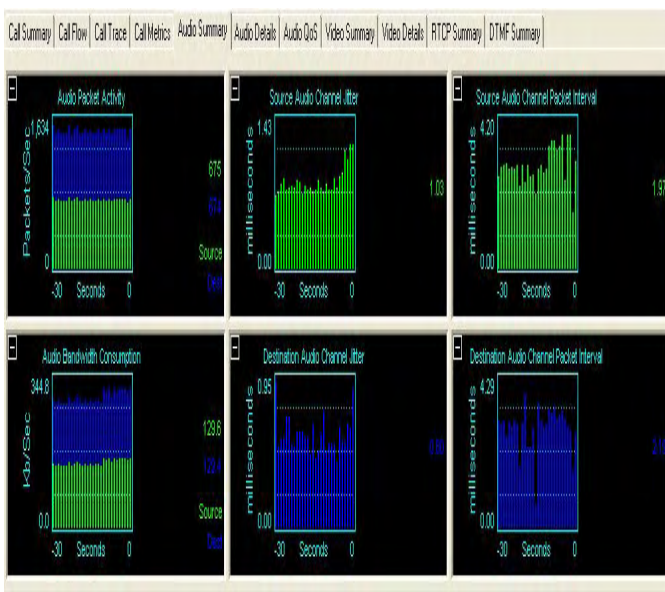


Fig. 5. Graphic form of the audio stream parameters

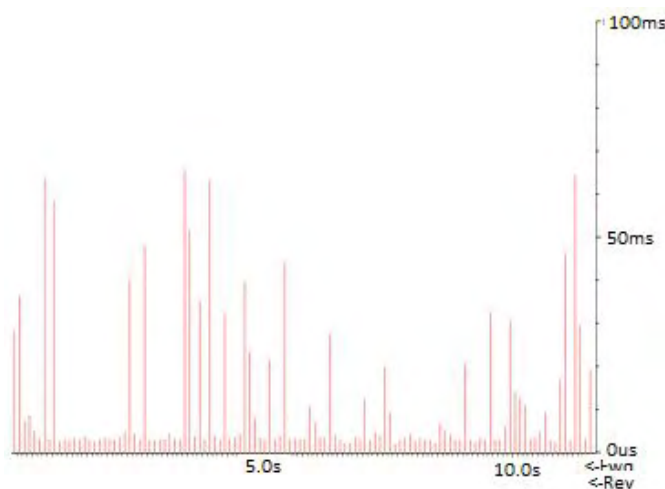


Fig. 6. Delays packages using G.711

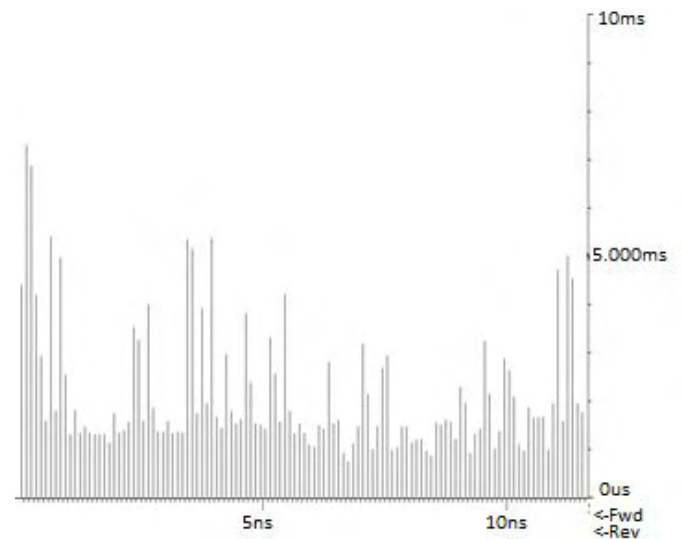


Fig. 7. Jitter of the individual packages

G.722.1: The coding system uses subband adaptive differential pulse-code modulation (SB-ADPCM) at a speed of 64kbps. The system has three modes of operation for 7 kHz audio encoding, corresponding to the bit rate which is 64 or 56, or 48kbps. For the study a speed mode 64kbps is used.

G.723.1 6,3kbps: This recommendation of the ITU, has two standard speeds - 5, 3 and 6,3kbps. The higher speed, which is used in the study, gives better quality. Дължината The length of the voice frame is 30ms and further prediction of 7,5ms allows algorithmic delay of 37,5ms. Encoder is designed to compress a voice with good quality with reduced complexity and it is not suitable for music [4].

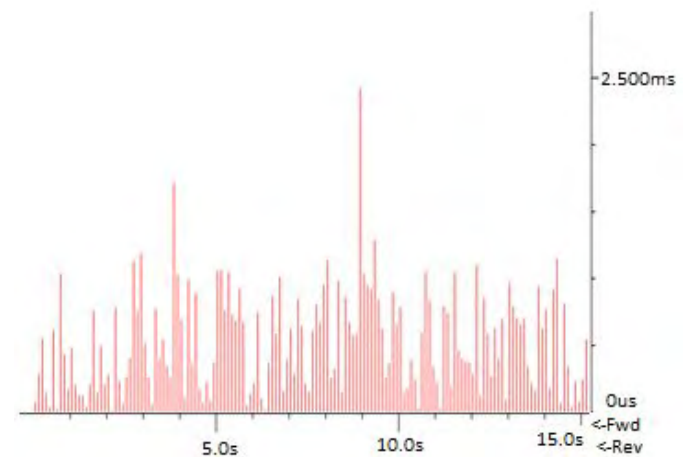


Fig. 8. Delays of packets in G.723.1 6,3kbps

To assess the MOS G.711 a study was conducted with SIP signaling on a heavily loaded network (Fig. 9). Despite the large number of rejected packets and large network delay, MOS assessments for forward and reverse direction of the conversation are fully comparable with the requirements.

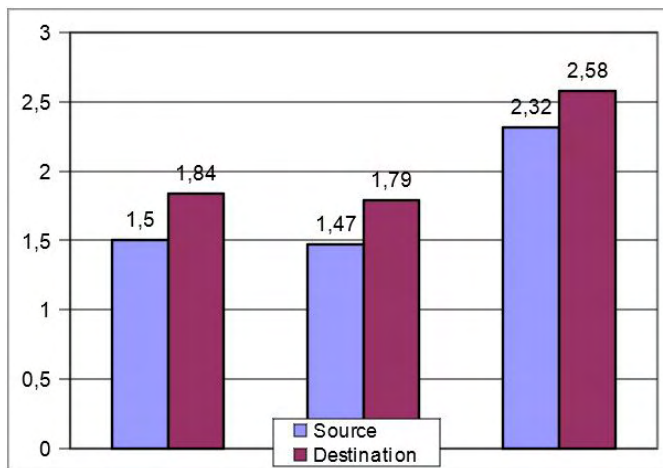


Fig. 9. MOS results in loaded wireless network

Data on network delays of packets, jitter, packet intervals were obtained for real SIP call. The calls have good quality – understandable, without echo and without noise.

### III. CONCLUSION

The testing of the realised Vo802.11 network shows its ability to compete PSTN under comparable conditions: the use of uncompressed voice and the appropriate for the conditions protocols. The possibility of using compressed formats of the media stream while maintaining the best operational parameters gives sizable potential of the system and makes it flexible for network designers and advanced signaling provide high intelligence of the entire network. These factors make it possible to quickly build stable and reliable systems providing quality services to the end users.

In the process it became clear that the disadvantages associated with the range of wireless networks and their security cannot be a real obstacle to the implementation of Vo802.11 voice transmission in real time.

The main contradictions that still remain to the fore are those related to QoS. To show in practice, in real conditions, the behavior of a system Vo802.11, a wireless tract was examined for two major signaling protocol SIP and H.323, and for the only possible at the moment transport protocol - RTP. The obtained results and the comparisons are an evidence of work together of SIP and H.323 - SIP realizing a signaling over packet switched networks, and H.323 building interface via gateways to the PSTN.

Taking into account the obtained data, concerning different aspects of the behavior of the investigated system, and recognizing that they are not used special measures to provide better QoS, such as RSVP, DiffServ, then the results obtained can guarantee the operability of Vo802.11 for voice communications.

### REFERENCES

- [1] "Switching to VoIP", Theodore Wallingford, O'Reilly 2005
- [2] Quality of Service (QoS) in the New Public Network Architecture", A. Agarwal, IEEE Canadian Review
- [3] Resource Reservation Protocol", Cisco Systems, <http://www.cisco.com>
- [4] "Softswitch: Architecture for VoIP", F. Ohrtman, McGraw-Hill



# Contemporary Trends for Increasing the Reliability of the SCADA systems Communication Level

Emiliya Dimitrova<sup>1</sup> and Vasil Dimitrov<sup>2</sup>

**Abstract** – In this paper the benefits of the implementation of the open transport networks in the communication level of SCADA systems are given. OTN ensures absolute reliability. Its main features are simplicity, openness, easy integration, ability to cover very large distances, as well as the future-proof design.

**Keywords** – SCADA (Supervisory Control and Data Acquisition systems), Optical Open Transport Network (OTN).

## I. INTRODUCTION

SCADA systems are highly distributed systems used to control of geographically dispersed objects, often scattered over thousands of square kilometers, where centralized data acquisition and control are critical to system operation. They are used in distribution systems such as electrical power grids (production and distribution of electricity), railway transportation systems, water distribution systems, oil and natural gas pipelines, etc. The purpose of SCADA systems is centralized data acquisition, information processing and visualization and generation of managerial impacts. The centralized monitoring and control of many non-concentrated subobjects (including monitoring alarms and processing status data) is performing in the Central Dispatcher Post (CDP) by long-distance communications networks. Based on information received from remote stations, automated or operator-driven supervisory commands can be pushed to remote station control devices, which are often referred to as field devices. They control local operations such as opening and closing breakers and valves, collecting data from sensor systems, and monitoring the local environment for alarm conditions. The correct and reliable functioning of SCADA system is crucial for the functioning of the whole system [1].

## II. STRUCTURE OF THE SCADA SYSTEMS

Each SCADA system is built on three formal levels [1, 2]: Upper (Dispatching) level, Communication level and Lower (Object or Field) level (Fig. 1). It consists of both hardware and software. Typical hardware includes following elements:

- Master terminal unit (MTU) – server placed at the CDP. In addition to the servers, this level includes video-walls, Dispatchers working places, personal computers, printers and the data historian, which are all connected by a LAN.

<sup>1</sup>Dr. Emiliya Dimitrova is with the Faculty of Communications and Electrical Equipment in Transport at Todor Kableshkov University of Transport, 158 Geo Milev Str., Sofia, Bulgaria.

<sup>2</sup>Assoc. Prof. Dr. Vasil Dimitrov is with the Faculty of Communications and Electrical Equipment in Transport at Todor Kableshkov University of Transport, E-mail: vdimitroff@abv.bg

- Communications equipment (e.g., radio, telephone line, cable, or satellite) and access monitoring and surveillance systems (video cameras, sensors, and various types of identification systems).

- Geographically distributed field sites consisting of Remote terminal units (most often Programmable Logic Controllers - PLCs), sensors for measurement, actuators (i.e. breakers, switches, control valves, and motors) and Intelligent Electronic Devices (IED) – i.e. multifunction protection relays. PLC monitors sensors and controls actuators through appropriate algorithms.

The IEDs are “smart” sensors/actuators containing the intelligence required to acquire data, communicate to other devices, and perform local processing and control. Many IEDs combine analog input sensors, digital (relay) inputs, analog and digital outputs, low-level control capabilities, a communication system, and program memory in one device. The use of IEDs in SCADA systems allows for automatic control at the local level (IED act without direct instructions from the CDP). The local PLCs poll the IEDs to collect the data and pass it to the SCADA Server.

The communications hardware allows the transfer of information and data back and forth between the MTU and PLCs. The MTU stores and processes the information from objects, displays information to the monitors and video-walls, and may generate actions based upon detected events. The server is also responsible for centralized alarming, trend analyses, and reporting.

The software is programmed to tell the system what and when to monitor, what parameter ranges are acceptable, and what response to initiate when parameters change outside acceptable values.

The communication level consists of Ethernet network in the CDP, transmission medium between the CDP and object and Ethernet network in any object.

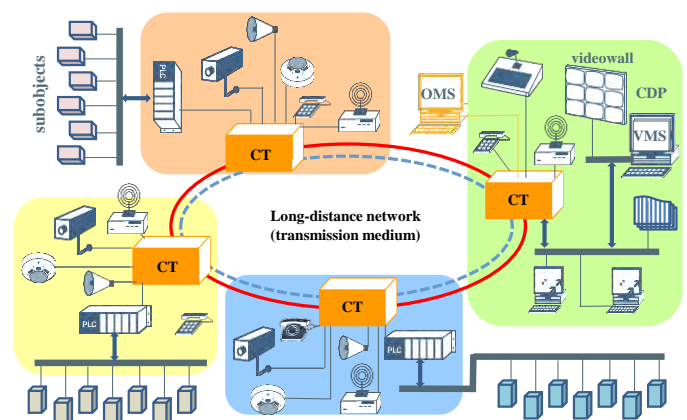


Fig. 1. Example of a SCADA system



There are different network characteristics for each layer within the system hierarchy. The fieldbus network links sensors and other devices to a PLC. Use of fieldbus technologies eliminates the need for point-to-point wiring between the controller and each device. The devices communicate with the fieldbus controller using a variety of protocols. The messages sent between the sensors and the controller uniquely identify each of the sensors. The control network connects the supervisory control level to lower-level control modules. The communications terminals (CT in Fig. 1) transfer messages between two networks. Common uses for CTs include connecting MTUs and RTUs to a long-distance network medium for SCADA communication. A firewall protects devices on a network by monitoring and controlling communication packets using predefined filtering policies. Modems are devices used to convert between serial digital data and a signal suitable for transmission over a telephone line to allow devices to communicate. Modems were often used in SCADA systems to enable long-distance serial communications between MTUs and remote field devices.

The communicational level for contemporary SCADA systems is constructed in compliance with the principle of the Optical Open Transport Network (OTN).

### III. SPECIFIC INDICATORS OF RELIABILITY OF SYSTEMS FOR INFORMATION TRANSMISSION

The reliability of the systems for transmission of information (including SCADA systems) has to be evaluated by specific indicators, in addition to the general technical reliability indicators. This is required by their structure and by the necessity of time reserve for processing and transfer of the information. These systems are often multifunctional and have complex hierarchical structure. This is why a big part of failures is partial and does not lead to the failure of the whole system, but only of part of it [2 – 5]. The presence of a time reserve leads to the fact that not all failures and accidents lead to failure of the functioning of the whole system, thus the term “reliability of operation” must be considered.

Probability of successful execution of the task can be used as an indicator of the system reliability if there is a reserve of time to process and transmit information. The execution of the task requires a certain time  $t_{task}$  within the acceptable range – the allowed time  $t_{all}$  (i.e.  $t_{task} \leq t_{all}$ ). There are two options for the system operation and the corresponding indicators of reliability of operation:

1) Failures of functioning do not occur in the following cases:

- There is no loss of information when system failure occurs, i.e. the system work continues after restoration in the failure point. The compliance with the condition  $t_{task} \leq t_{all}$  is needed for the successful solution of the problem.

- There is a complete loss of information when system failure occurs, i.e. the system work continues from beginning after restoration. In this case, a trouble-free operation for a time greater than  $t_{task}$  over time  $t_{all}$  is required.

Indicators of reliability are the probability of occurrence of these events.

2) Another situation arises when failures of system elements do not lead immediately to a failure of its functioning, and there is any particular or occasional time reserve for development of the failure. This kind of situations occurs in the following cases:

- Inertial objects are controlled (interruptions of the object control are allowed for a while). Indicator of reliability is the probability of execution of the control task.

- Not urgent messages are transmitted. Indicator of reliability is the cumulative time of the message retention because of a failure. This cumulative time depends on the recovery time and whether the channel is busy.

The time delay  $t_d$  because of one failure can be calculated by the equation:

$$t_d = \frac{k}{(1-k)} \frac{\tau_r}{t_m} (\tau_r - t_m), \quad (1)$$

where  $\tau_r$  is the recovery time,

$t_m$  – time for transmission of one message,

$k$  – part of time to occupy the communication channel for transmission of one message

In most cases the following presumption can be assumed:

$$t_m \ll \tau_r \quad (2)$$

Therefore:

$$t_d \sim \alpha \cdot \tau_r^2, \quad (3)$$

where

$$\alpha = \frac{k}{(1-k) \cdot t_m} \quad (4)$$

The recovery time generally is a random variable with mathematical expectation  $T_r$  and dispersion  $D_r$ . Then the cumulative time delay  $T_d$  because of one failure is equal to:

$$T_d = \alpha \cdot (T_r^2 + D_r). \quad (5)$$

The average cumulative time message delay  $T_{d-av}$  can be calculated by the formula:

$$T_{d-av} = \alpha \cdot \omega \cdot (T_r^2 + D_r). \quad (6)$$

where  $\omega$  is a parameter of the failures flow.

Under an exponential distribution of the recovery time  $D_r = T_r^2$ . Therefore

$$T_{d-av} = 2 \cdot \alpha \cdot \omega \cdot T_r^2. \quad (7)$$

The maximum value of the time delay assessment can be calculated by the equation (7), because the exponential distribution has the biggest dispersion.

The reliability of multifunctional information transmission systems can be assessed either by the execution of individual functions or by the whole system. There are two borderline cases:

1) The system provides performance of closely interconnected processes, then the failure of any one function leads to violation of all processes and the system reliability can be assessed as a standalone system.

2) The system performs quite poorly interconnected functions. If all the functions are equally important, reliability

can be estimated by mathematical expectation of the number of dysfunctions. A disadvantage of this indicator is that the simultaneous failure of several elements leads to disturbing the performance of several functions, which can lead to disproportionately heavier consequences.

The coefficient of keeping the effectiveness can be used as a general reliability indicator of multifunctional information transmission systems - it characterizes the impact of system reliability on its technical efficiency [4]. It is the ratio of the index of technical efficiency at real reliability to the maximum possible value of this indicator at absolute system reliability.

#### IV. INCREASING THE COMMUNICATION LEVEL RELIABILITY BY OTN

OTN is a digital communication network for real-time and mission-critical applications [2-7]. Basically, OTN is a communication backbone built up by nodes which are interconnected by a double optical fiber ring (primary and secondary rings). The transmission of signals in optical fiber instead of copper wire has several advantages: less weight and size, larger information capacity, absence of electrical connections and circuits, higher protection of the information transmitted. There is no electromagnetic interference, it is cheaper. The larger bandwidth allows transmitting all types of information on a single network. Typical communication requirements are speech (audio), computer data, LAN info, control information, video, etc. [6, 7]

OTN network architecture is based on four major system components: the fiber optic backbone, the OTN nodes, the interface cards that provide the user access to the system and the network management system (Network Control Center NCC or OTN Management System OMS). The configuration of the OTN node is influenced by the transmission medium, the distance between nodes and the topology. Different types of fiber can be used for OTN: multimode fibers 62.5/125  $\mu\text{m}$ , 50/125  $\mu\text{m}$  or 100/140  $\mu\text{m}$  or single mode fiber 9/125  $\mu\text{m}$ . Multimode fiber optic connectors are much cheaper (purchase and installation) than single mode ones. The available bandwidth  $\times$  length product of multimode fiber is more than sufficient for short to medium inter-node distances with many connectors on the way. The bandwidth version also influences the choice of the fiber type. For longer distances, the only option available is single mode fiber. Here, dispersion characteristics and the overall attenuation of the fiber are the dictating parameters.

The time required by a message to travel from the transmitter to the receiver is determined by:

a. The OTN basic delay (ring delay)  $T_{RDb}$ , i.e. the run-through times of the fiber, the TRMs (Transceiver Modules), and DRAs (Dual Ring Adapters) or BORAs (Broadband Optical Ring Adapters).

b. The delay on the interface cards themselves.

The OTN basic delay  $T_{RD}$  can be calculated as follows:

$$T_{RDb} = n \cdot t_N + L \cdot t_L, \quad (8)$$

where:  $n$  is the number of nodes in the ring,  
 $t_N$  – the node delay,

$L$  – fiber length in km (for a ring with loopback, also the part along the secondary ring must be considered),

$t_L = 5 \mu\text{s}/\text{km}$  – the run-through time of the fiber.

The node delay may adopt the following values:

$t_N = 2.2 \mu\text{s}$  for nodes OTN-2500,

$t_N = 2.6 \mu\text{s}$  for nodes OTN-600 and

$t_N = 2.9 \mu\text{s}$  for nodes OTN-150 at a ring with loopback and

$t_N = 1.2 \mu\text{s}$  for nodes OTN-2500,

$t_N = 1.6 \mu\text{s}$  for nodes OTN-600 and

$t_N = 1.9 \mu\text{s}$  for nodes OTN-150 at a ring without loopback.

Newer version OTN-X3M is based on "Node add/drop bandwidth" technology and the node delay time is less than defined above values.

Communication between OTN nodes on the ring is based on Time Division Multiplexing (TDM). This access scheme allows multiple users on the ring, sharing the transmission medium. With TDM, the time continuum is divided into repetitive periods called frames. These frames are subdivided into time slots. In OTN, all time slots are one bit long. Devices can communicate when they are allocated a number of time slots during which they can transmit signals. TDM-frames are circulating on both the primary and the secondary ring at an overall rate of 147.456 Mbps, 589.824 Mbps or 2488 Mbps. These frames have a length of 31.25  $\mu\text{s}$  and contain 4608 bits, 18432 bits or 73728 bits for OTN-150, OTN-600 and OTN-2500 respectively. "Channels" and "bits per frame" are equivalent for OTN.

The fact that always a whole number of frames circulates on the ring is not considered in the Eq. (8) due to the extra delay in the master node's DRA or BORA. Subsequently, the ring delay is always exceeding or equal to  $T_{RDb}$ , and equal to:

$$T_{RD} = 62.5 \mu\text{s} + m \cdot 31.25 \mu\text{s} \quad (9)$$

where  $m$  is the frame number (this a whole number exceeding or equal to 0). The complete ring delay  $T_{RD}$  is considered, where the complete and possibly looped-back ring is run through once, without any interface cards.

Comparing Eq. (9) and Eq. (7) it can be noted that OTN offers very high performance.

Because of the double ring structure, OTN guarantees an unparalleled degree of reliability. Should one fiber ring or node fails because of any reason (fire, rupture etc.), then the other ring and nodes take over immediately, thus keeping the whole network operational round the clock (Fig. 2).

The system will always find a way around any problem without affecting its users. The OTN nodes also offer redundancy of the main components such as power supplies, common logic cards and optical modules. Therefore, a section failure does not lead to system failure.

There are many SCADA systems in Bulgaria, which are functioning especially in very important spheres as transport and energetics. A contemporary and efficient SCADA-system is in operation in the metropolitan – Sofia and covers all of the technological processes [8, 9].

It consists of several independent systems, the most important of which are the system for operational dispatching on the train traffic DISIM-V and this one for monitoring and control on electrical equipment and traction substations DISIM-E.

## V. CONCLUSION

SCADA systems are usually designed to be fault-tolerant systems with significant redundancy built into the system architecture. The system reliability can be remarkably increased by implementing optical OTNs which transfer all available communication applications, data, voice, digital video etc. and provides Accident Prevention. The environment that is ideally suited for OTN is an environment characterized by a multitude of data/ LAN/ voice/ Video communication services, distributed throughout a large area. The high availability of the OTN system is a key asset. The higher the criteria (mixed/distance) for a particular application, the better the OTN system is suited

## REFERENCES

- [1] Stouffer K., J. Falco, K. Kent, „Guide to Supervisory Control and Data Acquisition (SCADA), and Industrial Control Systems Security”, NIST, Gaithersburg, 2011
- [2] Goranov, E., E. Dimitrova. *Remote Control and Management Systems in Transport*, Todor Kableshkov University of Transport, Sofia, 2010
- [3] Clarke G., D.Reynders, E.Wright, *Practical Modern SCADA Protocols*, Newness, ISBN: 978-0-7506-5799-0, 2004
- [4] Simankov V., P. Sundeev, System analysis of the functional stability of mission-critical information systems, Krasnodar, 2003
- [5] Sokolov B. The optimal structure reconfiguration control in an intelligent system. Int. Sc. Conf., Pskov, 2008
- [6] OTN systems – structure, features, advantages, OTN Systems, 2011, available at: <http://www.otnsystems.com/>
- [7] OTN Systems – Dedicated Networks for Critical Environments. Available on <http://www.otn.be/>
- [8] Technical documentation of SCADA-system DISIM-V, 2012
- [9] Technical documentation of SCADA-system DISIM-E, 2012
- [10] Plovdiv-Svilengrad Railway Electrification & Upgrading of Corridors IV & IX Signalling, Telecommunication and SCADA Systems for the whole line, Thales–Alcatel-Lucent, Sofia, 2011
- [11] Martin, C., Lalanda, P., Donsez, D.: A MDE Approach for Power Distribution Service Development. 3rd International Conference on Service Oriented Computing, LNCS 3826, 2005
- [12] System automation in substation, available at: <http://www.roconbg.com/en/engineering-services/system-automation-in-substation-12>

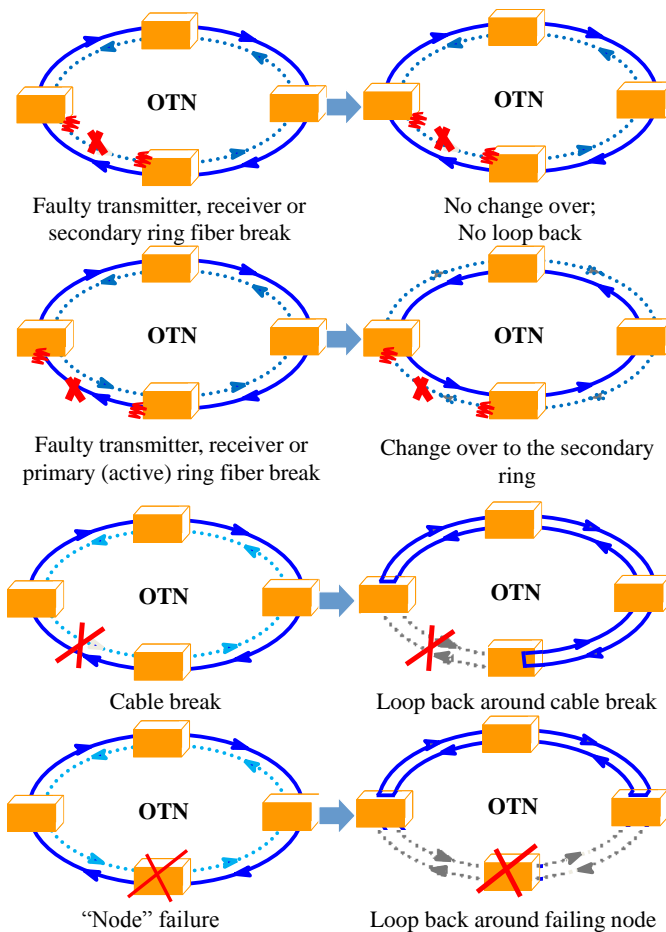


Fig. 2. Reconfiguration

The communication level on the first metro line comprises a pair of modems, 4-wire line connectivity and an expandable system for serial communication ports (Equinox), which supports connection to a central computer in the CDP through a standard protocol and to the devices on the physical level through asynchronous protocol for data transmission.

The communication level on the newer second line is based on the optical OTN.

The newest SCADA systems in Bulgarian railway transport are also constructed in compliance with the principle of the Optical OTN [10]. They have been designed and constructed in Plovdiv-Svilengrad section of European Corridors IV & IX.

OTN offers the best opportunity for SCADA-systems in electric distribution and transmission networks. In view of the existing different kind of grid infrastructures, there are several innovative concepts for substation automation, which allows consumers to swiftly adapt the electric power system to the new conditions at present and in the future [11, 12].

# Transmission of multimedia data in Wireless Sensor Networks

Miloš Kosanovic<sup>1</sup> and Mirko Kosanovic<sup>2</sup>

**Abstract** – In recent years, Wireless Sensor Networks (WSN) has become the most efficient tool for solving many humanity problems. The increasing development of Micro Electro-Mechanical Systems (MEMS) and the availability of low-cost hardware, such as CMOS cameras and microphones, has fostered the development one other kind of WSN, named Wireless Multimedia Sensor Networks (WMSNs). However, the need to transfer large amounts of data from sensor nodes to the primary base station, due to limited bandwidth and limited power consumption in multi hop wireless communication, is a major problem to be solved. In this paper, we give a comprehensive description of the problem, principles and standards that are important for transmission of multimedia information in applications such as real-time multimedia streaming, video conferencing, conversational services (video-conferencing), surveillance, overhear reality etc. The paper presents analysis which indicates how savings in the amount of data, that is sent, affects to the lifetime of sensor node and therefore the entire lifetime of WSN.

**Keywords** – wireless sensor node, multimedia data, efficient transmission, packets

## I. INTRODUCTION

Wireless Sensor Networks (WSNs) has become a big challenging issue for researcher community in the last few years. They represents one kind of wireless network composed of small sensor nodes (SNs) with limited computer resources, which are capable for gathering, data processing and communicating among each other completely independently [1]. WSNs are cost effective and distributed solutions implemented in various environments where conventional networks are impractical. WSNs have currently deployed in a wide range of applications as a sensor is becoming smaller and production cost is smaller. On the other side, the increasing development of Micro Electro-Mechanical Systems (MEMS), the availability of low-cost hardware, such as CMOS cameras and microphones, has fostered the development the other kind of WSN, named Wireless Multimedia Sensor Networks (WMSNs). They are capable to sample the physical environments at unprecedented spatiotemporal resolutions. Because of that, they can recorded many multimedia information such as images, video and audio streams from the environment, which allows their applicability into many *data-intensive applications*. All of this

information requires an enormous amount of data to be sent through the network. Unlike traditional wired networks, the reliability of data transfer in WSNs is not a critical parameter. Namely, for most WSNs (monitoring environment, object tracking, etc.) some occasional amount of data losses is acceptable. However, as WSNs become ubiquitously deployed, such as multimedia and process control WSNs, reliable data transfer, RDT, is required. In this kind of applications every byte of the data packet has to be, reliable delivered to the destination. On the other side, the need to transport the massive amount of data generated by sensor nodes to base stations poses a fundamental challenge due to the limited node bandwidth and high power consumption of multi-hop wireless topologies. This implies that more conflict in traffic, more congestion in accesses, and more packet retransmissions among SNs are needed. As a result, the SN's power consumption increases, while its lifetime period becomes shorter.

In this paper, we look at the problem of efficient and reliable data transport in WMSN. We propose a new solution for information delivery in both directions – upstream (from sensor node to master node – sink) and downstream (form master node to sensor node). Our solution ensures reliable data transport in both directions with much better relation between payload data and header (reduced header) in spite of well known current solutions [2]. In the remainder of this paper, in section 2 we explain some important features that characterize data transmission in WMSM. The third section gives a suggestion for new solution intended for reducing the number of bytes in the header of the transport protocol. Analysis of its success from the standpoint of increased lifetime of WSN is given in the fourth section. The fifth section concludes this paper.

## II. CHARACTERISTICS OF MULTIMEDIA DATA TRANSMISSION

Due to the large differences in data transmission, a large number of network protocols developed for the classical WSN are not usable or require specific changes in WMSN [3]. The main difference lies in the fact that the majority of protocols developed for WSN focused on reducing energy consumption (low power) of SN in the network, because the flow of data in such networks is very low. Unlike them, WMSN protocols, in addition to the conditions (energy-efficiency), are faced with two new additional problems to solve: significantly increased the intensity of traffic and the need for greater communication quality (QoS). Usually wireless environment is error prone environment, what means that is an environment with a large percentage of errors in transmission (40%-70%) [4]. Therefore a reliable and efficient transmission of such data is

<sup>1</sup>Milos M. Kosanovic is with the College of Applied Technical Science Nis, Aleksandra Medvedeva 20, 18000 Nis, Serbia, E-mail: [milos.kosanovic@vtsnis.edu.rs](mailto:milos.kosanovic@vtsnis.edu.rs).

<sup>2</sup>Mirko R. Kosanovic is with the College of Applied Technical Science Nis, Aleksandra Medvedeva 20, 18000 Nis, Serbia, E-mail: [mirko.kosanovic@vtsnis.edu.rs](mailto:mirko.kosanovic@vtsnis.edu.rs)



a very complex problem [5]. Hereafter, we will point to some important features in the transmission of data in WMSN, which are representing the basic things on which we must pay attention, and that require the efficient solutions.

#### A. The increased number of packets

In addition to significantly increased amount of data that the network needs to transmit, there is a problem in the speed with which data is sent. Thus, while the data rate in WSN is usually about 40 kb/s, in WMSN we must have min. 64 kb/s to send audio data, and min. 500 kb/s to send the video signal [6]. Bearing in mind that the delivered packets are with the length of payload data limited around of 30 bytes, it is clear that here we have a very large amount of packages. On the other side, the amount of packets further increases because we have many redundant packets coming from other neighboring SNs, due to multi-hop communication. The traffic intensity increases with in-sink approach, because it collected all the data coming from the peripheral SNs. Another problem is the different intensity of traffic, that can be classified as periodic or event triggered (event driven).

#### B. The degree of reliable packet

In most cases in WSN, the degree of reliable (reliability) data was not great. This has arisen from the fact that most applications are not required that each packet is received. It is sufficient to receive only one correct package from SNs from a region, not all, because in most case, they have the same or similar information. It gives us more freedom in the development of network protocols for WSN. But in the WMSN it is not the some case, because the nature of the information transmission through network must be provided with completely reliable transmission. This fact making them very closely to standard wired networks. On the other hand, however, all multimedia data, images, video or audio information consists of data that have different standards for secure data transmission. To achieve a data rate of multimedia information that had previously been described, SNs can not send all data from which this information is made. They sent only some relevant frames that contain some parts of this information, while the rest of the information will be obtained through different methods of interpolation at the destination, i.e. at sink. For example, it is critical to lose during transmission the packet with ROI (Region Of Interest) data if it is coding with JPEG 2000, or I-frame with MPEG compression [7]. If this happens, we irretrievably loss the transmitted information, and interpolation method is not possible. In WSN's two methods to solve this problem are used: HtH (Hop-to-Hop) and EtE (End-to-End). HtH method is demonstrated better results than the traditional EtE mechanism, because it reduces the density of traffic and thereby saving power consumption of SNs, which means long lifetime of WSN. Due to the large number of packages making up the information in WMSN better results have shown when a combination of both methods is used [7].

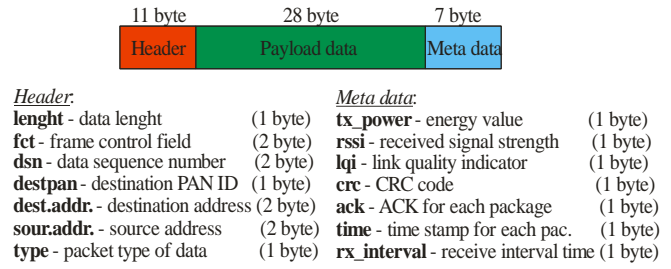


Fig. 1. Structure of message\_t package in TinyOS

#### C. Packet size

In most applications, communications in WMSN consists of a large number of packets which must be sent to only one frame. The size of these packages is limited to the applied operating system and the size of the transmitter buffer in SNs. Thus, TinyOS, which has become almost standard operating system for WSN, use program structure message\_t for communication between the SN's (see Fig. 1) [8]. It consists of the standard size of 28 bytes for useful data size and a 11 byte for header. On the other hand, the standard communication chip CC2420, standard IEEE 802.15.4, use 128 bytes for packet length. In both cases this size can be changed by program operation [9]. Bearing in mind that the standard package, that is sent, have 28-110 byte useful (payload) data (message\_t in TinyOS), it turns out that the number of packets which must be sent for only one picture are very large and exceeds the number of over a thousand packages. As in these packages we have great part of "useless" data header and meta data (28/18), is of great importance that these "useless" data reduce to a minimum. Primary role in it should be played by the development of new transport protocols designed for multimedia transmission of information, which should reduce the amount of data that is sent for header and metadata [9].

#### D. Control of traffic density

Effect of increased traffic in these networks should be emphasized. Due to large amounts of data required to send a multimedia information, and taking into account that the packets are sent from these data are limited to 28-110 byte useful (payload) data, it is clear that the number of packets are very large. On the other hand, the number of packets is further increase with many redundant packets arriving from other SN. The traffic intensity increases with the approach of the main sensor node - sink, because it collected all the data coming from the peripheral SNs. A special problem to solving this problem is the different intensity of traffic that can be classified as periodic or event triggered (event driven). This is that the traffic becomes very intensive only when there are changes in the monitored region. For any other time, the intensity is very small, i.e. it is almost nonexistent. Due to the nature WMSN, the detection of changes in time, there is a large number of redundant data. It follows from the fact that a large number of SNs can detect the same event. To effectively

solve this problem we need to develop more algorithms that will contribute to: the effective selection of the observed events and the elimination of redundant data, detection and elimination of traffic congestion, congestion avoidance and control traffic flow in WMSN. Most of these algorithms base their decisions on the utilization of the buffer control in the SNs, as well as to control the frequency/speed of sending data (ACC-Active Congestion Control) [4].

### III. PROPOSED SOLUTION

During development of the new transport protocol we were aware that the protocol must primarily meet the requirements of energy-aware protocol type, and is simple enough to be implemented in hardware and software of a large number of WMSN applications. In addition we were aware that in the future, because of the increased need to carry more multimedia data will need a protocol that will enable reliable transmission in both directions. Therefore, the proposed solution is treated equally and provides reliable data transfer (reliable transport of data) in both directions, up-streams and down-streams. In most applications that need to transmit multimedia information, communication consists of a large number of packets to be sent for only one frame. Such as, for sending of a typical video images, we need about 25-30 MB (28 888 bytes in JPEG or 319 254 BMP-byte code) [6]. Having in mind that the standard package that is sent take from 28-110 bytes of useful (payload) data it turns out that the number of packets to be sent to only one picture is about 263-1032 packets. As in these packages is usually a very small number of useful data, in relation to information relating to the header and metadata (28/18), if we succeed to reduce the size of "useless" data, we will save a considerable amount of energy. In this way, we will increase the useful data and the total packet length would stay the same, which would lead to a significant reduction in the number of sent packets. As for the specific topology of these networks, communication mechanism is multi-hop (hop-by-hop), every SN's in addition to basic functions, detection of an event, should meet the function of the router, i.e. to redirects the large number of packets passing through it. This means that if we have multi-hop communication we have much more packets between SNs. Saving only one byte in packets which are transmitted, we can save a large amount of power energy due multiplies number of skips between SNs. On the other hand, reducing the number of packets reduces the intensity of traffic, and thereby reducing the percentage of errors. All this contributes to saving large amounts of power energy in the SN's, which is the main goal of any protocol in WSN [10].

The basic idea of our solution consists in reducing the header, and increases the space for transmits useful information. This is done in header in which we dropped 2B which are used to denote the unique data sequence number of packets being sent (field *dsn* on Fig.1). To maintain unique identification of packets sent, it is compensated by multiple different addresses that are dynamically allocated to SNs. All addresses within a segment of the WMSN dispose the main (master) node. From the standpoint of IP addressing this node acts as a DHCP server. It dynamically assigned addresses to

SNs, depending on the traffic intensity, quality, quantity and direction data. One node in the sensor network can have one, two, four or more addresses ( $2^n$  address), which are awarded based on traffic load, the number of SN's that have data to sent, and the amount of data being sent. If exists only down-stream communication, master SN all available addresses assigned to SN's. In the case of up-stream communication each SNs has only one base address, while all the remaining addresses are assigned to the master SN. Each message can be divided into multiple packets; each of them becomes a unique packet that can be sent independently. The number of packets depends on the number of addresses assigned to the SN. Destination SN collects all the packages and create message that is sent. The order of packets arrival at the destination is not important, because all packets is with different addresses and actual reconstruction of the message can be done. If the size of message that is sent is greater than the number of packets, i.e. address range assigned to the SN's, then a message is divided into blocks where each block has as many packets as the addresses assigned to SN's.

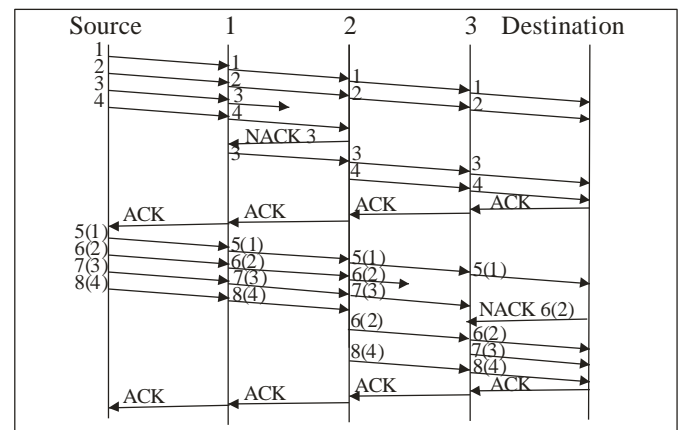


Fig.2. Transport sequence of frames among SNs in WSN

The Fig. 2 shows the process of sending one message which consists of two blocks of 4 packets (SN destination addresses have been assigned to 4), as well as the procedure is carried out with the bad reception of a frame. The Fig. 2 shows that the transient SN uses technique NACK, meaning that the transient SN can accept packets in any order.

If you accept a framework that is in the order, NACK message is sent only after the expiration of a timer that will later be explained. Each node accepts the framework is its buffering and passes it to the next node. Preferably the buffer size in each node is the same for all SN's. After the final destination SN receives all packets from the first block (in our case, all 4 packets), it sends an ACK message to the original SN who started this sending. On this basis, each SNs can reset its buffer in preparation for receiving new data block. Due to the nature of WMSN, primarily due to the large number of errors in communication, it is possible that many transient SNs do not receive the correct packets. In order to still allow each SNs that can accommodate all packets sent and to reset its buffer and receives a new frame with the same address, our solution introduced the three delay counters on each SNs. The role of the first counter is to disable the receipt of the frame

currently sending NACK messages, unless it receives a frame in the sequence, because it assumes that it can also be delayed by multiple pathways. The second counter is introduced to allow sufficient time to each SNs that can be requested to resend the frame. It has a role to control the speed of sending the frame and thus the intensity of traffic in the WMSN. Each detection NACK message causes increase the value of the delay counter SN who received the message. In the initial state of the state of the counter is set to 0, and its value increases depending on the number of NACK messages that have reached him. The greater the number of received NACK message, the greater the value of the counter because it is assumed that there was deterioration in a relationship: higher traffic density, or communication problems. The third counter is reset to enable the role of the receiving buffer if the connection is asymmetric and difficult conditions, has not received a return ACK message. The timer is activated upon receipt of all necessary frameworks from one block (4 packets in our case). Now, if it receives a new data set with the same addresses and the value of the counter has expired, it is possible to reset the receive buffer and a new block of data regardless of whether we have received confirmation ACK from the destination node. When the ACK message arrives to the source, SN continues sending the next block of data that are stored in boxes that have the same address as the previous one, but with other information (now the packets 5, 6, 7 and 8 have addresses 1, 2, 3 and 4 respectively). The procedure is repeated until the entire message is sent.

#### IV. ANALYSIS OF PROTOCOL SUCCESS

In order to evaluated our solution and the main savings for assigning multiple addresses to each SN wictc gets the required reduction in the header bytes in the frame to be sent. Now there is no need to send the data sequence number of the packet because it is uniquely determined by the address of SN. We will assume that the sensor network operating in ideal conditions, i.e. that no additional retransmissions and analyze the total number of packets are sent in function of number of pictures. We will study the case when the packet size is standard length of 46 bytes (11 byte header, 28 bytes of payload data bytes and 7 meta data) for standard message\_t and compare with our solution.

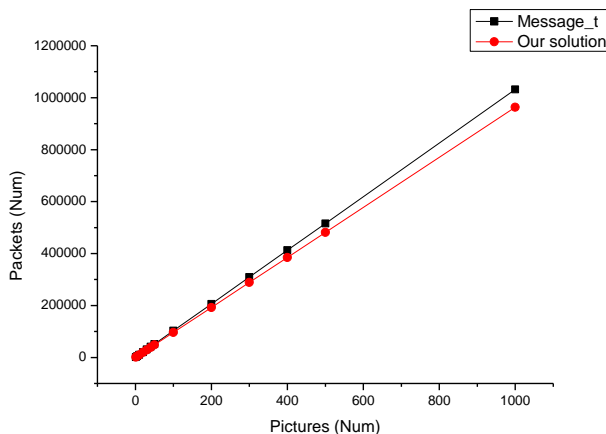


Fig.3. Number of packets in function of transmitted pictures

According to the graphics presented in Fig. 3. it can conclude that our solution always give better results. As the number of transmitted image increases, the difference in the transferred messages in favor of our proposed solutions, increase.

#### V. CONCLUSION

Transmission of multimedia data in environments that are subject to a number of errors is a very complex problem. When you add very poor resources of SNs, complex nature and unpredictable conditions in which they work problem is even more complicated. Just because of this, issue is very interested many researchers and currently represents one of the most researched areas in the WMSN. This paper has attempted to group together all the issues and parameters that should be taken into account for this transfer, to assist in the further development of an efficient protocol for transmission of multimedia data. We propose a simple solution, with reduction of the header for the only two bytes. Our calculation show us that can be achieve to savings about 1.8-6.66 % per packets (depends of packet length: 128 or 46 bytes) and thus extend the life of the entire WMSN. It should be noted that in the paper we shows the example when it comes to single-hop data transmission. If it is applying multi-hop transmission, that is standard for the WMSN, these will be obtained even better results because the number of packets transmitted increases even more.

#### REFERENCES

- [1] I.F.Akyildiz, T.Melodia, K.R.Chowdhury, "Wireless Multimedia Sensor Networks: Applications and Testbeds", Proceeding of the IEEE, Vol.96, No.10, 2008
- [2] Taner Cevik, Alex Gunagwera, Nazife Cevik, "A Survey of Multimedia Streaming in Wireless Sensor Networks: Progress, Issue and Design Challenges", International Journal of Computer Networks & Communications, Vol.7, No.5, 2015
- [3] Daniel G. Costa, Luiz Affonso Guedes, "A Survey on Transport Protocols for Wireless Multimedia Sensor Networks", <http://www.itis.org/digital-library/manuscript/286>, 20.03.2016
- [4] Z.Rosberg, R.P.Liu, A.Y.Dong, L.D.Tuan,S.Jha, "ARQ with Implicit and Explicit ACKs in WSN", <http://www.ict.csiro.au/staff/ren.liu/publications/2008-rosberg-glob.pdf>, acc. 20.01.2016
- [5] S.Misra, M.Reisslein, G.Xue, "A Survey of Multimedia Streaming in WSN", [http://www.public.asu.edu/~ssatyaja/papers/ieee\\_survey.pdf](http://www.public.asu.edu/~ssatyaja/papers/ieee_survey.pdf), acc. 20.01.2016
- [6] H.Öztarak, A.Yazici, D.Aksoz, R. George,"Multimedia Processing in WSN", IEEE 4th International Conference Innovations in Information Technology, page78-82, 2007.
- [7] Joa-Hyoung Lee and In-Bum Jung, "Reliable Asynchronous Image Transfer Protocol in Wireless Multimedia Sensor Networks", Sensors 2010, 10, 1486-1510
- [8] P.Levis, TEP 111, message\_t, <http://www.tinyos.net /tinyos-2.x/doc/html/tep111.html>, acc. 20.01.2012
- [9] M.Kosanović, M.Stojčev, "Reliable Transport Protocol with multi address sensor nodes", INFOTEH 2010, Jahorina 2010
- [10] S.Soro, W.Heinzelman, "A survey on Visual Sensor Networks", Advances in Multimedia, Volume 2009



# Performance analysis of FSO transmission over M atmospheric turbulence channels

Petar Spalević<sup>1</sup>, Bojan Prlinčević<sup>2</sup>, Stefan Panić<sup>3</sup>, Mladen Veinović<sup>4</sup>, Aleksandra Kokić Arsić<sup>2</sup>

**Abstract** – In this paper, we have derived closed form expressions for standard performance measures criteria of free space optics (FSO) communication systems operating over M atmospheric turbulence channels. Since M model represents general turbulence model, which can be reduced to other well-known turbulence models, obtained closed-form expressions for bit-error rate (BER) and ergodic channel capacity (CC) could be used for carrying out performance analysis over various FSO transmission scenarios. Capitalizing on obtained expressions, values for BER over given modulation formats, and ergodic CC are numerically evaluated, graphically presented and discussed in the function of FSO link parameters, such as number of large-scale cells of the scattering process, amount of fading, amount of scattering power coupled to the line-of-sight (LOS) component and average power of the total scatter components.

**Keywords** – Free Space Optics, M-distribution model, Bit Error Rate, Ergodic Channel Capacity.

## I. INTRODUCTION

Free space optics (FSO) transmission has emerged recently as efficient solution for obtaining secure, high data rate, wide bandwidth communication, due to its lack of licensing requirements, non-susceptibility to interferences and cost-effectiveness [1]. However, mentioned advantages of FSO transmission are significantly reduced by the occurrence of scintillation in the optical channel. Variations of the refractive index, atmospheric turbulence, lead to irradiance fluctuations of the received optical signal-scintillation. Channel modeling is very important in order to carry out accurate estimation of system performances in the FSO link designing process. In that purpose, extensive research has been conducted, and numerous mathematical models for the probability density function (PDF) of the irradiance are proposed for various fluctuation scenarios [2-8]. First, models which describe the weak irradiance fluctuation conditions such are K-distribution,

Nakagami-m, the normalized K-distribution, as well as a log-normal distribution [3] are proposed. Mathematical model which describes wide range of turbulence conditions is Gamma-Gamma model [4]. In order to describe the fluctuations of widest possible level Gama-Gamma model has long been the most convenient and most described in the literature. As a result of the quest for obtaining a general and mathematically simpler model to describe the widest possible range of fluctuations, in the paper [9] is presented a generalized M model. This model is applicable to a large number of FSO fluctuation conditions. Because of its generality, the generalized M model can be reduced to large number of other known models, such are Rice-Nakagami [2], Gamma [4], HK, Gamma-Gamma, Gama-Rician, lognormal [9]. In this paper, performance analysis of FSO transmission over atmospheric channel modeled with generalized M model will be carried out. Closed form expressions will be derived for bit-error rate (BER) over Binary Differentially Phase Shift Keying (BDPSK) modulation schemes. Further, closed form expressions for ergodic channel capacity (CC) will also be derived. Based on these expressions, average BER and ergodic CC values will be evaluated, graphically presented and discussed in the function of FSO link parameters.

The remainder of this paper is structured as follows: section II describes system model; section III provides performance analysis; numerical results are presented in section 4; conclusion remarks are given in section 5.

## II. SYSTEM MODEL

FSO channel modeled by generalized M distribution, describes the FSO transmission carried out through three signal components, line-of-sight (LOS) component  $U_L$  and two scattered components, denoted as  $U_S^{cop}$  and  $U_S^{ind}$ . The first scattered component  $U_S^{cop}$ , is assumed to be coupled to the LOS component, representing quasi-forward optical signal. Component denoted as  $U_S^{ind}$  is statistically independent of first two components and denotes optical field scattering, occurring from the energy which is scattered by the off-axis eddies. Composed optical field can now be represented as [9]:

$$U = (U_L + U_S^{cop} + U_S^{ind}) \exp(\chi + jS) \quad (1)$$

with  $\chi$  and  $S$  being real random variables representing the log-amplitude and phase fluctuations of the optical field. LOS components is defined as  $U_L = \sqrt{G} \sqrt{\Omega} \exp(j\phi_L)$ , while scattered components are defined as

<sup>1</sup>Petar Spalević, is with the Faculty of Technical Sciences, University of Pristina, Knjaza Milosa 7, 38220 K. Mitrovica, Serbia, E-mail: petar.spalevic@pr.ac.rs.

<sup>2</sup>Bojan Prlinčević and Aleksandra Kokić Arsić are with the Higher Technical Professional School in Zvečan, Nušićeva 6, Zvečan, Serbia, E-mail: b.prlincevic@vts-zvecan.edu.rs, akokicster@gmail.com.

<sup>3</sup>Stefan Panić is with the Faculty of Natural Science and Mathematics, University of Pristina, Lole Ribara 29, 38220 K. Mitrovica, Serbia, E-mail: stefanpnc@yahoo.com.

<sup>4</sup>Mladen Veinović, is with the Faculty of Electrical and Computer Engineering, Singidunum University, Danijelova 32, 11000 Belgrade, Serbia, E-mail: mveinovic@singidunum.ac.rs.



$U_s^{cop} = \sqrt{\rho}\sqrt{G}\sqrt{2b_0} \exp(j\phi_c)$  and  $U_s^{ind} = \sqrt{(1-\rho)}U_s'$  with parameter  $\Omega = E[|U_L|^2]$ , total scattered component denoted as  $2b_0 = E[|U_s^{cop}|^2 + |U_s^{ind}|^2]$ . Parameter  $\rho$ ,  $0 \leq \rho \leq 1$ , denotes the factor expressing the amount of scattering power coupled to the LOS component and depends on the propagation path length, while  $U_s'$  denoting circular Gaussian complex random variable, and  $G$  denoting Gamma random process with unit mean value. Constants  $\phi_L$  and  $\phi_c$  denote deterministic phases of the LOS and coupled-to-LOS scatter components.

Received irradiance can be expressed as:

$$I = |U_L + U_s^{cop} + U_s^{ind}|^2 \exp(2\chi) = YX \quad (2)$$

where  $X = \exp(2\chi)$  denotes large-scale fluctuations and  $Y = |U_L + U_s^{cop} + U_s^{ind}|^2$  denotes small-scale fluctuations, distributed according to (20) and (21) from [9]. It is further shown that received irradiance is distributed according to Generalized M as:

$$f(I) = A \sum_{k=1}^{\beta} a_k I^{\frac{\alpha+k}{2}-1} K_{\alpha-k} \left( 2\sqrt{\frac{\alpha\beta I}{\gamma\beta + \Omega'}} \right) \quad (3)$$

where,

$$\left\{ \begin{aligned} A &= \frac{2\alpha^{\frac{\alpha}{2}}}{\gamma^{1+\frac{\alpha}{2}}\Gamma(\alpha)} \left( \frac{\gamma\beta}{\gamma\beta + \Omega'} \right)^{\beta+\frac{\alpha}{2}} \\ a_k &= \binom{\beta-1}{k-1} \frac{(\gamma\beta + \Omega')^{1-\frac{k}{2}}}{\Gamma(k)} \left( \frac{\Omega'}{\gamma} \right)^{k-1} \left( \frac{\alpha}{\beta} \right)^{\frac{k}{2}} \end{aligned} \right. \quad (4)$$

where  $K_\nu(\cdot)$  [10, (8.432.2)] denotes the modified Bessel function of the second kind where,  $\Gamma(\cdot)$  denotes the Gamma function [10, (8.310.1)] and  $\binom{n}{k}$  represent the binomial coefficient.

Parameter  $\alpha$  represents the effective number of large-scale cells of the scattering process, while parameter  $\beta$  represents the effective number of small-scale effects, in the same form as was explained. Also stands  $\Omega' = \Omega + \rho 2b_0 + 2\sqrt{2b_0\Omega\rho} \cos(\phi_L - \phi_c)$ ,  $\gamma = 2b_0(1-\rho)$ .

After setting corresponding values, to parameters  $\rho$ ,  $\gamma$ ,  $\Omega'$ ,  $\alpha$ ,  $\beta$ ,  $X$ ,  $G$  and  $U_L$  this general M distribution model can be reduced to other well-known irradiance models, i.e.: a) by setting  $\rho=0$ , and  $\text{Var}[|U_L|]=0$  M model reduces to Rice-Nakagami model; b) by setting  $\rho=0$ , and  $\gamma=0$  it reduces to Gamma model; c) by setting  $\text{Var}[G]=0$ ,  $\rho=0$ , and  $X=\gamma$  it reduces to HK distribution model; d) by setting  $\rho=1$ ,  $\gamma=0$  and  $\Omega'=1$  it reduces to Gamma-Gamma distribution; e) by setting  $\text{Var}[X]=0$ , it reduces to Shadowed-Rician distribution; f) by setting  $\rho=0$ ,  $\text{Var}[|U_L|]=0$ , and  $\gamma \neq 0$  it reduces to Lognormal model; g) by setting  $\Omega=0$  and  $\rho=0$  or  $\beta=1$  it reduces to K distribution; h) by setting  $\Omega=0$ ,  $\rho=0$ , and  $\alpha \rightarrow \infty$  it reduces to

Exponential distribution; and i) by setting  $\beta \rightarrow \infty$  it reduces to Gamma-Rician distribution.

### III. PERFORMANCE ANALYSIS

The average BER can be evaluated directly by averaging the conditional BER,  $\bar{P}_e$ , depending on the type of modulation, i.e.:

$$P_e = \int_0^{+\infty} f_I(I) \bar{P}_e dI \quad (5)$$

and after substituting  $\bar{P}_e$  expression for BDPSK modulation scheme, it yields:

$$P_e = \int_0^{+\infty} \frac{1}{2} \text{erfc} \left( \sqrt{\frac{P_T}{\sigma_N}} \sqrt{I} \right) f_I(I) dI \quad (6)$$

In order to evaluate BER performance values of the BDPSK FSO system in M turbulence induced fading channel, after expressing integrands from (6) in terms of the Meijer's

G-function as:  $e^{-x} = \frac{1}{\sqrt{\pi}} G_{0,1}^{1,0} \left[ x \middle| \begin{smallmatrix} - \\ 0 \end{smallmatrix} \right]$  [11, (8.4.3/1)],

$K_\nu(x) = \frac{1}{2} G_{0,2}^{2,0} \left[ \frac{x^2}{4} \middle| \begin{smallmatrix} - & - \\ \frac{\nu}{2} & -\frac{\nu}{2} \end{smallmatrix} \right]$  [11, (8.4.23/1)], and we obtain:

$$P_e = \frac{1}{8} A \sum_{k=1}^{\beta} a_k \int_0^{+\infty} I^{\frac{\alpha+k}{2}-1} G_{0,2}^{2,0} \left[ \frac{\alpha\beta I}{\gamma\beta + \Omega'} \middle| \begin{smallmatrix} - & - \\ \frac{\alpha-k}{2} & -\frac{\alpha-k}{2} \end{smallmatrix} \right] G_{0,1}^{1,0} \left[ \frac{P_T}{\sigma_N} I \middle| \begin{smallmatrix} - \\ 0 \end{smallmatrix} \right] dI \quad (7)$$

Using [11, (2.24.1/1)] along with [11, (8.2.2/14)], the closed form of the average BER for the BDPSK FSO system over M turbulence induced fading can be expressed as:

$$P_e = \frac{1}{4\pi} A \sum_{k=1}^{\beta} a_k \left( \frac{\alpha\beta}{\gamma\beta + \Omega'} \right)^{-\frac{\alpha+\beta+1}{2}} G_{2,1}^{1,2} \left[ \frac{P_T}{\sigma_N} \frac{\gamma\beta + \Omega'}{\alpha\beta} \middle| \begin{smallmatrix} 1-2\alpha & 2-2k \\ 2 & 2 \end{smallmatrix} \right] \quad (8)$$

where  $G_{p,q}^{m,n}[\cdot]$  denotes the Meijer's G-function [10, (9.301)].

Assuming that the FSO channel is memoryless, stationary and ergodic with perfect channel state information (CSI) available at both the transmitting lasers and the aperture receivers, the normalized ergodic CC can be defined as:

$$\frac{C}{B_{FSO}} = \int_0^{+\infty} \log_2(1+I) f_I(I) dI \quad (\text{bit/s/Hz}) \quad (9)$$

where  $B_{FSO}$  is the channels bandwidth.

In order to derive ergodic CC of the FSO system over M turbulence induced fading, integrands from (10) should be expressed in terms of the Meijer's G-function

$$\ln(1+x) = G_{2,2}^{1,2} \left[ x \mid \begin{matrix} 1,1 \\ 1,0 \end{matrix} \right] \quad [11, (8.4.6/5)], \quad K_\nu(x) = \frac{1}{2} G_{0,2}^{2,0} \left[ \frac{x^2}{4} \mid \begin{matrix} - \\ \frac{\nu}{2}, -\frac{\nu}{2} \end{matrix} \right] \quad [11, (8.4.23/1)], \text{ and we obtain:}$$

$$\frac{C}{B_{FSO}} = \frac{1}{8\pi \log 2} A \sum_{k=1}^{\beta} a_k \int_{k=1}^{+\infty} I^{\frac{\alpha+k-1}{2}} G_{0,2}^{2,0} \left[ \frac{\alpha\beta I}{\gamma\beta + \Omega'} \mid \begin{matrix} - \\ \frac{\alpha-k}{2}, -\frac{\alpha-k}{2} \end{matrix} \right] G_{2,2}^{1,2} \left[ \frac{P_T}{\sigma_N} I^2 \mid \begin{matrix} 1,1 \\ 1,0 \end{matrix} \right] \quad (10)$$

Using [11, (2.24.1/1)] along with [11, (8.2.2/14)], the closed form of the FSO CC in M turbulence induced fading can be expressed as:

$$\frac{C}{B_{FSO}} = \frac{1}{8\pi \log 2} A \sum_{k=1}^{\beta} a_k 2^{\alpha+k} \left( \frac{\alpha\beta}{\gamma\beta + \Omega'} \right)^{-\frac{\alpha+\beta+1}{2}} G_{6,2}^{1,6} \left[ 16 \frac{P_T}{\sigma_N} \left( \frac{\gamma\beta + \Omega'}{\alpha\beta} \right)^2 \mid \begin{matrix} 1,1, \frac{1-2\alpha}{4}, \frac{2-2\alpha}{4}, \frac{1-2k}{4}, \frac{2-2k}{4} \\ 1,0 \end{matrix} \right] \quad (11)$$

#### IV. NUMERICAL RESULTS

In this section, numerical results for BER of BDPSK FSO communication systems over M turbulence induced fading channels and for CC are presented.

Figs. 1-3 illustrate obtained BER values for the BDPSK FSO system in M turbulence induced fading. At Fig. 1. is shown the influence of  $\alpha$  and  $\beta$  parameters on obtained BER. Obtained BER values are lower when parameter  $\beta$  is increasing and/or when parameter  $\alpha$  is increasing. At Fig. 2. is the influence of  $\Omega'$  and  $\gamma$  parameters on obtained BER. The performances are better, obtained BER values are lower when parameter  $\Omega'$  is taking higher values and/or when parameter  $\gamma$  takes lower values. At Fig. 3. is shown the dependence between  $\rho$  and  $\gamma$  parameter values and obtained BER.

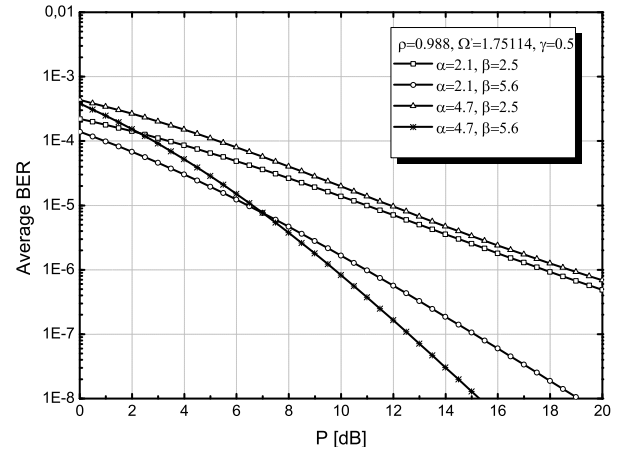


Fig. 1. BER of the BDPSK FSO system over M turbulence induced fading when  $\rho$ ,  $\Omega'$  and  $\gamma$  are constant.

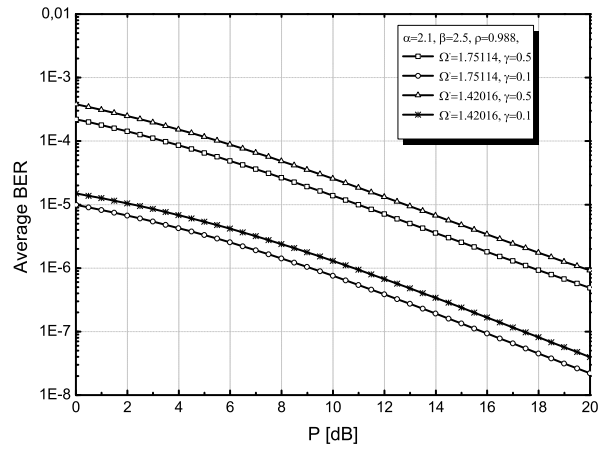


Fig. 2. BER of the BDPSK FSO system over M turbulence induced fading when  $\alpha$ ,  $\beta$  and  $\rho$  are constant.

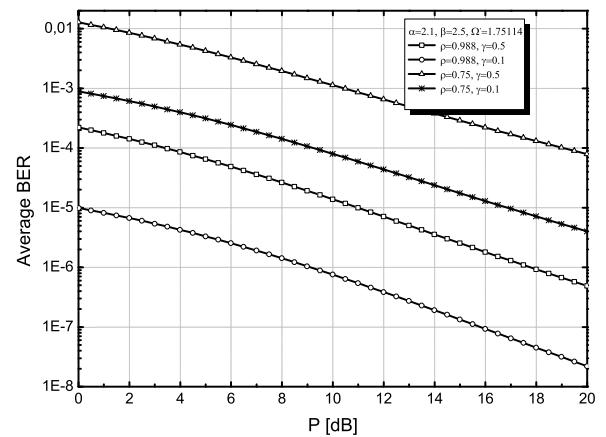


Fig. 3. BER of the BDPSK FSO system over M turbulence induced fading when  $\alpha$ ,  $\beta$  and  $\Omega'$  are constant.

Fig. 4. illustrate the normalized CC of the FSO system over M turbulence induced fading. At Fig. 4. is shown dependence

between values of  $\alpha$  and  $\beta$  parameters and obtained normalized CC. The normalized CC is higher when parameter  $\alpha$  increase. Also is visible that when parameter  $\beta$  increases normalized CC increase.

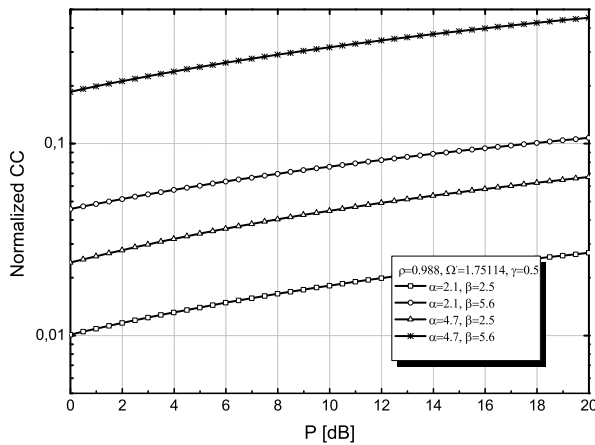


Fig. 4. Normalized CC of the FSO system over M turbulence induced fading when  $\rho$ ,  $\Omega'$  and  $\gamma$  are constant.

## V. CONCLUSION

In this paper FSO transmission over M distributed atmospheric channel was analyzed through standard performance criterions such are BER and ergodic CC. Capitalizing on closed form expressions for BDPSK BER and ergodic CC derived in this paper, observed FSO link performances were analyzed in the function of FSO transmission parameters, such are number of large-scale cells of the scattering process, amount of irradiance, amount of scattering power coupled to LOS component and average power of the total scatter components. Major contribution of this work is its generality, because results from this paper could be applied to various other FSO transmission scenarios.

## ACKNOWLEDGEMENT

This paper is carried out within the project research of Ministry of Education, Science and Technological Development of the Republic of Serbia [III 4406 and TR 32003].

## REFERENCES

- [1] K. Kazaura, K. Wakamori, M. Matsumoto, et. al., "A universal platform for convergence of fiber and free-space optical communication networks", *IEEE Comm. Mag.* **48**, 130 (2010).
- [2] L. C. Andrews and R. L. Phillips, "Laser Beam Propagation through Random Media", (SPIE, 2005).
- [3] X. Zhu and J.M. Kahn, "Free-space optical communication through atmospheric turbulence channels", *IEEE Trans. Commun.* **50**, 1293–1300 (2002).

- [4] M.A. Al-Habash, L.C. Andrews and R.L. Phillips, "Mathematical model for the irradiance probability density function of a laser beam propagating through turbulent media", *Opt. Engineering* **40**, 1554–1562 (2001).
- [5] J. H. Churnside and S. F. Clifford, "Log-normal Rician probability-density function of optical scintillations in the turbulent atmosphere", *J. Opt. Soc. Am. A* **4**, 1923–1930 (1987).
- [6] M. Toyoshima, H. Takenaka and Y. Takayama, "Atmospheric turbulence-induced fading channel model for space-to-ground laser communications links", *Opt. Express* **19**, 15965–15975 (2011).
- [7] A. Jurado-Navas, J.M. Garrido-Balsells, J.F. Paris, et al, "A unifying statistical model for atmospheric optical scintillation", invited chapter in "Numerical Simulations of Physical and Engineering Processes," 181–206 (In-Tech, 2011).
- [8] N.D. Chatzidiamantis, H.G. Sandalidis, G.K. Karagiannidis, et al. "Inverse Gaussian modeling of turbulence-induced fading in free-space optical systems", *J. Lightwave Technol.* **29**, 1590–1596 (2011).
- [9] J. M. Garrido-Balsells, A. Juado-Navas, J. F. Paris, et. al. "Novel formulation of the M model through the Generalized-K distribution for atmospheric optical channels", *J. Opt. Soc. Am.*, Vol. 23, No. 5, DOI:10.1364/OE.23.006345-23.006358.
- [10] I. S. Gradshteyn and I.M. Ryzhik, "Table of Integrals, Series and Products" (Elsevier, 2007).
- [11] A. P. Prudnikov, Y. A. Brychkov, and O. I. Marichev, "Integral and Series". 2<sup>nd</sup> edition, Vol. 3. Moskva, Fizmatlit 2003, ISBN 5-9221-0325-3.

# Performance Analysis of Wireless Macrodiversity SSC Receiver in Gamma-Shadowed nLoS Fading Channel

Caslav Stefanovic<sup>1</sup>

**Abstract** – Wireless mobile macrodiversity (MaD) system with switch and stay (SSC) receiver and two microdiversity (MiD) selection combining (SC) structures operating over Gamma shadowed non line-of-sight (nLoS) multipath fading environment is proposed. Probability density function (PDF) and cumulative distribution function (CDF) of MaD SSC receiver output signal envelope in the form of the sum of one-folded integrals are efficiently evaluated and used for derivation of outage probability (OP) of the proposed system. The results are graphically presented and discussed regarding different system model parameters.

**Keywords** – CDF; MaD; MiD; Nakagami-m; OP; PDF.

## I. INTRODUCTION

In order to diminish simultaneously short- and long-term fading macro-diversity system (MaD) with MaD receiver and two or more micro-diversity (MiD) branches is often proposed in the literature [1-2]. MaD receiver are used to reduce shadowing while MiD receivers are used to mitigate multipath fading on the system performances. Accordingly, MaD system can be efficiently applied in composite fading environment to improve performances of wireless communication system such as outage probability (OP), level crossing rate (LCR) and average fade duration (AFD).

Since, there are publications considering statistical measures of MaD selection combining (SC) system with MiD maximal-ratio-combining (MRC) or SC structures [3-5], there are scarcely papers focusing on the system performances of MaD system with switch and stay (SSC) receiver and MiD SC receivers [6]. SC and SSC are often discussed and applied diversity technique and due to its low implementation complexity, can be advantageous than other diversity techniques [7]. MaD SSC system with MaD SSC receiver and two MiD SC receivers selects MiD SC receiver to provide service to user for the time when signal envelope at the outputs is larger than the given threshold. When signal envelope average power falls below the given threshold, SSC MaD selects the other SC to give service to user.

The Nakagami-m distribution describes signal envelope variation in non line-of-sight (nLoS) multipath fading channels

when signals propagate in the environment with two or more clusters. Moreover, this distribution fits well with experimental data and is often applied in multipath environment [8].

On the other hand, log-normal distribution and Gamma distribution are used for the purpose of describing shadowing in wireless communication channel. Since multipath fading and shadowing can be present at the same time in wireless channel, it is advantageous to include both of them simultaneously into channel model. Moreover, multipath fading is confronted by MiD SC receivers at single BSs while shadowing is confronted by MaD SSC receiver combining signals from two BSs.

In this paper, the sum of one-folded integrals expression of cumulative distribution function (CDF) of MaD SSC receiver output signal envelope in the presence of Gamma shadowed Nakagami-m multipath fading is efficiently calculated and used for derivation of outage probability (OP) of the proposed model. Gamma distribution is proposed, since it can be advantageous than long-normal distribution due to its mathematical tractability. Numerical results are presented graphically and the effect of various system parameters on the proposed MaD model are examined and discussed.

## II. PDF OF MAD SSC RECEIVER OUTPUT SIGNAL

The MaD system with two MiD SC structures feeding the MaD SSC, as shown on Fig. 1. is considered. Proposed wireless channel is under the influence of Gamma shadowed Nakagami- $m$  multipath fading.

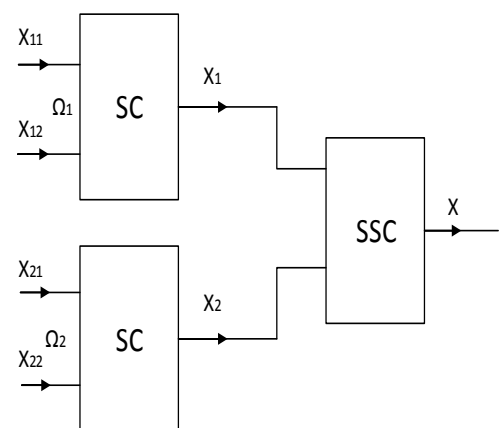


Fig. 1. The MaD system with MaD SSC receiver and two MiD SC structures

<sup>1</sup>Caslav Stefanovic is with the Faculty of Electronic Engineering, University of Nis, Aleksandra Medvedeva, 18000, Nis, Serbia, E-mail: caslav.stefanovic@gmail.com.



Probability density function (PDF) of Nakagami- $m$  signal envelope at the inputs of the first SC of the first and second branch are respectively [8]:

$$p_{x_{1i}}(x_{1i}) = \frac{2}{\Gamma(m)} \left(\frac{m}{\Omega_1}\right)^m x_{1i}^{2m-1} e^{-\frac{m}{\Omega_1} x_{1i}^2}, i = 1, 2; \quad (1)$$

where  $m$  is fading severity factor,  $\Omega_1$  is related to the local mean power of  $x_{1i}$ .

Further, CDF of Nakagami- $m$  signal envelope at the inputs of the first SC can be expressed as [11]:

$$\begin{aligned} F_{x_{1i}}(x_{1i}) &= \int_0^{x_{1i}} p_{x_{1i}}(t) dt \\ &= \frac{1}{\Gamma(m)} \gamma\left(m, \frac{m}{\Omega_1} x_{1i}^2\right), x_{1i} > 0, i = 1, 2; \end{aligned} \quad (2)$$

where  $\Omega_1$  is signal envelope average power at the input of first micro-diversity SC and  $\gamma(m, x)$  is incomplete Gamma function [10]. Since, Nakagami- $m$  fading at the input of the first SC micro-diversity is identical and independent PDF of the first SC output signal envelope is:

$$\begin{aligned} p_{x_1}(x_1) &= p_{x_{11}}(x_1) F_{x_{12}}(x_1) + p_{x_{12}}(x_1) F_{x_{11}}(x_1) \\ &= 2p_{x_{11}}(x_1) F_{x_{12}}(x_1) \\ &= \frac{4}{\Gamma(m)^2} \left(\frac{m}{\Omega_1}\right)^m x_1^{2m-1} \gamma\left(m, \frac{m}{\Omega_1} x_1^2\right) e^{-\frac{m}{\Omega_1} x_1^2}. \end{aligned} \quad (3)$$

Similarly, PDF of the second SC output is:

$$p_{x_2}(x_2) = \frac{4}{\Gamma(m)} \left(\frac{m}{\Omega_2}\right)^m x_2^{2m-1} e^{-\frac{m}{\Omega_2} x_2^2} \frac{1}{\Gamma(m)} \gamma\left(m, \frac{m}{\Omega_2} x_2^2\right) \quad (4)$$

where  $\Omega_2$  is signal envelope average power at the input of second MiD SC.

Under the assumption that both MiD SC structures are shadowed with the same obstacles, correlated long term Gamma fading is taken into account. Signal envelope average power at inputs of two MiD SC structures,  $\Omega_1$  and  $\Omega_2$  are Gamma distributed [12]:

$$\begin{aligned} p_{\Omega_1 \Omega_2}(\Omega_1 \Omega_2) &= \frac{(\Omega_1 \Omega_2)^{\frac{c-1}{2}}}{\Gamma(c)(1-\rho^2)\rho^{c-1}\Omega_0^{c+1}} e^{-\frac{\Omega_1 + \Omega_2}{\Omega_0(1-\rho^2)}} \\ &\quad \times I_{c-1}\left(\frac{2\rho}{\Omega_0(1-\rho^2)}(\Omega_1 \Omega_2)^{\frac{1}{2}}\right) \\ &= \frac{1}{\Gamma(c)(1-\rho^2)\rho^{c-1}\Omega_0^{c+1}} \sum_{i=0}^{\infty} \left(\frac{\rho}{\Omega_0(1-\rho^2)}\right)^{2i+c-1} \\ &\quad \times \frac{1}{i! \Gamma(i+c)} \Omega_1^{i+c-1} \Omega_2^{i+c-1} e^{-\frac{\Omega_1 + \Omega_2}{\Omega_0(1-\rho^2)}}, \end{aligned} \quad (5)$$

where  $\rho$  is correlation parameter of the shadowing,  $c$  is shadowing severity parameter and  $\Omega_0$  is mean value of  $\Omega_1$  and  $\Omega_2$ .

Finally, PDF of MaD SSC receiver output signal envelope is:

$$P_x(x) = \frac{1}{2} \int_{v_T}^{\infty} d\Omega_1 \int_0^{\infty} P_{x_1}(X|\Omega_1) p_{\Omega_1 \Omega_2}(\Omega_1 \Omega_2) d\Omega_2$$

$$\begin{aligned} &+ \frac{1}{2} \int_0^{v_T} d\Omega_1 \int_0^{\infty} P_{x_2}(X|\Omega_2) p_{\Omega_1 \Omega_2}(\Omega_1 \Omega_2) d\Omega_2 \\ &+ \frac{1}{2} \int_{v_T}^{\infty} d\Omega_2 \int_0^{\infty} P_{x_2}(X|\Omega_2) p_{\Omega_1 \Omega_2}(\Omega_1 \Omega_2) d\Omega_1 \\ &+ \frac{1}{2} \int_0^{v_T} d\Omega_2 \int_0^{\infty} P_{x_1}(X|\Omega_1) p_{\Omega_1 \Omega_2}(\Omega_1 \Omega_2) d\Omega_1 \\ &+ \frac{1}{2} \int_0^{v_T} d\Omega_2 \int_0^{\infty} P_{x_1}(X|\Omega_1) p_{\Omega_1 \Omega_2}(\Omega_1 \Omega_2) d\Omega_1. \end{aligned} \quad (6)$$

where  $P_{x_1}(X|\Omega_1)$  and  $P_{x_2}(X|\Omega_2)$  are PDFs at the first and second input of SSC, respectively. After substitution Eq. (3), Eq. (4) and Eq. (5) in Eq. (6) and using additional mathematical manipulation [10], PDF at the output of SSC becomes:

$$\begin{aligned} P_x(x) &= \left(\frac{2m^m x^{2m-1}}{\Gamma(m)^2 \Gamma(c)(1-\rho^2)\rho^{c-1}\Omega_0^{c+1}} \sum_{i=0}^{\infty} \left(\frac{\rho}{\Omega_0(1-\rho^2)}\right)^{2i+c-1}\right) \\ &\quad \times \frac{1}{i! \Gamma(i+c)} (\Omega_0(1-\rho^2))^{i+c} \Gamma(i+c) \\ &\quad \times \int_{v_T}^{\infty} \Omega_1^{i+c-1} \gamma\left(m, \frac{m}{\Omega_1} x^2\right) e^{-\frac{m}{\Omega_1} x^2 - \frac{\Omega_1}{\Omega_0(1-\rho^2)}} d\Omega_1 \\ &\quad + \\ &\quad \left(\frac{2m^m x^{2m-1}}{\Gamma(m)^2 \Gamma(c)(1-\rho^2)\rho^{c-1}\Omega_0^{c+1}} \sum_{i=0}^{\infty} \left(\frac{\rho}{\Omega_0(1-\rho^2)}\right)^{2i+c-1}\right) \\ &\quad \times \frac{1}{i! \Gamma(i+c)} (\Omega_0(1-\rho^2))^{i+c} \gamma\left(i+c, \frac{v_T}{\Omega_0(1-\rho^2)}\right) \\ &\quad \times \int_0^{\infty} \Omega_1^{i+c-1} \gamma\left(m, \frac{m}{\Omega_1} x^2\right) e^{-\frac{m}{\Omega_1} x^2 - \frac{\Omega_1}{\Omega_0(1-\rho^2)}} d\Omega_1. \end{aligned} \quad (7)$$

### III. CDF OF MAD SSC RECEIVER OUTPUT SIGNAL

Since Nakagami- $m$  fading at the input of the first dual branches SC structure is identical and independent, CDF of Nakagami- $m$  random variable  $x_1$  is:

$$\begin{aligned} F_{x_1}(x_1) &= F_{x_{11}}(x_1) F_{x_{12}}(x_1) \\ &= \frac{1}{\Gamma(m)^2} \left(\gamma\left(m, \frac{m}{\Omega_1} x_1^2\right)\right)^2. \end{aligned} \quad (8)$$

where  $F_{x_1}$  is CDF of Nakagami- $m$  random variable at the output of the first MiD SC. Similarly, CDF of random variable  $x_2$  is:

$$F_{x_2}(x_2) = \frac{1}{\Gamma(m)^2} \gamma\left(m, \frac{m}{\Omega_1} x_2^2\right)^2. \quad (9)$$

CDF of MaD SSC receiver output signal envelope is:

$$F_x(x) = \frac{1}{2} \int_{v_T}^{\infty} d\Omega_1 \int_0^{\infty} F_{x_1}(X|\Omega_1) p_{\Omega_1 \Omega_2}(\Omega_1 \Omega_2) d\Omega_2$$

$$\begin{aligned}
 & + \frac{1}{2} \int_0^{v_T} d\Omega_1 \int_0^\infty F_{x_2}(X|\Omega_2) p_{\Omega_1\Omega_2}(\Omega_1\Omega_2) d\Omega_2 \\
 & + \frac{1}{2} \int_{v_T}^\infty d\Omega_2 \int_0^\infty F_{x_2}(X|\Omega_2) p_{\Omega_1\Omega_2}(\Omega_1\Omega_2) d\Omega_1 \\
 & + \frac{1}{2} \int_0^{v_T} d\Omega_2 \int_0^\infty F_{x_1}(X|\Omega_1) p_{\Omega_1\Omega_2}(\Omega_1\Omega_2) d\Omega_1. \quad (10)
 \end{aligned}$$

where  $F_{x_1}(X|\Omega_1)$  is CDF at the first input of SSC and  $F_{x_2}(X|\Omega_2)$  is CDF at the second input of SSC. Finally CDF of MaD output signal envelope becomes [10]:

$$\begin{aligned}
 F_x(x) = & \frac{1}{\Gamma(m)^2 \Gamma(c) (1-\rho^2) \rho^{c-1} \Omega_0^{c+1}} \sum_{i=0}^{\infty} \left( \frac{\rho}{\Omega_0(1-\rho^2)} \right)^{2i+c-1} \\
 & \times \frac{1}{i! \Gamma(i+c)} (\Omega_0(1-\rho^2))^{i+c} \Gamma(i+c) \\
 & \times \int_{v_T}^\infty \Omega_1^{i+c-1} \gamma\left(m, \frac{m}{\Omega_1} x^2\right)^2 e^{-\frac{\Omega_1}{\Omega_0(1-\rho^2)}} d\Omega_1 \\
 & + \\
 & \left( \frac{1}{\Gamma(m)^2 \Gamma(c) (1-\rho^2) \rho^{c-1} \Omega_0^{c+1}} \sum_{i=0}^{\infty} \left( \frac{\rho}{\Omega_0(1-\rho^2)} \right)^{2i+c-1} \right) \\
 & \times \frac{1}{i! \Gamma(i+c)} (\Omega_0(1-\rho^2))^{i+c} \\
 & \times \gamma\left(i+c, \frac{v_T}{\Omega_0(1-\rho^2)}\right) \int_0^\infty \Omega_1^{i+c-1} \gamma\left(m, \frac{m}{\Omega_1} x^2\right)^2 e^{-\frac{\Omega_1}{\Omega_0(1-\rho^2)}} d\Omega_1 \quad (11)
 \end{aligned}$$

#### IV. NUMERICAL RESULTS

PDF of MaD SSC receiver output signal envelope for various system model parameters such as: fading severity parameter  $m$ , shadowing severity parameter  $c$ , correlation parameter  $\rho$ , average power  $\Omega_0$  and threshold value  $v_T$  is calculated using software package mathematica and presented on Figures 2-4. Curves show that for lower values of the signal envelope  $x$ , PDFs increase, reach its maximums, and then decrease for higher values of  $x$ .

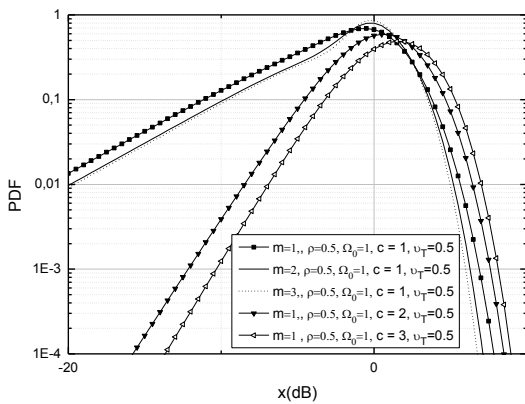


Fig. 2. PDF of MaD SSC output for different values of parameters  $m$  and  $c$  and constant values of  $\rho$ ,  $v_T$  and  $\Omega_0$ .

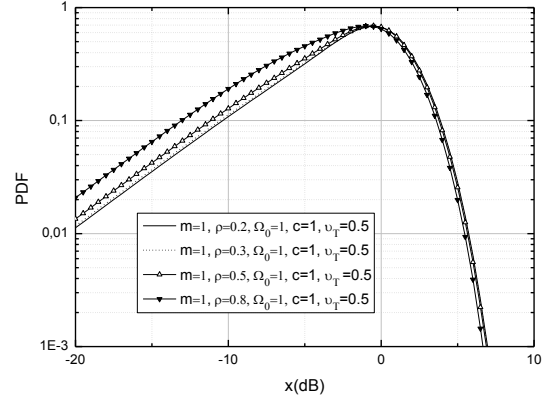


Fig. 3. PDF of MaD SSC output for different values of parameter  $\rho$  and constant values of  $m$ ,  $c$ ,  $v_T$  and  $\Omega_0$ .

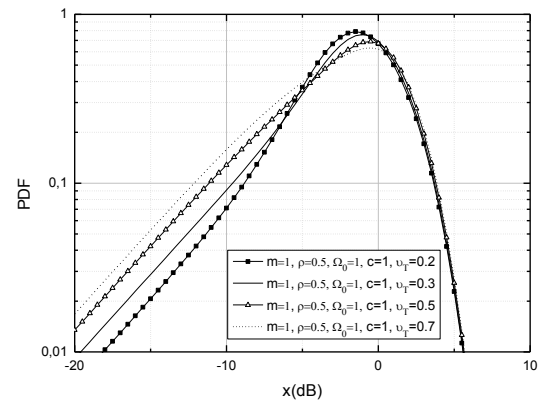


Fig. 4. PDF of MaD SSC output for different values of parameter  $v_T$  and constant values of  $m$ ,  $c$ ,  $\rho$  and  $\Omega_0$ .

Outage probability (OP), defined as the probability that the output signal envelope of MaD SSC receiver drops below given outage threshold  $v_T$ , is given as [9],

$$P_{out} = \int_0^{v_T} F_X(t) dt = F_X(v_T), \quad (12)$$

After substitution Eq. (11) in Eq. (12) and also using software package mathematica OP for various system parameters is presented on Figures 5-7.

OP of MaD SSC system for various values of  $m$  and constant  $c$ ,  $\rho$ ,  $\Omega_0$  and  $v_T$  is presented on Fig. 5. It can be seen that by increasing parameter  $c$ , the performances improves, since OP decreases. Moreover, the system performance improvement is evident by increasing parameters  $m$ . It's also evident that parameter  $c$  has greater impact on OP then parameter  $m$ . Fig. 6. shows OP of MaD SSC system for constant values  $m$ ,  $c$ ,  $v_T$  and  $\Omega_0$  and different values of  $\rho$ . Since, there is no correlation for  $\rho=0$ , the influence of correlation on the MaD SSC system performances increases for higher values of  $\rho$ , resulting in system performance degradation, as expected. Since, OP degrade by decreasing the parameter  $\rho$  negative influence of correlation on the performances for higher values of  $\rho$  is obvious. OP of MaD SSC system for constant values  $m$ ,  $c$ ,  $\rho$

and  $\Omega_0$  and different values of  $v_T$  is shown on Fig. 7. OP improves by increasing the SSC threshold  $v_T$ , which means that lower values of  $v_T$  contribute to the stability of MaD system.

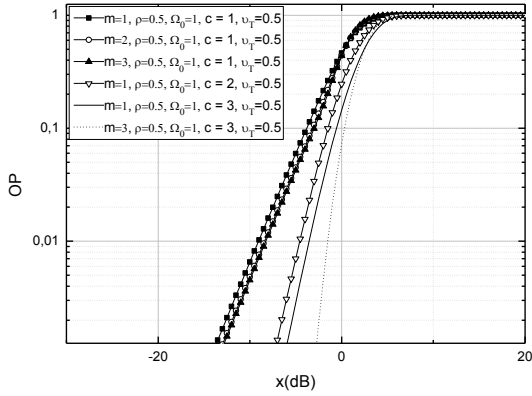


Fig. 5. OP of MaD SSC receiver for different values of parameters  $m$  and  $c$  and constant values of  $\rho$ ,  $v_T$  and  $\Omega_0$ .

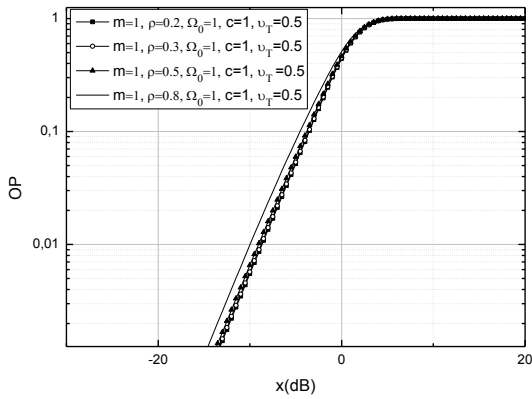


Fig. 6. OP of MaD SSC receiver for different values of parameter  $\rho$  and constant values of  $m$ ,  $c$ ,  $v_T$  and  $\Omega_0$ .

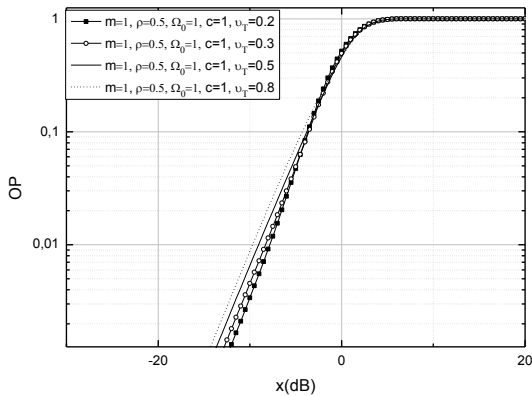


Fig. 7. OP of MaD SSC receiver for different values of parameter  $v_T$  and constant values of  $m$ ,  $c$ ,  $\rho$  and  $\Omega_0$ .

## V. CONCLUSION

MaD system with MaD SSC receiver with two dual MiD SC structures is proposed, since it has lower complexity implementation than other MaD systems. PDF and CDF of output signal envelope are calculated and used for derivation of OP of SSC MaD system in the form of the sum of one-folded integrals. The results are graphically presented regarding different system model parameters. The system is theoretically more efficient by designing with higher values of parameter  $c$  and  $m$ . Further, influences of correlation parameter  $\rho$  and  $v_T$  are also given and discussed. The best possible scenario in theory can be achieved by increasing  $c$  and  $m$  and by decreasing  $\rho$  and  $v_T$ .

## REFERENCES

- [1] P. M. Shankar, "Macrodiversity and microdiversity in correlated shadowed fading channels", *IEEE Trans. on Vehicular Technol.*, vol. 2, no. 58, pp. 727-732, 2009.
- [2] B. Zhang, Z. Zhong, B. Ai, R. He, "Impact of Shadowing Correlation on Microdiversity and Macrodiversity of Cellular System in High-speed Railway Environments", In *Proc. Progress In Electromagnetics Research Symposium*, Guangzhou, China, Aug. 25-28, 2014, pp 2159-2163.
- [3] N. Sekulović, M. Stefanović, "Performance analysis of system with micro- and macrodiversity reception in correlated gamma shadowed Rician fading channels", *Wireless Personal Communications*, vol. 65, no.1, pp. 143-156, 2012.
- [4] D. Djosic, D. Stefanovic, C. Stefanovic, "Level Crossing Rate of Macro-diversity System with Two Micro-diversity SC Receivers over Correlated Gamma Shadowed  $\alpha$ - $\mu$  Multipath Fading Channels", *IETE Journal of Research*, pp. 1-6, 2015.
- [5] S. Panić, D. Stefanovic, I. Petrović, M. Stefanović, J. Anastasov, D. Krstic "Second order statistics of selection macro-diversity system operating over Gamma shadowed  $k$ - $\mu$  fading channels", *EURASIP Journal on Wireless Communications and Networking*, vol. 2011, Oct. 2011.
- [6] Č. Stefanović, "Macrodiversity system with macrodiversity SSC receiver and two microdiversity receivers in the presence of composite fading environment", *23<sup>rd</sup> Telecommunications forum - TELFOR 2015, Proceedings of papers*, Belgrade, Serbia, pp. 321-324, 24-26 Nov. 2015.
- [7] J. Zhao, P. Fan, N. C. Beaulieu, X. Lei, "Switching rates of selection diversity and switch-and stay diversity on mixed high-speed train channels", In *Proc. IEEE International Workshop on High Mobility Wireless Communications*, Xi'an, China, 21-23 Oct. 2015, pp 51-55.
- [8] Nakagami, M., "The m-distribution-A general formula of intensity distribution of rapid fading", *Statistical Method of Radio Propagation*, 1960.
- [9] S. Panic, M. Stefanovic, J. Anastasov, P. Spalevic, *Fading and Interference Mitigation in Wireless Communications*. CRC Press: New York, 2013.
- [10] I. S. Gradshteyn, I. M. Ryzhik, *Table of Integrals, Series, and Products*, 6<sup>th</sup> ed., New York: Academic, 2000.
- [11] G.L. Stüber, *Principles of Mobile Communications*, Kluwer Academic Publishers: Massachusetts, USA, 1996.
- [12] Xekalaki E., Panaretos J., Psarakis S. (2003) A predictive model evaluation and selection approach—the correlated gamma ratio distribution. In: Panaretos J. (Ed.) *Stochastic musings: Perspectives from the pioneers of the late 20th century*. Psychology Press, USA, pp. 188–202.

# Outage Performance of Satellite-Terrestrial Multiuser Networks with Fixed Gain AF Relay

Aleksandra M. Cvetkovic<sup>1</sup>, Jelena A. Anastasov<sup>1</sup> and Nemanja M. Zdravkovic<sup>1</sup>

**Abstract** – We investigate the outage performance of satellite-terrestrial multiuser network with single fixed gain amplify-and-forward relay. Driven by opportunistic scheduling strategy, new expression for the cumulative distribution function of the highest end-to-end signal-to-noise ratio (SNR) is derived. Based on this result, the outage probability of the system over shadowed Rician fading satellite-relay link and Rayleigh fading relay-user channels, is evaluated. Compact numerical results, analyzing the impact of number of service users, shadowing effects and the average SNR values, are also given together with independently performed simulations.

**Keywords** – hybrid satellite-terrestrial network, multiusers, opportunistic scheduling, outage probability, relaying technology

## I. INTRODUCTION

Deployment of satellite systems over challenging wireless environments is a great solution in providing high data rates, extended coverage area and variety of services for terrestrial users [1]. Disaster recovery is one of the important issues that interpreted basic utilization of geostationary satellites for broadband communications. Due to the fact that the line-of-sight (LoS) links between the satellite and terrestrial destination points are often blocked by bad weather conditions or surrounding obstacles, the usage of relaying technology is required. Among the different cooperative strategies, amplify-and-forward (AF) is the most effective relaying strategy according to its low complexity, sufficient performance gain and low implementation cost [2].

The main concept of relaying technology refers to the scenario with point-to-point link configuration having a single source and a single destination terminal. In order to establish communication between the source and many remote or scattered multiusers, point-to-multipoint dual-hop AF configuration can be utilized [3-4]. In point-to-multipoint multiuser applications, selecting the mobile user with the strongest channel i.e. opportunistic scheduling strategy can provide larger performance gains.

In open technical literature, there are numerous published papers of end-to-end performance of hybrid satellite-terrestrial relay systems over different fading or shadowed fading environments. Capitalizing on the analysis over shadowed fading links, various performance evaluations of dual-hop AF systems with a satellite are proposed. In [5] focusing on the AF relaying over shadowed Rician fading paths, the derived analytical results are given in the form of infinite series. In

[6], the hybrid land mobile satellite (LMS) system with satellite-relay shadowed Rician link and a relay-to-destination free space optical link was analysed.

Also, the hybrid satellite-terrestrial networks with single user scenario have been obtained in [7-10]. A LMS system with the source-to-relay path under shadowed Rician and the relay-destination link under Nakagami- $m$  fading conditions was investigated in [7] and [8]. In [7], the outage probability performance of hybrid networks was provided. The error probability evaluations of the similar hybrid systems were proposed in [9]–[10]. Since future development of satellite communications implies servicing a number of terrestrial users, the multiuser relaying techniques were also considered in related works. The performance analysis of hybrid multiuser system has been reported in [11]. The analytical expressions for evaluating ergodic capacity and the outage probability (in the form of infinite series) of multiuser system applying channel-state-information based AF relay, were presented.

In this paper, we propose the outage probability performance evaluation of hybrid satellite terrestrial multiuser network with fixed gain relay. The analysis is performed under the assumption that satellite-relay channel can be modeled by shadowed Rician distribution and relay-to mobile users paths as Rayleigh fading channels. Novel analytical expression is derived in closed-form for integer values of shadowing parameter  $m$ , which describes LoS component of satellite-relay link. From practical point of view, this result can be used in many real scenarios obtaining required engineering accuracy.

## II. SYSTEM MODEL

Dual-hop communication between satellite and multiple users is performed via fixed gain AF relay, as shown in Fig. 1. The satellite broadcasts the signal to the relay in the first transmission phase, the relay amplifies the signal and in the second phase relay resends the signal to  $N$  users. The satellite transmits the modulated signal  $s(t)$  over the link between the satellite and relay with channel complex fading coefficient  $h_1$ .

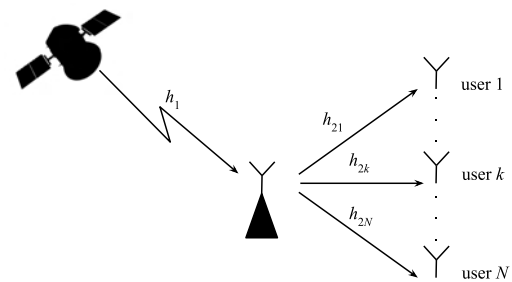


Fig. 1. System model for hybrid satellite-terrestrial multiuser system

<sup>1</sup>Aleksandra M. Cvetković, Jelena A. Anastasov and Nemanja Zdravkovic, are with the University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mails: [aleksandra.cvetkovic@elfak.ni.ac.rs](mailto:aleksandra.cvetkovic@elfak.ni.ac.rs), [jelena.anastasov@elfak.ni.ac.rs](mailto:jelena.anastasov@elfak.ni.ac.rs), [nemanja.zdravkovic@elfak.ni.ac.rs](mailto:nemanja.zdravkovic@elfak.ni.ac.rs)



The received signal at the relay is given by

$$y_R = \sqrt{E_s} h_1 s(t) + n_R, \quad (1)$$

where  $E_s$  denotes the source average energy and  $n_R$  is additive white Gaussian noise (AWGN) with variance  $\sigma_R^2$ . The signal at the relay is multiplied by a gain  $G$ . The relay transmits amplified signal to the  $k$ -th user over channel with complex fading coefficient  $h_{2k}$ . The received signal at the  $k$ -th user is expressed as

$$y_k = \sqrt{E_R} h_{2k} G \left( \sqrt{E_s} h_1 s(t) + n_R \right) + n_{Dk} \quad (2)$$

where  $E_s$  denotes the relay output average energy and  $n_{Dk}$  is AWGN at the  $k$ -th user with variance  $\sigma_{Dk}^2$ .

The fixed gain relay is implemented in the system so the gain depends on the statistical channel state information of the satellite-relay link in a way

$$G^2 = \frac{E_R}{C \sigma_R^2} = \frac{E_R}{\sigma_R^2 + E[E_s |h_1|^2]} = \frac{E_R}{\sigma_R^2 (1 + \bar{\gamma}_1)}, \quad (3)$$

where  $C$  is a constant determined by the relay gain and  $\bar{\gamma}_1$  is average signal-to-noise ratio (SNR) of the satellite-relay link ( $E[\cdot]$  denotes expectation).

The instantaneous SNR at the  $k$ -th user can be determined as

$$\gamma_{eqk} = \frac{E_s |h_1|^2 E_R |h_{2k}|^2 G^2}{E_R |h_{2k}|^2 G^2 \sigma_R^2 + \sigma_{Dk}^2}. \quad (4)$$

With the introduction of instantaneous SNR of the satellite-relay and relay-user link as  $\gamma_1 = E_s |h_1|^2 / \sigma_R^2$  and  $\gamma_{2k} = E_R |h_{2k}|^2 / \sigma_{Dk}^2$  respectively, and substituting (3) in (4), the equivalent SNR at  $k$ -th user has the form [4]

$$\gamma_{eqk} = \frac{\gamma_1 \gamma_{2k}}{\gamma_{2k} + C}. \quad (5)$$

From (3), the constant  $C$  is  $C = 1 + \bar{\gamma}_1$ .

### III. CHANNEL MODEL

We assume that a hybrid satellite-terrestrial system, from the satellite to the users, consists of two links. First, LoS link between satellite and relay experiences Rice fading where the LoS amplitude fluctuation is modeled by Nakagami- $m$  distribution. The relay-user links are assumed to follow Rayleigh fading.

#### A. Satellite-relay channel

The shadowed Rician distribution was proposed in [12] for modeling the signal amplitude over LMS channel and the probability density function (PDF) of channel gain has a form

$$p_{|h_1|^2}(x) = \alpha \exp(-\beta x) {}_1F_1(m; 1; \delta x). \quad (6)$$

where  ${}_1F_1(\cdot; \cdot; \cdot)$  denotes the confluent hypergeometric function [13, (9.210.1)], and parameters  $\alpha$ ,  $\beta$  and  $\delta$  can be calculated as

$$\alpha = \frac{1}{2b} \left( \frac{2bm}{2bm + \Omega} \right)^m, \quad \beta = \frac{1}{2b}, \quad \delta = \frac{\Omega}{2b(2bm + \Omega)}, \quad (7)$$

wherein  $2b$  is the average power of the multipath components,  $\Omega$  is average power of the LoS component and  $m$  is Nakagami- $m$  parameter.

The PDF of the instantaneous SNR,  $\gamma_1 = E_s |h_1|^2 / \sigma_R^2$ , is [11]

$$p_{\gamma_1}(x) = \frac{\alpha}{E_s / \sigma_R^2} \exp\left(-\beta \frac{x}{E_s / \sigma_R^2}\right) {}_1F_1\left(m, 1, \delta \frac{x}{E_s / \sigma_R^2}\right). \quad (8)$$

Average SNR in this case has the following form

$$\bar{\gamma}_1 = \frac{E[|h_1|^2]}{\sigma_R^2} E_s = (2b + \Omega) \frac{E_s}{\sigma_R^2}. \quad (9)$$

For integer values of the fading parameter  $m$  the function  ${}_1F_1(\cdot; \cdot; \cdot)$  using [14, (07.20.03.0009.01)] and [14, (05.02.06.0005.01)] can be presented in a simple mathematical form

$${}_1F_1(m; 1; z) = e^z \sum_{k=0}^{m-1} \binom{m-1}{m-1-k} \frac{1}{k!} z^k. \quad (10)$$

Substituting (10) into (8) the PDF can be written as

$$p_{\gamma_1}(\gamma_1) = \frac{\alpha}{E_s / \sigma_R^2} e^{-\frac{\beta-\delta}{E_s / \sigma_R^2} \gamma_1} \sum_{k=0}^{m-1} \frac{1}{k!} \binom{m-1}{m-1-k} \left( \frac{\delta}{E_s / \sigma_R^2} \right)^k \gamma_1^k. \quad (11)$$

The expression of the PDF given by (11) only applies to integer values of  $m$ , and it can be used to determine the upper and lower bounds of system performance for any value of parameter  $m$ .

Based on (11), the cumulative density function (CDF) can be determined using [13, (3.351.1)]

$$F(\gamma) = \alpha \sum_{k=0}^{m-1} \frac{1}{k!} \binom{m-1}{m-1-k} \frac{\delta^k}{(\beta - \delta)^{k+1}} \times \left( \Gamma(k+1) - \Gamma\left(k+1, \gamma \frac{\beta - \delta}{E_s / \sigma_R^2}\right) \right), \quad (12)$$

where  $\Gamma(\cdot)$  is the gamma function [13, (8.310.1)] and  $\Gamma(\cdot, \cdot)$  is the incomplete gamma function [13, (8.350.2)].

#### B. Relay-users channels

Multiple users are linked to the relay over Rayleigh fading channels. The corresponding PDF and CDF of the instantaneous SNR over Rayleigh fading channel between relay and  $k$ -th user has a form

$$p_{\gamma_{2k}}(\gamma_2) = \frac{1}{\bar{\gamma}_2} \exp\left(-\frac{\gamma_2}{\bar{\gamma}_2}\right) \text{ and } F_{\gamma_{2k}}(\gamma) = 1 - \exp\left(-\frac{\gamma}{\bar{\gamma}_2}\right) \quad (13)$$

respectively. We assume that all average SNRs are equal to  $\bar{\gamma}_2 = E_R |h_{2k}|^2 / \sigma_{Dk}^2$ .

Opportunistic multiuser scheme is adopted and the relay selects the user which channel has the highest instantaneous SNR, i.e.  $\gamma_2 = \max_{1 \leq k \leq N} \gamma_{2k}$ . Then, the CDF of selected end-to-end SNR is

$$F_{\gamma_2}(\gamma) = \prod_{k=1}^N F_{\gamma_{2k}}(\gamma). \quad (14)$$

In the case of accepting an opportunistic protocol in relay-multiuser communication, the equivalent SNR can be defined as in (5), wherein  $\gamma_{2k}$  is replaced by  $\gamma_2$ , i.e.  $\gamma_{eq} = \gamma_1 \gamma_2 / (\gamma_2 + C)$ .

#### IV. OUTAGE PERFORMANCE EVALUATION

Outage probability is one of the performance measures that can be useful in system design. The outage performance can be obtained as the probability that the end-to-end SNR falls below a threshold.

The outage probability for hybrid system with fixed gain AF relay can be evaluated in the following way

$$F_{\gamma_{eq}}(\gamma_{th}) = \int_0^{\gamma_{th}} P_r \left( \gamma_2 > \frac{C\gamma_{th}}{\gamma_1 - \gamma_{th}} \middle| \gamma_1 \right) p_{\gamma_1}(\gamma_1) d\gamma_1 + \int_{\gamma_{th}}^{\infty} P_r \left( \gamma_2 < \frac{C\gamma_{th}}{\gamma_1 - \gamma_{th}} \middle| \gamma_1 \right) p_{\gamma_1}(\gamma_1) d\gamma_1. \quad (16)$$

The closed-form solution for the first integral, namely for

$$I_1 = \int_0^{\gamma_{th}} P_r \left( \gamma_2 > \frac{C\gamma_{th}}{\gamma_1 - \gamma_{th}} \middle| \gamma_1 \right) p_{\gamma_1}(\gamma_1) d\gamma_1 = \int_0^{\gamma_{th}} p_{\gamma_1}(\gamma_1) d\gamma_1, \quad (17)$$

is given by (12).

Substituting (11) and (14) in the second integral of (16), under the assumption of equal average SNRs over all relay-users channels, we get

$$I_2 = \int_{\gamma_{th}}^{\infty} \left( 1 - e^{-\frac{C\gamma_{th}}{\gamma_2(\gamma_1 - \gamma_{th})}} \right)^N p_{\gamma_1}(\gamma_1) d\gamma_1 = \frac{\alpha}{E_s / \sigma_R^2} \sum_{k=0}^{m-1} \frac{1}{k!} \binom{m-1}{m-1-k} \times \frac{1}{k!} \left( \frac{\delta}{E_s / \sigma_R^2} \right)^k \int_{\gamma_{th}}^{\infty} \left( 1 - e^{-\frac{C\gamma_{th}}{\gamma_2(\gamma_1 - \gamma_{th})}} \right)^N e^{-\frac{\beta-\delta}{E_s / \sigma_R^2} \gamma_1} \gamma_1^k d\gamma_1. \quad (18)$$

Applying binomial expansion and introducing a change of variables,  $x = \gamma_1 - \gamma_{th}$  in (18), the integral  $I_2$  becomes

$$I_2 = \frac{\alpha}{E_s / \sigma_R^2} e^{-\frac{\beta-\delta}{E_s / \sigma_R^2} \gamma_{th}} \sum_{k=0}^{m-1} \sum_{l=0}^N \sum_{p=0}^k \frac{1}{k!} \binom{m-1}{m-1-k} \binom{N}{l} \binom{k}{p} \times (-1)^l \left( \frac{\delta}{E_s / \sigma_R^2} \right)^k \gamma_{th}^{k-p} \int_{\gamma_{th}}^{\infty} e^{-\frac{Cl\gamma_{th}}{\gamma_2 x}} e^{-\frac{\beta-\delta}{E_s / \sigma_R^2} x} x^p dx. \quad (19)$$

For  $l=0$ ,  $I_2$  can be solved in closed-form as

$$I_2^{l=0} = \frac{\alpha}{E_s / \sigma_R^2} e^{-\frac{\beta-\delta}{E_s / \sigma_R^2} \gamma_{th}} \sum_{k=0}^{m-1} \sum_{p=0}^k \frac{1}{k!} \binom{m-1}{m-1-k} \binom{k}{p} \left( \frac{\delta}{E_s / \sigma_R^2} \right)^k \gamma_{th}^{k-p} \left( \frac{\beta-\delta}{E_s / \sigma_R^2} \right)^{-1-p} \Gamma(p+1), \quad (20)$$

while for  $l \neq 0$  the closed-form solution of  $I_2$  has a following form

$$I_2^{l \neq 0} = \frac{2\alpha}{E_s / \sigma_R^2} e^{-\frac{\beta-\delta}{E_s / \sigma_R^2} \gamma_{th}} \sum_{k=0}^{m-1} \sum_{l=1}^N \sum_{p=0}^k \frac{1}{k!} \binom{m-1}{m-1-k} \binom{N}{l} \binom{k}{p} (-1)^l \left( \frac{\delta}{E_s / \sigma_R^2} \right)^k \left( \frac{Cl\gamma_{th} E_s / \sigma_R^2}{\gamma_2 (\beta-\delta)} \right)^{\frac{1+p}{2}} \gamma_{th}^{k-p} K_{-1-p} \left( 2 \sqrt{\frac{Cl\gamma_{th} (\beta-\delta)}{\gamma_2 E_s / \sigma_R^2}} \right) \quad (21)$$

with  $K_n(\cdot)$  denoting the modified Bessel function of the second kind [13, (8.407.1)]

Finally, the expression for outage probability of the hybrid satellite-terrestrial multiuser system with fixed gain AF relay can be evaluated as

$$P_{out}(\gamma_{th}) = \alpha \sum_{k=0}^{m-1} \frac{1}{k!} \binom{m-1}{m-1-k} \frac{\delta^k}{(\beta-\delta)^{k+1}} \times \left( \Gamma(k+1) - \Gamma\left(k+1, \gamma_{th} \frac{\beta-\delta}{E_s / \sigma_R^2}\right) \right) + \frac{\alpha}{E_s / \sigma_R^2} e^{-\frac{\beta-\delta}{E_s / \sigma_R^2} \gamma_{th}} \sum_{k=0}^{m-1} \sum_{p=0}^k \frac{1}{k!} \binom{m-1}{m-1-k} \binom{k}{p} \times \left( \frac{\delta}{E_s / \sigma_R^2} \right)^k \gamma_{th}^{k-p} \left( \frac{\beta-\delta}{E_s / \sigma_R^2} \right)^{-1-p} \Gamma(p+1) + \frac{2\alpha}{E_s / \sigma_R^2} e^{-\frac{\beta-\delta}{E_s / \sigma_R^2} \gamma_{th}} \sum_{k=0}^{m-1} \sum_{l=1}^N \sum_{p=0}^k \frac{1}{k!} \binom{m-1}{m-1-k} \binom{N}{l} \binom{k}{p} (-1)^l \left( \frac{\delta}{E_s / \sigma_R^2} \right)^k \left( \frac{Cl\gamma_{th} E_s / \sigma_R^2}{\gamma_2 (\beta-\delta)} \right)^{\frac{1+p}{2}} \times \gamma_{th}^{k-p} K_{-1-p} \left( 2 \sqrt{\frac{Cl\gamma_{th} (\beta-\delta)}{\gamma_2 E_s / \sigma_R^2}} \right). \quad (22)$$

Novel expression (22) is derived in closed-form solution. Regardless it looks quite robust, numerical computation of (22) in many of softwares like *Matlab* or *Mathematica* is easy, because it contains simple and well-tractable mathematical functions.

#### V. NUMERICAL RESULTS

In this section we present outage probability numerical and simulation results for hybrid satellite-terrestrial multiuser system. Numerical results are obtained based on the derived analytical results (22) and accompanied with independent simulation results.

In order to analyze the impact of fading and shadowing effects on outage probability we assume the heavy and average shadowing conditions over the satellite link. For heavy shadowing condition the parameters are set as  $\{m, b, \Omega\} = \{1, 0.063, 0.0007\}$  and for average shadowing as  $\{m, b, \Omega\} = \{5, 0.251, 0.279\}$  [15].

Outage probability for different values of  $E_s/\sigma_R^2$  as a function of  $\bar{\gamma}_2$ , is shown in Fig. 2. Relying on obtained results we can conclude that for a certain value of  $E_s/\sigma_R^2$  in the low range of  $\bar{\gamma}_2$ , increasing of  $\bar{\gamma}_2$  leads to outage probability decreasing. However, the outage floor appears at high values of  $\bar{\gamma}_2$ . We can also notice that this floor is only defined by the value of  $E_s/\sigma_R^2$  and is totally independent of the number of users.

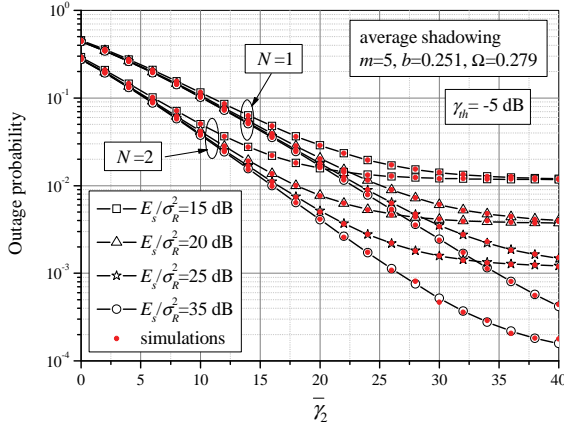


Fig. 2. Outage probability versus average relay-user SNR for various  $E_s/\sigma_R^2$  and users number

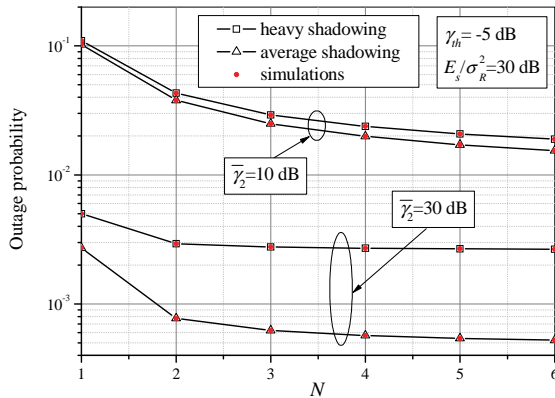


Fig. 3. Outage probability versus number of users

Outage probability versus the number of users in the hybrid multiuser system is shown in Fig. 3. Outage performance is presented for heavy and average shadowing condition over satellite link and for different average SNR of relay-user channels. With the increase of number of users, the performance gain is decreased. This gain is the greatest when we increase the number of users from one to two, when we select the user with the best relay-user channel. By increasing the average relay-user SNR, the effect of shadowing on outage probability is more pronounced. For example, for  $N = 4$  and  $\bar{\gamma}_2 = 10$  dB, the outage probability in heavy shadowing conditions is 0.02374, while in average shadowing conditions is the same order of magnitude, 0.01987. For  $\bar{\gamma}_2 = 30$  dB, changes in shadowing conditions cause the change in probability in almost of an order of magnitude (from

$5.69392 \times 10^{-4}$  to  $2.71 \times 10^{-3}$ , for the same change of conditions).

## VI. CONCLUSION

In this paper, outage performance for hybrid satellite-terrestrial multiuser network with the fixed gain AF relay is analyzed. Analytical expression for outage probability is derived for integer fading parameter value of shadowed LoS component over satellite channels. The obtained results are in excellent agreement with the simulation results which were performed independently. We showed the effect of shadowing over the satellite-relay link, number of users, and average SNR values on the overall outage system performance.

## ACKNOWLEDGEMENT

This work was supported by Ministry of science and technology development of Republic of Serbia TR-32052.

## REFERENCES

- [1] P. Chini, G. Giambene, and S. Kota, "A survey on mobile satellite systems," *Int. J. Satellite Commun. Netw.*, vol. 28, no. 1, pp. 29–57, January/February, 2009.
- [2] M. Dohler, and Y. Li, *Cooperative Communications: hardware, channel & phy*, John Wiley & Sons, United Kingdom, 2010.
- [3] C.-J. Chen and L.-C. Wang, "A unified capacity analysis for wireless systems with joint multiuser scheduling and antenna diversity in Nakagami fading channels," *IEEE Trans. Commun.*, vol. 54, no. 3, pp. 469–478, March, 2006.
- [4] N. Yang, M. ElKashlan, and J. Yuan, "Outage probability of multiuser relay networks in Nakagami- $m$  fading channels," *IEEE Trans. Veh. Technol.*, vol. 59, no. 5, pp. 2120–2132, June, 2010.
- [5] A. M. Cvetkovic, J. Anastasov, S. Panic, M. Stefanovic, and D. Milic, "Performance of dual-hop relaying over shadowed Ricean fading channels," *J. Elect. Eng.*, vol. 62, no. 4, pp. 244–248, July, 2011.
- [6] M. R. Bhatnagar and M. K. Arti, "Performance analysis of hybrid satellite terrestrial FSO cooperative system," *IEEE Photon. Technol. Lett.*, vol. 25, no. 22, pp. 2197–2200, September, 2013.
- [7] A. Iqbal and K. M. Ahmed, "A hybrid satellite-terrestrial cooperative network over non identically distributed fading channels," *J. Commun.*, vol. 6, no. 7, pp. 581–589, October, 2011.
- [8] M. R. Bhatnagar and M. K. Arti, "Performance analysis of AF based hybrid satellite-terrestrial cooperative network over generalized fading channels," *IEEE Commun. Lett.*, vol. 17, no. 10, pp. 1912–1915, October, 2013.
- [9] K. An, M. Lin, J. Ouyang, Y. Huang, and G. Zheng, "Symbol error analysis of hybrid satellite-terrestrial cooperative networks with co-channel interference," *IEEE Commun. Lett.*, vol. 18, no. 11, pp. 1947–1950, November, 2014.
- [10] M. K. Arti and M. R. Bhatnagar, "Beamforming and combining in hybrid satellite-terrestrial cooperative systems," *IEEE Commun. Lett.*, vol. 18, no. 3, pp. 483–486, March, 2014.
- [11] K. An, M. Lin, and T. Liang, "On the Performance of Multiuser Hybrid Satellite-Terrestrial Relay Networks With Opportunistic Scheduling," *IEEE Commun. Lett.*, vol. 19, no. 10, pp. 1722–1725, October 2015.
- [12] A. Abdi, W. Lau, M.-S. Alouini, and M. Kaveh, "A new simple model for land mobile satellite channels: first and second order statistics," *IEEE Trans. Wireless Commun.*, vol. 2, no. 3, pp. 519–528, May 2003.
- [13] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed. San Diego, CA, USA: Academic, 2007.
- [14] The Wolfarm Functions Site, 2008. [Online] Available: <http://functions.wolfarm.com>
- [15] N. I. Miridakis, D. D. Vergados, and A. Michalas, "Dual-Hop Communication Over a Satellite Relay and Shadowed Rician Channels," *IEEE Trans. Veh. Technol.*, vol. 64, no. 9, pp. 4031–4040, September 2015

# Responsive Web Design in Online-CADCOM platform

Ognyan Chikov<sup>1</sup>

**Abstract** – In this paper the software development of Responsive Web Design layout in Online-CADCOM platform is presented. The approach for “responsive” design of pages layout in Online-CADCOM is presented, as well as approach for “responsive” design in Content Management System pages’ layout. Challenges and benefits of Responsive Web Design in Online-CADCOM are described.

**Keywords** – Software development, Responsive Web Design, Layout, Responsive Elements, Styles.

## I. INTRODUCTION

Online-CADCOM platform with included Content Management System (CMS) is an online assisted platform for computer-aided design in communications that is designed to work not only with standard computer devices with unified resolutions. Furthermore, the platform is designed to work with different devices and Operating Systems (OS) and large scale of resolutions. This approach gives a reliable method to manage all materials and resources that Online-CADCOM stores and it is frequently called “responsive” design, which becomes more important in World Wide Web because the amount of traffic from non-standard devices (desktop computers and notebooks) is half of total internet traffic [1]. Responsive Web Design is an approach for making web designs provide an optimal viewing and user interaction experience – easy reading and navigation with a minimum of resizing, panning and scrolling which is very important for Online-CADCOM due to the platform complexity and large number of online content. The platform is designed to adapt the layout to the viewing environment by using fluid, proportion-based grids, flexible images and CSS3 (Cascading Style Sheets) media queries and media rules that are dynamically adapted according to the used device. Fluid grid concept is a method for resizing and adapting the layout to the viewing environment in relative units like percentages, rather than the traditional method that uses pixels and points [1] which cannot be applied in the platform due to the large amount of different type of content. Also flexible images in Online-CADCOM are resized in relative units, so as to prevent them from displaying outside their containing element. Online-CADCOM also uses media queries that allow the platform to use different CSS style rules based on characteristics of the devices that the platform is displaying.

First challenges and benefits of using Responsive Web Design are presented, and then responsive build of both

Online-CADCOM platform and Content Management System is presented.

## II. RESPONSIVE WEB DESIGN DEVELOPMENT IN ONLINE-CADCOM

Responsive Web Design in Online-CADCOM platform was developed with the rule “keep it simple” and it fits literally all screen sizes no matter if it is desktop, laptop, tablet, smartphone or wearable in landscape or portrait mode. With this kind of design, there is need to maintain only one website because while the layout changes, the content stays the same across different devices and platforms. In standard approach, different variations of one platform are created for different layouts, which cost development time and money. That’s why this approach is not used in developing the Responsive Web Design in the platform.

One of the biggest challenges related with creating Online-CADCOM Responsive Web Design is to be highly user friendly and in the same time to provide as much information as the full layout provide. The platform uses consistent style and consistent content because users don’t expect it to be different just because they use a different device and operating system and this method solves the problem that a lot of web platform has – limited functionalities, lack of information and missing content in “responsive” designs.

Redirects in the platform do not exist in fact that for all devices same page URLs (Uniform Resource Locator) deliver content to all users that use the platform. This method eliminates the problem with redirections to different pages and incompatibilities between different devices and ensures that a link can be certain access directly, no matter from where or what kind of device. This solves the problem with possible email sending from mobile device when a link is open and the need of device switching for this operation.

Load times in Online-CADCOM, when the platform is loaded through a mobile device are significantly decreased because of Responsive Web Design methods for layout, media and media-queries adaptations to the browsing devices. This approach saves the time needed for the platform pages to be loaded and also the 3G and 4G traffic bytes of mobile users.

Another challenge of developing the Responsive Web Design is that older browsers do not recognize media-queries, especially Internet Explorer version 8 and older. Currently about 14% of web users worldwide still use IE8 in accordance with [2]. These 14% are potential users portion of Online-CADCOM audience, which is a reason to solve the problem. Online-CADCOM has additional CSS source code for proper displaying in IE8 and older versions, which makes the platform completely multi usable not only with new trends in web browsing, but also with old methods and versions.

It is not a secret that when a flexible image is resized and scaled in relative units, it quickly loses details and thus its

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



meaning. This challenge in creating Responsive Web Design in Online-CADCOM was solved with using both for normal and “responsive” layout – vector pictures with low pixel size and in the same time with high dots detail. This gives to the “responsive” design a quality near to the ordinary layout.

### III. SOFTWARE DEVELOPMENT OF ONLINE-CADCOM FRONT VIEW RESPONSIVE LAYOUT

Online-CADCOM platform has managed the CAD tools based on Daniel-Gajsky Y model of design levels and has two main panels – Panel 1 and Panel 2 equipped with buttons related to two design areas from the Y model and interconnections between them [4]. Also each tool is characterized and a tool passport is created. Besides the two panels, Online-CADCOM develops and supports a lot of different content [6]. This large amount of different type of content should be properly described and scaled for responsive view in different devices. Even more, this content should be presented properly to the users with all specifications and functions which offer and user should be able to use their full functions. In Online-CADCOM this approach is achieved as exporting all code rules and adjustments in a file in the “front\_end” folder with name – “style.css”. This file contains all colors, element specifications and placement and also “responsive” rules for different devices.

First and most difficult part of developing a “responsive” layout for Online-CADCOM is adjusting the screen resolution of the platform, when a request is placed. Most of the smart devices offer a function for switching from portrait to landscape orientation when the device is rotated. This changes the parameters of the screen and visible part and often can break the layout of the platform. In Online-CADCOM the method used for this situation is to define all rules for displaying platform elements in according to specific resolution. Code bellow shows how to properly define rules for different device categorization based on resolution using media queries. Media queries are simple filters that are applied to Online-CADCOM CSS styles. They make it easy to change platform styles based on the characteristics of the device rendering the content, including the display type, width, height, orientation and even resolution. They enabled the platform to create a responsive experience, where specific styles are applied to small screens, large screens and anywhere in between. Online-CADCOM styles uses min-width tags instead of old and deprecated min-device-width. The first tag is based on the size of the browser window, whereas min-device-width is based on the size of the screen. Unfortunately, some browsers, including the legacy Android browser may not report the device width properly and instead report the screen size in device pixels instead of the expected viewport width. Sample code is shown below:

```
/* Small devices (tablets, 768px and up) */
@media (min-width:768px) {
    .fixed-header-on.boxed .header.fixed {
        width: 750px;
```

```
    }
}
/* Medium devices (desktops, tablets, 992px and up) */
@media (min-width:992px) {
    .fixed-header-on .header .logo {
        display: block;
    }
    .fixed-header-on.boxed .header.fixed {
        width: 970px;
    }
}
/* Large devices (Large desktops 1200px and up) */
@media (min-width:1200px) {
    .fixed-header-on.boxed .header.fixed {
        width: 1170px;
    }
}
/* Small devices (tablets, phones less than 767px) */
@media (max-width:767px) {
    /* Fixed header position */
    .header.fixed {
        position: relative !important;
        top: auto !important;
    }
}
```

Device categorization and properly adjusting screen resolution of the platform is not the single step to properly develop Responsive Web Page. Online-CADCOM is optimized for a variety of devices and include a meta viewport element in the head of the “style.css” document. A meta viewport tag gives the browser instructions on how to control the page's dimensions and scaling. The problem is that to attempt to provide the best experience, mobile browsers will render the platform page at a desktop screen width (usually about 980px, though this varies across devices), and then try to make the content look better by increasing font sizes and scaling the content to fit the screen [3]. This is strictly true as regards to Apple devices. For users of the platform, this means that font sizes may appear inconsistently and they have to double-tap or pinch-to-zoom in order to see and interact with the navigation panels and content. Additional rule must be provided for the described case. The code bellow shows, how this malfunction is solved in the platform:

```
<meta name="viewport" content="width=device-width,
initial-scale=1">
```

Using the meta viewport value width=device-width instructs the platform pages to match the screen's width in device-independent pixels. This allows the page to reflow content to match different screen sizes, whether rendered on a small mobile phone or a large desktop monitor. Some browsers will keep the page's width constant when rotating to landscape mode, and zoom rather than reflow to fill the screen. Adding the attribute initial-scale=1 instructs browsers to establish a 1:1 relationship between CSS pixels and device-independent pixels regardless of device

orientation, and allows the platform pages to take advantage of the full landscape width.

Part of the responsive layout of Online-CADCOM first page is presented on Figure 1.

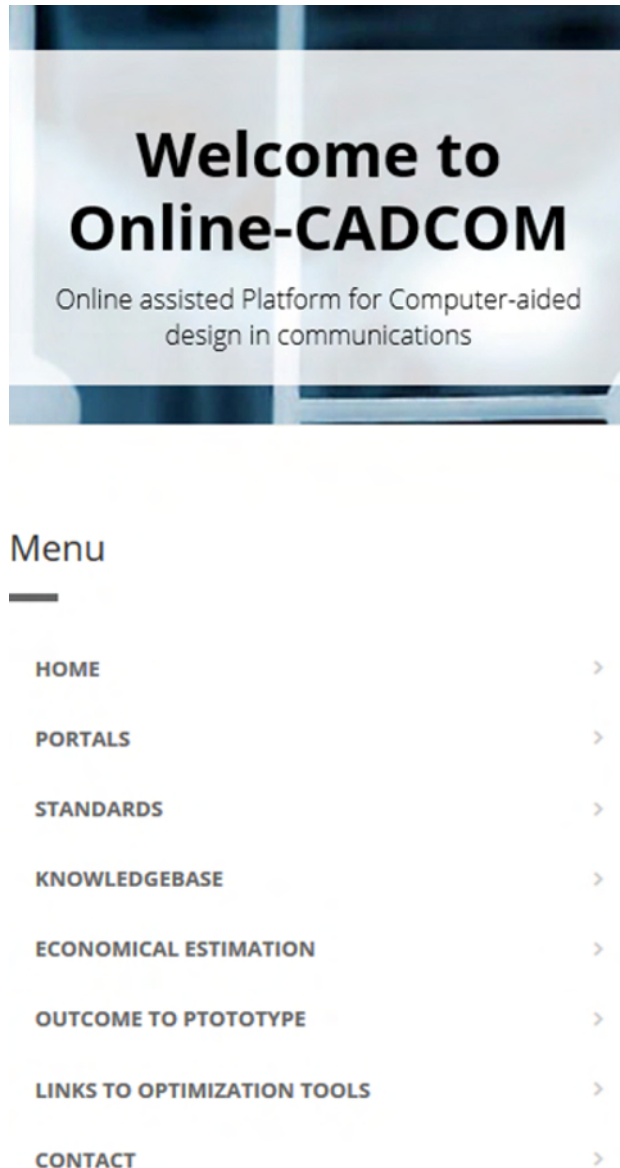


Fig. 1. Responsive layout of Online-CADCOM first page.

#### IV. SOFTWARE DEVELOPMENT OF CONTENT MANAGEMENT SYSTEM FRONT VIEW RESPONSIVE LAYOUT

Content Management System is part of Online-CADCOM which enables dynamically adding content without using any programming skills [5, 6]. This requires the development of Responsive Web Design of Content Management System also. Like the above-described method for Online-CADCOM loading styles method, also CMS platform has its own folder

with name “front\_end” that stores the file with all styling rules – “style.css”. Furthermore, in CMS platform we have two more additional style files that take care about the breakpoints. CMS uses breakpoints to define a point to which content will respond to provide the users with the best possible layout to consume the information in the platform. The approach used for creating the Content Management System is – “mobile first” [3]. Approaching the design “mobile first” is the best approach for a responsive design [3]. It allows CMS to layout the content that is most important to users in a clear and logical way on the smallest screen real-estate. This helps to identify what is the most important information on the platform and how to be properly displayed [7]. CMS content was designed to fit on a small screen size first, and then the screen was expanding until a breakpoint becomes necessary. This optimizes breakpoints based on content and maintains the fewest number of breakpoints possible. The first step of designing the layout is to fit good on small screen with size of 320x480 pixels [8]. Next the browser is resized until there is too much white space between the elements and they do not look good. The decision is somewhat subjective, but above 600px is certainly too wide for the platform instruments for content management. This requires the creation of two additional files for breakpoints like described above. Index page html contains the following code for inserting the style files:

```
<link rel="stylesheet" href="style.css">
<link rel="stylesheet" media="(max-width:600px)"
href="style-2-small.css">
<link rel="stylesheet" media="(min-width:601px)"
href="style-2-large.css">
```

Common platform styles such as fonts, icons, basic positioning and colors are placed in style.css. Specific layouts for the small screen are then placed in style-2-small.css and large screen styles are placed in style-2-large.css. CMS platform has good responsive layout because between the major breakpoints, it has adjustments for margins, paddings and font size which make feel the platform more natural in the layout.

Screen size of devices is not a definitive of what a user may want to see or use [9]. That’s why CMS platform does not hide content from the standard layout just because it can’t fit on screen. Instead, each element layout is strictly designed for responsive view, even text. Optimize text for reading is another “responsive” task that is completely done with creating a CMS platform with “responsive” design. Classic readability theory suggests that an ideal column should contain 70 to 80 characters per line (about 8 to 10 words in English) [3]. Thus each time the width of a text block grows past about 10 words, a breakpoint should be considered. This means that the font of CMS will work perfectly on smaller screens giving 10 words per line, but larger screens will require a breakpoint. In this case, if the browser width is greater than 575px, the ideal content width is 550px. Code fragment showed below is for extra small devices layout:

```
@media (max-width: 575px) {
```

```

article {
width: 550px;
margin-left: auto;
margin-right: auto;
}
}

```

Figure 2 shows how the “Dashboard” of CMS looks. It is part of the screen dynamically generated on screen with size 560x640 pixels:

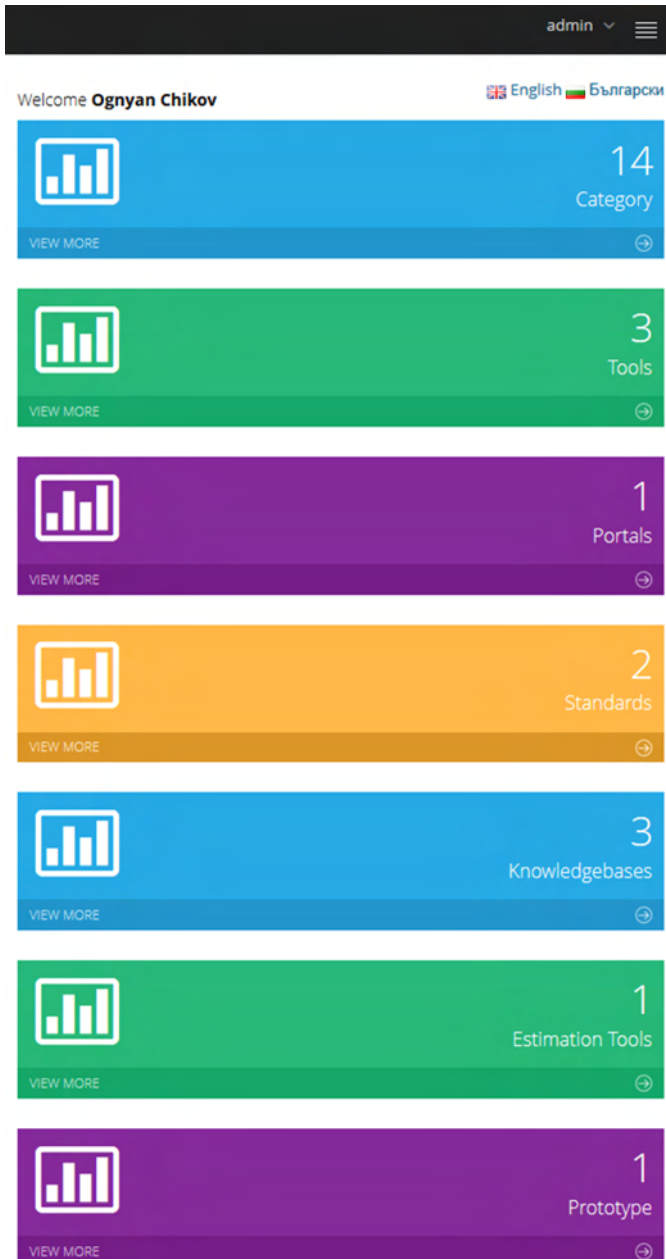


Fig. 2. Responsive layout of Content Management System Dashboard.

## V. CONCLUSION

Online-CADCOM platform with included Content Management System offers to users a completely “responsive” layout for perfect user experience when using the platform. The solution consists of a mix of flexible grids and layouts, images and an intelligent use of CSS media queries which dynamically change the content according to user device specifications. The approach and methods used ensure proper content visualization, easy content management and provides highly effective usability for the platform itself.

## ACKNOWLEDGEMENT

This research is partly supported by Project №152PD0039-07 in Technical University – Sofia.

## REFERENCES

- [1] Responsive Web Design  
[https://en.wikipedia.org/wiki/Responsive\\_web\\_design](https://en.wikipedia.org/wiki/Responsive_web_design)
- [2] Usage report for Internet Explorer “Internet Explorer is back from the dead”  
<http://www.webdesignerdepot.com/2016/01/ie8-is-back-from-the-dead/>
- [3] Flexible Web Design: Creating Liquid and Elastic Layouts with CSS – Third Edition, Zoe Mickley Gillenwater, New Riders Publishing 2013, ISBN-13: 978-0321553843
- [4] G. Marinova, V. Guliashki, O. Chikov, “Concept of Online Assisted Platform for Technologies and Management in Communications – OPTIMEK”, Int.l Conference in Computer Science, Information System and Telecommunication, ICCSIST 2014, 7-9 November 2014, Durrës, Albania, 2014, pp.55-62
- [5] O. Chikov, G. Marinova, MVC Framework Approach in the Online assisted platform form Computer-aided Design in Communications – Online-CADCOM, Computer and Communications Engineering, Submitted.
- [6] G. Marinova, O. Chikov, Methodology for tools integration in the Online Assisted Platform for Computer-aided design in communications, Proc. of papers of L International scientific conference on information, communication and energy system and technologies, ICEST’2015, 24-26 June 2015, Sofia, pp.31-36, ISBN:978-619-167-182-3
- [7] CMS Design Using PHP and jQuery, Kae Verens, Packt Publishing, 2010, ISBN: 978-1-849512-52-7
- [8] Develop Responsive HTML  
<https://responsivedesign.is/develop/html>
- [9] Responsive Web Design with HTML5 and CSS3, Ben Frain, Packt Publishing 2012, ISBN-13: 978-1849693189

# Signal Processing





# Statistical Analysis of Round-off Noise in First-Order IIR DSP Systems

Zlatka Valkova-Jarvis<sup>1</sup>, Kamelia Nikolova<sup>1</sup> and Venera Dimitrova<sup>1</sup>

**Abstract** – The quantisation of multiplication products in digital signal processing (DSP) recursive systems causes parasitic noise, resulting in deterioration of the system characteristics. The low-sensitivity of DSP systems helps to decrease this noise, thereby improving system performance. This work presents the experimental results of an investigation into the quantisation of the multiplication product at all possible placements of poles for some of the most commonly-used low-sensitivity recursive first-order digital systems. The study presented here enables the areas of best performance of each of the examined DSP systems to be determined.

**Keywords** – quantisation effects, round-off errors, first-order LP and HP DSP networks, all-pass first-order digital systems.

## I. INTRODUCTION

Due to recursion, infinite impulse response (IIR) DSP systems have considerable advantages over finite impulse response (FIR) ones. Recursive systems are much preferred in practice owing to their very high efficiency, lower order, faster signal processing, and improved operation at high sampling rates. Despite the above advantages, IIR digital systems have more complex structures, pose greater challenges in design and investigation, may have linear phase responses in only part of the relevant frequency range, and can display instability.

Most DSP applications in telecommunications, especially real-time processing ones, use fixed-point binary arithmetic and sign-magnitude, one's complement or two's complement binary encoding. Despite the narrow dynamic range, fixed point representation has advantages such as lower cost of hardware, higher speed processing and reduced energy consumption [1].

Ideally, all the discrete signals and parameters in both FIR and IIR DSP systems are of infinite length; hence the digital system will be absolutely quiet with no noise or distortion. In practice, delivering a working recursive DSP system based purely on theoretical design requires the use of a limited number of bits to represent the system's coefficients and processed signals. This process is called *quantisation* and transforms the ideal linear DSP system into a real nonlinear DSP system. Hence, finite word length errors will appear, possibly resulting in a catastrophic impact on the performance of IIR digital systems and thus rendering them inoperative. For example, unexpected oscillations may appear in the output of a nonlinear digital system – these are called limit cycles. They are typically a problem for recursive systems with poles located very close to the limit of stability (the unit circle) [1].

Quantisation can be performed either by *rounding* or by *truncation*, both having significant impact on the accuracy of

the presentation of the numbers as well as the characteristics of digital circuits.

Quantisation effects are also called *parasitic* and, for fixed-point digital systems, can be summarised as follows: overflow saturation; arithmetic rounding; coefficient quantisation; data scaling; limit cycling. Numerous methods are used for reducing parasitic effects and they are primarily directed towards one particular problem. However, general corrective techniques are also available and take the form of: scaling the input or coefficients, increasing the system's word length, and selecting an alternative digital structure [1], [2].

In DSP the following fixed-point parameters and signals are quantised by rounding or by truncation:

- The input and output signals;
- Signals resulting from intermediate calculations (also called inner products);
- Multiplier coefficients of the digital system.

The DSP system coefficients are quantised only once, remaining unchanged throughout the entire system operation but also causing the digital system's characteristics to diverge from their ideal form. If the system specifications are no longer met, the quantisation design must be optimised by allocating more bits or choosing a less sensitive realisation.

Quantisation of the results of intermediate calculations is a repetitive process that is reiterated on each cycle of the digital system action. As a result, the output of the system accumulates quantisation errors that worsen the DSP system performance. It is of great importance to assess these errors in order to develop methods and approaches to reduce them [2]. One of the most effective approaches to counteract quantisation errors is the use of very low-sensitivity digital structures, such as cascade or parallel combinations of first- and second-order sections.

In addition to being the building blocks of cascade or parallel DSP structures, first-order sections may also be autonomously operating DSP units. Examples of this are, inter alia, digital integrators, differentiators, averagers, oscillators, DC blockers, maximum-flat group delay adjustable and fixed fractional-delay digital filters [1], [3] digital Hilbert transformers; each performing as frequency selective low-/high-pass or all-pass systems [4].

In this work, classic frequency selective low-pass-cum-high-pass and all-pass first-order low-sensitivity digital systems are studied. The comparative analysis conducted here helps to define the best performance areas of the investigated bilinear DSP systems with respect to the errors due to quantisation of multiplication products. Our analysis allows recommendations to be made for the optimum use of the first-order digital sections which were investigated, these being currently regarded as the best in terms of sensitivity.

<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: zvv@tu-sofia.bg, ksi@tu-sofia.bg, vdimitrova@tu-sofia.bg.

## II. MULTIPLICATION ROUND-OFF NOISE ANALYSIS – A THEORETICAL BACKGROUND

The nonlinear nature of the process of quantisation makes it difficult to assess the effects of the finite word length on the performance of DSP systems. Hence, the analysis of parasitic effects can only be estimated by making a number of assumptions regarding the quantisation being linear in nature. Statistical assessment of the quantisation errors is a method that, following these assumptions, allows quantisation of the values represented by fixed point to be regarded as a linear process with a given accuracy.

However in such an approximate approach the results should also be treated as approximate.

The statistical assessments of quantisation errors that are most often used are:

- Probability Density Function  $P(e)$  (PDF);
- Variance  $\sigma_e^2$  or its square root standard deviation  $\sigma_e$  of the quantisation error;
- Signal-to-Noise Ratio (SNR).

In order to regard the quantisation of a signal  $m(n)$ , resulting in the occurrence of the quantisation error  $e(n)$ , as a linear process, the following assumptions are necessary:

1. The error signal  $e(n)$  is uncorrelated with the quantised signal  $m(n)$ .
2. The random discrete sequence of the quantisation error  $e(n)$  is an independent process and its samples are uncorrelated with one another.
3. The Probability Density Function  $P(e)$  of the quantisation error of each sample of the discrete error sequence  $e(n)$  has a normal distribution and constant value in a range having a width equal to the quantisation step size  $\delta$ .

These stipulations make practical sense if the values of the discrete quantised signal  $m(n)$  vary considerably from sample to sample, which can occur when the quantisation step is small.

The inner DSP products result from two arithmetic operations - summation and multiplication – both produce fixed-point results with longer word length; these then have to be quantised, especially in the case of real-time signal processing.

When summing fixed-point binary numbers, any increase will occur at the most significant bit thus possibly resulting in an integer part; the quantisation of this integer will lead to overflow – catastrophic in its effect on recursive digital systems.

In this work, we will investigate the quantisation of an inner signal resulting from intermediate multiplication. The word length of the multiplication product is the sum of those of the multiplication factors. In line with the stipulations listed above, a linear statistical noise model of multiplication product quantisation (Fig. 1) can be developed. Assuming that quantisation by rounding is used, which is most often the case in practice, the model is called the *round-off noise model*.

The signal  $m(n)$  and the constant  $a$  are quantised to  $B$  bits before being multiplied. The discrete sequence of the multiplication  $u(n)=am(n)$  is quantised by rounding in order to reduce its word length to  $B$  bits. The quantisation device is replaced by the simple linear model shown in Fig. 1, under which the quantised multiplication  $\hat{u}(n)$  will be the result of the

sum of non-quantised multiplication  $u(n)$  and the quantisation error  $e(n)$  (more commonly called *quantisation noise*, because it really does have the characteristics of noise). The generated parasitic noise signal  $e(n)$  is injected at a node after each multiplication quantisation in the digital structure and is referred to as *round-off noise* [2].

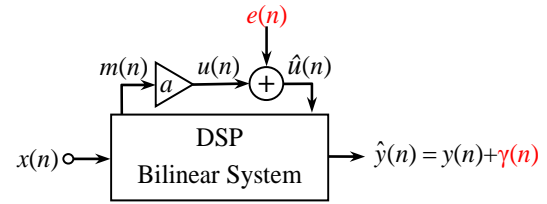


Fig. 1: Linear statistical noise model of multiplication product quantisation by rounding

The statistical quantisation noise model is regarded as a linear process when the samples of the signal  $m(n)$  have fast changing values. In this case, the quantisation noise signal samples are uncorrelated with each other, i.e.  $e_i(n)$  and  $e_i(k)$  are statistically independent for  $n \neq k$ . Hence, the round-off noise  $e(n)$  can be modelled as stationary white noise, uniformly distributed in the relevant interval. Since one multiplication is performed in the first-order DSP systems considered here, only one round-off noise source will appear. This error source develops a white noise  $\gamma(n)$  at the output of the bilinear digital system, which is added to the output signal  $y(n)$ , resulting in a output signal mixed with noise  $\hat{y}(n)$  (Fig.1).

When a fixed-point multiplication product is encoded by any of the three binary codes (sign-magnitude, one's complement or two's complement), round-off error will be normally distributed in the interval  $[-\delta/2 \div \delta/2]$ . When the quantisation error is normalised to the quantisation step size:  $e_n = e/\delta$ , the limits of the intervals of uniform distribution change. For the normalised round-off error the interval will be  $[-1/2 \div 1/2]$ .

The method that is used most frequently for a statistical representation of the quantisation error is determining variance  $\sigma_e^2$  or its square root standard deviation  $\sigma_e$ . In the case of fixed-point representation, the uniformly distributed round-off noise  $e(n)$  has zero mean value and the variance  $\sigma_e^2$  [1]:

$$\sigma_e^2 = \frac{\delta^2}{12} = \frac{2^{-2B}}{12}. \quad (1)$$

$B$  is the word length in bits (without the sign bit). The quantisation step size  $\delta$  (step of sampling) is the maximum value of the modulus of the error for fixed-point fractional numbers.

The variance of the noise that reaches the bilinear digital system output as a result of the multiplication quantisation has a steady-state (nominal) value  $\sigma_{\gamma}^2$  given by:

$$\sigma_{\gamma}^2 = \sigma_e^2 \frac{1}{2\pi j} \oint G(z)G(z^{-1})z^{-1}dz = \sigma_e^2 \sigma_{\gamma,n}^2, \quad (2)$$

where  $G(z)$  is defined as noise transfer function (NTF), i.e. the transfer function from the round-off noise source  $E(z)$  to the digital system output  $Y(z)$ . Clearly, the NTF depends on the structure of the digital filter.  $\sigma_{\gamma,n}^2$  denotes the noise gain and is also called normalised output noise variance [2], [5].

### III. NOISE MODELS OF FIRST-ORDER LOW-SENSITIVITY DSP SYSTEMS

As discussed, the use of very low-sensitivity digital structures is one of the most effective approaches to ensure successful DSP system performance in real-world conditions and to counteract the effects of finite word length. Different digital structures do not show the same sensitivity to coefficient accuracy and thus quantisation of coefficients results in unacceptable distortions in frequency response. In this work we investigate low-sensitivity bilinear digital structures with respect to the multiplication round-off noise for all possible pole locations.

In Fig. 2, graphs of signal flow due to multiplication product quantisation in universal LP/HP first-order MHNS (Fig. 2a) and LS1b (Fig. 2b) digital systems are shown [3].

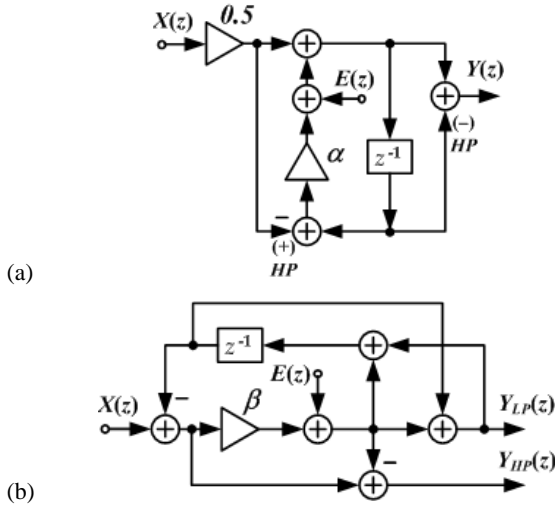


Fig.2: Statistical round-off multiplication noise model of bilinear LP/HP DSP systems (a) MHNS and (b) LS1b

Since round-off noise models contain one noise source  $E(z)$ , the noise transfer functions of bilinear LP/HP MHNS and LS1b systems are as follows:

$$G_{MHNS}^{LP}(z) = \frac{Y(z)}{E(z)} = \frac{1+z^{-1}}{1-\alpha z^{-1}}; \quad (3)$$

$$G_{LS1b}^{LP}(z) = \frac{Y_{LP}(z)}{E(z)} = \frac{1+z^{-1}}{1-(1-2\beta)z^{-1}}; \quad (4)$$

$$G_{MHNS}^{HP}(z) = \frac{Y(z)}{E(z)} = \frac{1-z^{-1}}{1-\alpha z^{-1}}; \quad (5)$$

$$G_{LS1b}^{HP}(z) = \frac{Y_{HP}(z)}{E(z)} = \frac{1-z^{-1}}{1-(1-2\beta)z^{-1}}. \quad (6)$$

Statistical round-off noise models for the three most low-sensitivity bilinear digital systems are depicted in Fig. 3 [4].

The all-pass bilinear sections (Fig. 2) noise transfer functions are as follows:

$$G_{MH1}(z) = \frac{Y(z)}{E(z)} = \frac{1+z^{-1}}{1-bz^{-1}}; \quad (7)$$

$$G_{SV}(z) = \frac{Y(z)}{E(z)} = \frac{-1+z^{-1}}{1+(1-c)z^{-1}}; \quad (8)$$

$$G_{ST1}(z) = \frac{Y(z)}{E(z)} = \frac{1+z^{-1}}{1-(1-a)z^{-1}}. \quad (9)$$

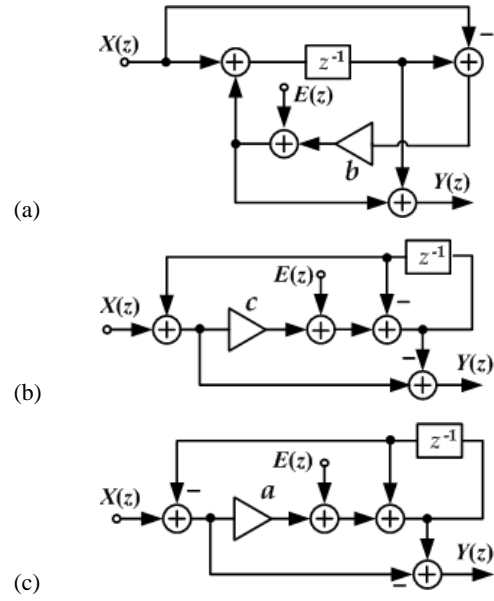


Fig.3: Statistical round-off multiplication noise model of bilinear All-pass DSP systems (a) MH1, (b) SV and (c) ST1

### IV. EXPERIMENTAL RESULTS

#### A. Low-pass / High-pass bilinear sections

The universal LP/HP MHNS and LS1b are limit-cycle-free DSP systems, stable when the coefficients are  $\alpha \in (-1 \div 1)$  and  $\beta \in (1 \div 0)$ . As per Eq. (2), the nominal round-off noise variances for LP and HP MHNS and LS1b systems (Eqs. (3), (4), (5) and (6)) are respectively as follows:

$$\sigma_{MHNS}^{LP} = \sigma_e^2 \frac{2}{1-\alpha}; \quad \sigma_{LS1b}^{LP} = \sigma_e^2 \frac{1}{\beta}; \quad (10)$$

$$\sigma_{MHNS}^{HP} = \sigma_e^2 \frac{2}{1+\alpha}; \quad \sigma_{LS1b}^{HP} = \sigma_e^2 \frac{1}{1-\beta}. \quad (11)$$

They are experimentally explored when the bilinear system fixed-point coefficient and the multiplication inner product are quantised by rounding to very short word length - 4, 5 and 6 bits, including the sign bit. Experimental results for round-off variance versus pole location are shown in Fig. 4a - for LP outputs (Eq. (10)), and in Fig. 4b - for HP outputs (Eq. (11)). The experiments are conducted on the pass-band frequency range, which is the essential one. One expected result is that the shorter the word length, the higher the value of the round-off variance. Due to its lower pass-band sensitivity [3], the LS1b bilinear system demonstrates lower multiplication round-off variance in comparison to the MHNS system for both LP and HP outputs. The shorter the word length, the bigger is the difference between these two sections. When the round-off quantisation removes more bits, the level of the round-off variance remains unchanged for larger pole location intervals. The variance levels change less frequently in the case of the low-sensitivity LS1b system.



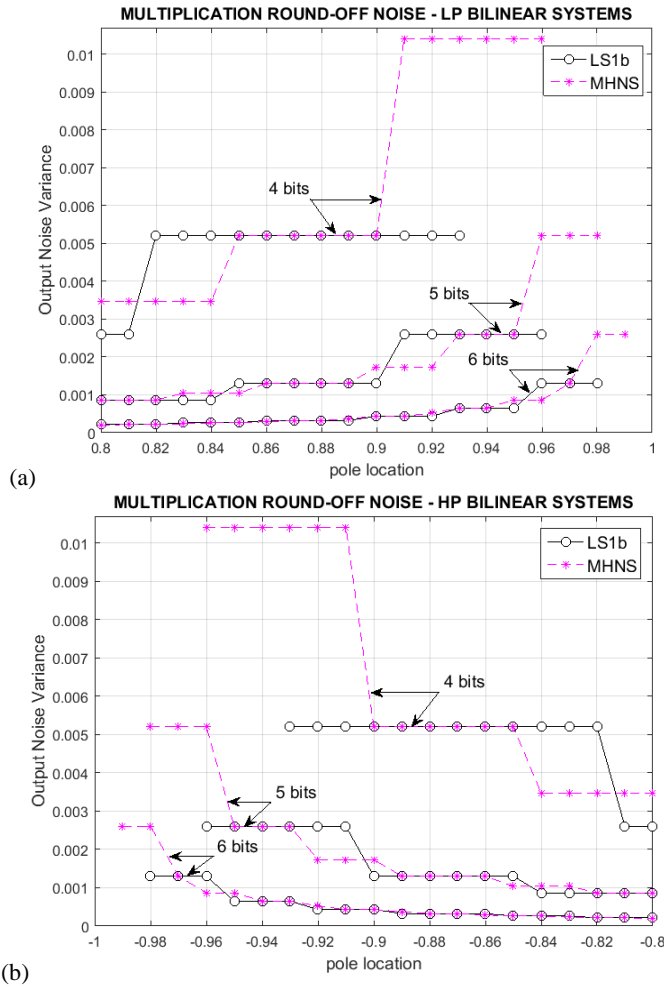


Fig. 4: Output round-off noise variance for bilinear (a) LP and (b) HP MHNS and LS1b DSP systems for different word lengths.

### B. All-pass bilinear sections

Analytical expressions of output round-off noise variance for the bilinear all-pass systems in Fig. 3 (Eqs. (7), (8) and (9)) are likewise calculated:

$$\sigma_{MH1}^2 = \sigma_e^2 \frac{2}{1-b}; \quad \sigma_{SV}^2 = \sigma_e^2 \frac{2}{c}; \quad \sigma_{ST1}^2 = \sigma_e^2 \frac{2}{a}. \quad (12)$$

Despite the word length, each all-pass system shows the best result for the pole location interval where it is most used in practice (Fig. 5). The MH1 bilinear system has low round-off variance for the pole located in  $(-0.5 \div 0.5)$ . Actually, in the last third of this interval, the SV system achieves a lower variance value, but this bilinear section is primarily useful when the pole is in  $(0.5 \div 1)$  where the SV section's variance is anyway the lowest. The ST1 all-pass system performs best when the pole is in the interval  $(-1 \div -0.5)$ , which corresponds to its existing implementation.

## V. CONCLUSION

Regardless of their undoubted advantages, IIR digital systems suffer from parasitic effects due to quantisation of multiplier coefficients, signals and inner products of digital processing. Since the low-sensitivity of DSP systems provides

better resistance to parasitic effects, this work investigates the round-off noise of the lowest sensitivity bilinear digital LP, HP and all-pass systems. Relevant assumptions are made about the source of quantisation noise that make the described linear noise model possible. The experiments performed confirm the lower the sensitivity the better the behaviour of the system in the frequency range which is important for the system.

The experimental work conducted here can be effectively employed for DSP systems of other types and orders, which will be helpful when choosing the best system for a particular application.

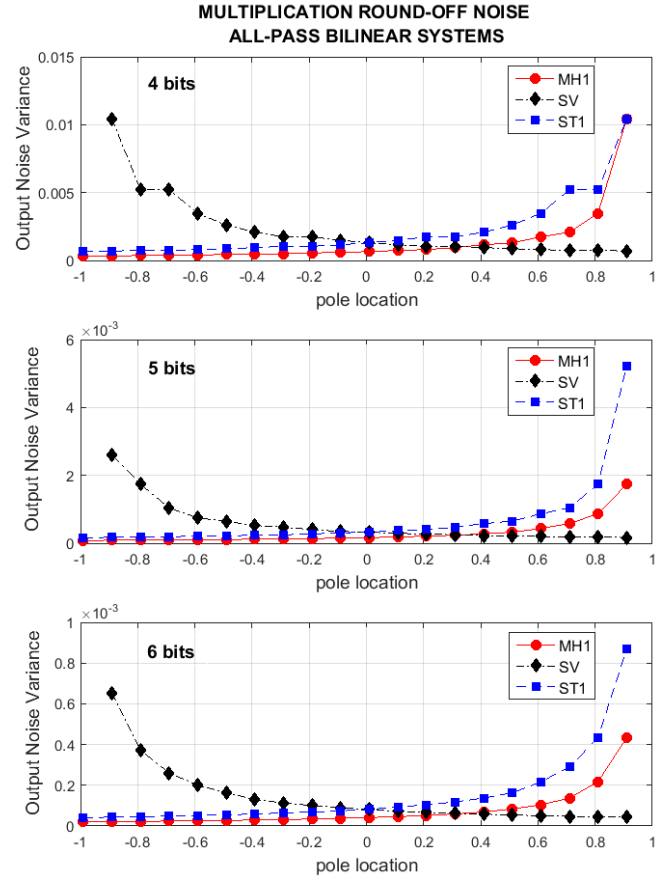


Fig. 5: Output round-off noise variance for bilinear all-pass MH1, SV and ST1 DSP systems for different word length

## REFERENCES

- [1] V. Udayashankara, *Modern Digital Signal Processing*, PHI Learning Pvt. Ltd., 2012.
- [2] S. K. Mitra, K. Hirano and H. Sakaguchi, "A simple method of computing the input quantization and multiplication roundoff errors in a digital filter", *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. ASSP-22, No. 5, pp. 326-329, Oct. 1974.
- [3] G. Stoyanov, M. Kawamata Z. Valkova, "New first- and second-order very low-sensitivity bandpass/bandstop complex digital filter sections" *Proc. IEEE Region 10th Annual Conf. "TENCON'97"*, Brisbane, Australia, vol. 1, pp. 61-64, 1997.
- [4] G. Stoyanov, and K. Nikolova, "Improved accuracy and low-sensitivity design of digital allpass based Hilbert Transformers," *Proc. Int. Conf. TELSIKS'2013*, Nish, Serbia, 2013, pp. 51-60.
- [5] Nikolova Z., D. Romanska, "Multiplication Products Quantization Noise Analysis for Orthogonal Complex IIR Digital Filters", *ICEST'08*, Nish, Serbia, pp. 540 – 543, 25-27 June 2008.

# Novel CIC FIR Filter Functions: Design and Influence of Added Filter Blocks

Biljana P. Stošić, Vlastimir D. Pavlović

**Abstract** — This paper presents new designs of selective multiplierless Cascaded-Integrator-Comb (CIC) finite impulse response (FIR) filter functions with applications in modern communication systems. The designs are based on cascading simple filters. The resulting filters possess an improved frequency response behaviour. The influence of cosine prefilter and compensation filter to frequency response characteristics of proposed novel CIC FIR filter functions is investigated here.

**Keywords** — CIC filters, FIR filters, linear phase, multiplierless structure, selective filters.

## I. INTRODUCTION

The term “Cascaded-Integrator-Comb (CIC)” filters was first reported in 1981 by E.B. Hogenauer [1]. He proposed a class of hardware-efficient linear phase finite impulse response (FIR) filters known as CIC or Hogenauer filters.

Because of the disadvantages of a CIC FIR filter such as not flat passband and a high passband droop, it is of a great interest to improve magnitude response characteristic. The idea of modified CIC filters is not new, it is proposed by the authors in [2]-[8]. This paper is focused on the design of novel class of CIC FIR filter functions in the explicit form. The properties of the designed novel class are demonstrated in detailed examples. The ideas of spreading the delays in the comb stages provide excellent results for increasing insertion loss in stopband and small changes in the bandwidth size of the filter function compared to classical CIC filter that has the same number of cascades and the same level of constant group delay. In general, the passband characteristic can be improved by adding a compensator filter in cascade with the original filter [9]-[12]. A cosine prefilter [13]-[14] can be added to improve the stopband characteristic. In this paper, influence of cosine prefilter and compensator blocks to frequency response characteristics of proposed novel CIC FIR filter functions is investigated.

The rest of the paper is organized as follows. Section II describes multiplierless CIC filter functions. The cosine prefilter and compensator blocks are elaborated in Section III and IV, respectively. In Section V, proposed novel CIC FIR filter functions are presented and analyzed. Section VI compares different modified CIC FIR filter functions present in the literature. Influence of added filter blocks (cosine prefilter and compensator) to modified CIC filters is described in Section VII. Concluding remarks of this paper are given in Section VIII.

Biljana P. Stošić and Vlastimir D. Pavlović are with the University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, Niš, Serbia, E-mail: biljana.stosic@elfak.ni.ac.rs

## II. MULTIPLIERLESS LINEAR-PHASE CIC FILTER FUNCTIONS

The design methodology of conventional CIC FIR filters is well known to the scientific community. The normalized CIC FIR filter function of one section in  $z$ -domain is defined with

$$H(N, z) = \frac{1}{N} \cdot H_c(N, z) \cdot H_i(z). \quad (1)$$

There are two basic building blocks: a comb and an integrator. The filter function of a comb in  $z$ -domain is  $H_c(N, z) = (1 - z^{-N})$ , therefore a difference equation in time domain is  $y_c(n) = x(n) - x(n - N)$ .  $H_i(z) = 1/(1 - z^{-1})$  represents the filter function of an integrator in  $z$ -domain and  $y_i(n) = x(n) + y(n - 1)$  is its time domain representation.

A poor magnitude characteristic of the CIC filter composed of one section, Eq. (1), is improved by cascading several identical CIC filters. The classical CIC FIR filter function of normalized amplitude response characteristic, represented in the  $z$ -domain, is defined as

$$H(N, K, z) = \left( \frac{1 - z^{-N}}{N \cdot (1 - z^{-1})} \right)^K, \quad (2)$$

where  $N$  is the decimation factor, and  $K$  is the number of sections (identical cascaded CIC filters of one section) [1]. They are positive integers.

## III. COSINE PREFILTER

To improve the magnitude characteristic of the CIC filter, the following cosine prefilter introduced in [13]-[14], can be used

$$H_{\cos}(z^N) = 0,125 \cdot (1 + z^{-2N}) \cdot (1 + z^{-N})^2. \quad (3)$$

It is combination of two cosine functions. Because this prefilter only employs unity coefficients, this prefilter is essentially multiplication free.

## IV. COMPENSATOR

Passband droop compensation is done by use of a multiplierless  $2N$ -order FIR filter with one free integer parameter  $b$  presented in [12]

$$G(z^N) = B \cdot [1 + A \cdot z^{-N} + z^{-2N}], \quad (4)$$

where  $B = -2^{-(b+2)}$  is a scaling factor ensuring unitary gain at the digital frequency zero, and  $A = -(2^{b+2} + 2)$ . The transfer function of compensator filter is a function of  $z^N$  and has only one coefficient  $A$  which can be realized using additions and shifts.

## V. PROPOSED NOVEL CIC FIR FILTER FUNCTIONS

A modification of classic CIC filters is here proposed. The filter function of normalized amplitude response characteristic of designed **novel class of CIC FIR filter functions** can be written in **non-recursive form** as follows

$$H(N, K, L, z) = \left[ \frac{1}{N-3} \cdot \frac{1}{N-2} \cdot \frac{1}{N-1} \cdot \frac{1}{N} \right]^L \cdot \left[ \left( \sum_{r=0}^{N-4} z^{-r} \right) \cdot \left( \sum_{r=0}^{N-3} z^{-r} \right) \cdot \left( \sum_{r=0}^{N-2} z^{-r} \right) \cdot \left( \sum_{r=0}^{N-1} z^{-r} \right) \right]^L \cdot \left[ \frac{1}{N+1} \cdot \frac{1}{N+2} \cdot \frac{1}{N+3} \right]^L \cdot \left[ \left( \sum_{r=0}^N z^{-r} \right) \cdot \left( \sum_{r=0}^{N+1} z^{-r} \right) \cdot \left( \sum_{r=0}^{N+2} z^{-r} \right) \right]^L$$

and  $K = 7L$ , (5)

where  $N$  and  $L$  are free integer parameters. The main idea is instead of using  $K$  times the same CIC core filter with a fixed order  $N$ , to use  $L$  times (where  $L = K/7$ ) a core combined from orders  $N-3$ ,  $N-2$ ,  $N-1$ ,  $N$ ,  $N+1$ ,  $N+2$ , and  $N+3$ .

Classical CIC filters have multiple zeros located on the unit circle. With this new idea, zeros are more spread around the stopband region of the unit circle leading to a better stopband attenuation while having the same number of zeros and a very similar circuit complexity. The idea provides an elegant solution to increase the stopband attenuation while having the same complexity.

**A recursive form of a novel class of CIC FIR filter functions is**

$$H(N, K, L, z) = \left( \frac{1-z^{-(N-3)}}{(N-3) \cdot (1-z^{-1})} \cdot \frac{1-z^{-(N-2)}}{(N-2) \cdot (1-z^{-1})} \cdot \frac{1-z^{-(N-1)}}{(N-1) \cdot (1-z^{-1})} \cdot \frac{1-z^{-N}}{N \cdot (1-z^{-1})} \right)^L \cdot \left( \frac{1-z^{-(N+1)}}{(N+1) \cdot (1-z^{-1})} \cdot \frac{1-z^{-(N+2)}}{(N+2) \cdot (1-z^{-1})} \cdot \frac{1-z^{-(N+3)}}{(N+3) \cdot (1-z^{-1})} \right)^L$$

and  $K = 7L$ . (6)

Frequency response characteristic is obtained by evaluating the filter function in the  $z$ -plane at the sample points defined by setting  $z = e^{j\omega}$ , where  $\omega = 2\pi \cdot f$  [rad/s]. The normalized amplitude response characteristic is defined as the magnitude of the complex filter frequency response  $H(N, K, L, z = e^{j\omega})$ . The magnitude response characteristic is obtained as absolute value of normalized amplitude response characteristic. The linear phase response characteristic is defined as the phase angle of the complex filter frequency response. The constant group delay response characteristic is expressed as

$$\tau(N, K, L, \omega) = (N-1) \cdot 7L / 2. \quad (7)$$

### A. Impulse Coefficients

Generally, the FIR filter function of the corresponding non-recursive implementation has the form

$$H(N, K, L, z) = \frac{1}{H_0(N, K, L)} \cdot \sum_{r=0}^{(N-1) \cdot K} h(N, K, L, r) \cdot z^{-r} \quad (8)$$

where  $H_0(N, K, L) = H(N, K, L, z \rightarrow 1)$  is normalized constant for unit magnitude response at  $f = 0$ . Vector  $h(N, K, L)$  is the vector of impulse response coefficients  $h(N, K, L, r)$ , where total number of elements in vector is  $M = (N-1) \cdot K + 1$  and  $h(N, K, L, r) = h(N, K, L, M-r)$ .

The non-recursive implementation of classical CIC filter functions  $H(N, K, z)$ , obtained for even value of integer parameter  $N = 4$  and  $K = 7$ , is observed here. The normalized constant is  $H_{0CIC}(4, 7) = 16384$ . The vector of the impulse response coefficients is

$$h_{CIC}(4, 7) = \left\{ 1, 7, 28, 84, 203, 413, 728, 1128, 1554, 1918, 2128, \right. \\ \left. 2128, 1918, 1554, 1128, 728, 413, 203, 84, 28, 7, 1 \right\} \quad (9)$$

The non-recursive forms of novel CIC filter functions,  $H(N, K, L, z)$ , obtained for even value of integer parameter  $N$  and  $K = 7$  (obtained for  $L = 1$ ), satisfied Eq. (8). The vector  $h(N, K, L)$  of impulse response coefficients is given in Eq. (10). The constant is  $H_0(4, 7, 1) = 5040$ .

$$h(4, 7, 1) = \left\{ 1, 6, 20, 49, 98, 169, 259, 359, 455, 531, 573, \right. \\ \left. 573, 531, 455, 359, 259, 169, 98, 49, 20, 6, 1 \right\} \quad (10)$$

The normalized constant for proposed filter class is smaller than that of classical CIC FIR filters. Also, impulse response coefficients of proposed filter class have smaller values in comparison to those of classical CIC filters.

### B. New Filter Functions and their Properties

In Table I, passband and stopband cut-off frequencies,  $f_{cp}$  and  $f_{cs}$ , maximum attenuation in the passband, chosen as  $\alpha_{\max} = 0.28$  dB, minimum attenuation in the stopband,  $\alpha_{\min}$  [dB], and group delay,  $\tau(f)$  [s], of the classical CIC filter functions  $H(N, K, z)$  from Eq. (2), are given. These parameters, as well as stopband cut-off frequency  $\hat{f}_{cs}$  and attenuation in the stopband area  $\hat{\alpha}_{\min}$  [dB] of the novel CIC filter functions  $H(N, K, L, z)$  from Eq. (5), are listed in Table II. The novel CIC filter functions have two peaks in the transition area of the classical filter (on frequency between the passband and stopband cut-off frequencies) as can be seen in Fig. 2. The attenuation of the lowest peak in the stopband region is assigned as  $\alpha_{\min}$  [dB] and given in Table II. Note that the attenuations of the novel functions in the stopband area  $\hat{\alpha}_{\min}$  [dB] given in Table II are higher than the attenuations of classical CIC filter in the stopband area  $\alpha_{\min}$  [dB] given in Table I. Filter functions are compared under the fair conditions: the same number of constant group delay and the same number of cascaded sections  $K$ .

The relative difference of passband cut-off frequencies in [%] is defined as ratio

$$\Delta f_{cp} = [(f_{cp\_CIC} - f_{cp\_novel\_class}) / f_{cp\_CIC}] \cdot 100 \quad (11)$$

and obtained to judge the improvement of the frequency response characteristics. The relative differences of passband

cut-off frequencies are  $\Delta f_{cp} = 5.25\%$  for  $L=1$ , i.e.  $K=7$ ,  $\Delta f_{cp} = 5.51\%$  for  $L=2$ , i.e.  $K=14$ , and  $\Delta f_{cp} = 5.52\%$  for  $L=3$ , i.e.  $K=21$ .

TABLE I

CUT-OFF FREQUENCIES IN PASSBAND AND STOPBAND, CONSTANT GROUP DELAY AND STOPBAND ATTENUATION OF CLASSICAL CIC FILTER FOR  $K \in \{7, 14, 21\}$  and  $N=6$

$K$	$f_{cp}$	$\alpha_{\max}$	$f_{cs}$	$\alpha_{\min}$	$\tau(f)$
7	0.00894	0.28	0.13354	86.979	17.5
14	0.00632	0.28	0.13354	173.958	35.0
21	0.00516	0.28	0.13354	260.937	52.5

TABLE II

CUT-OFF FREQUENCIES IN PASSBAND AND STOPBAND, CONSTANT GROUP DELAY AND STOPBAND ATTENUATION OF NOVEL CIC FILTERS FOR  $K \in \{7, 14, 21\}$  (OBTAINED FOR  $L \in \{1, 2, 3\}$ ),  $N=6$  AND  $\alpha_{\max} = 0.28\text{dB}$

$K$	$f_{cp}$	$f_{cs}$	$\alpha_{\min}$	$\hat{f}_{cs}$	$\hat{\alpha}_{\min}$	$\tau(f)$
7	0.00847	0.10956	80.855	0.13727	91.758	17.5
14	0.00599	0.10956	161.709	0.13727	183.516	35.0
21	0.00489	0.10956	242.564	0.13277	275.274	52.5

In order to illustrate clearly the achieved improvements of the new class, the normalized magnitude response characteristics in dB of the new functions and classical CIC filters are summarized in Fig. 1. It can be concluded that the attenuation in the stopband region is closely related to the parameter  $L$ . By increasing  $L$  for the constant value of  $N$ , the higher stopband attenuation is achieved.

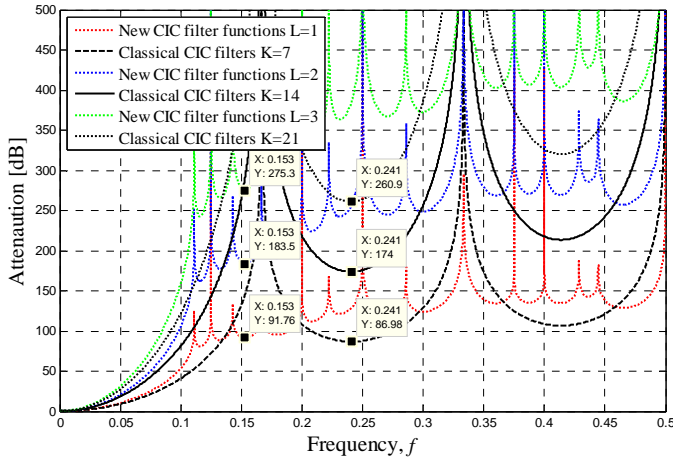


Fig. 1. Magnitude response characteristics in dB versus normalized frequency  $f = \omega / (2\pi)$  of novel CIC FIR filter functions and those functions of classical CIC filters for  $N=6$ , and different values of parameter  $L \in \{1, 2, 3\}$ .

## VI. COMPARISON AMONG FILTER FUNCTIONS

This kind of modified CIC filters has been widely studied in the literature in [2]-[8]. A detailed comparison of novel filter functions with the selected modified CIC filters from literature is carried out. The CIC FIR filter functions which closed-form design equations are presented in [2], [5], [8] and

[12] are designed to compare feature of that class with the new CIC FIR filter functions presented in this paper. Each class is designed for specified group delay  $\tau = 28s$ . Comparison is summarized in Fig. 2. Technique proposed in [12] includes passband droop compensator and stopband improvement filters. The new filter functions show higher selectivity in the transition area. The proposed filters have bigger attenuation in the stopband area without added additional filter for improvements versus solution given in [12] where stopband improvement filter is included.

In order to implement the design forms the following steps are to be followed: 1) choose constant group delay value, 2) identify free filter parameters for the chosen group delay, and 3) based on the form (3) or (4), design new filter class with the chosen parameters.

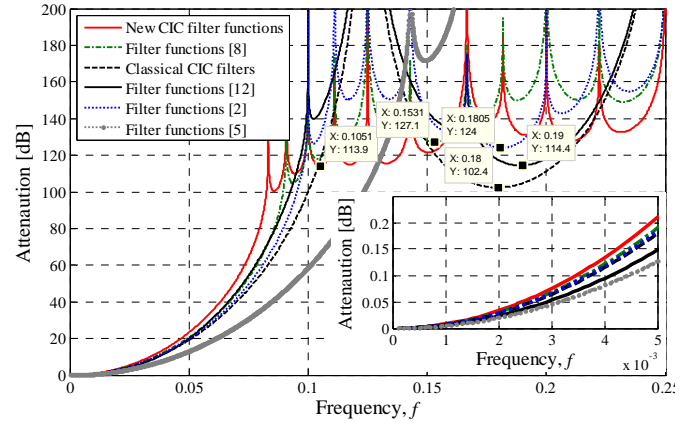


Fig. 2. Normalized magnitude response characteristics in dB (dashed black lines - classical CIC filter for  $K=8$ ,  $N=8$ ; solid red lines - novel filter functions for  $N=9$ ,  $K=7$ ,  $L=1$ ; green line - filter functions given in [8] for  $N=8$ ,  $L=1$ ,  $K=8$ ; black lines - filter functions given in [12] for  $M=8$ ,  $K=8$ ,  $N_1=3$ ,  $N_2=5$  and  $b=0$ ); dashed blue line - filter functions given in [2] for  $N=8$ ,  $L=1$ ,  $K=8$ ; dashed gray line - filter functions given in [5] for  $N=5$ ,  $L=2$ ,  $K=14$ ).

## VII. INFLUENCE OF ADDED FILTER BLOCKS

Influence of cosine prefilter and compensator blocks to frequency response characteristics of novel CIC FIR filter functions is investigated here.

### A. Compensator Influence

Like classical CIC filters, new designed CIC FIR filter functions exhibit a significant passband droop in the passband of interest, as can be seen from the passband detail given in Fig. 2, what is usually undesirable in many applications. Hence, it is of great interest to get a flat passband in some way, i.e. by connecting CIC compensator [12] in cascade with the original filter.



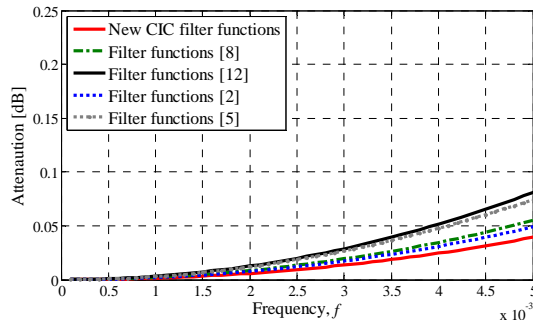


Fig. 3. Normalized magnitude response characteristics in passband in case with added compensator  $b = 0$ .

### B. Compensator and Prefilter Influence

The proposed filter structure is the cascade of the cosine prefilter structure (3), the new filter functions (6), and the compensator filter (4):

$$H_p(z) = H_{\cos}(z^N) \cdot H(N, K, L, z) \cdot G(z^N) \quad (12)$$

Fig. 4 demonstrates the improvement in the stopband and the increased passband droop of the resulting filter. The proposed filter has a very little increased attenuation in the folding bands. However, the passband exhibits a high passband droop. The passband characteristic of the proposed filter structure is inferior to that of novel CIC filter functions.

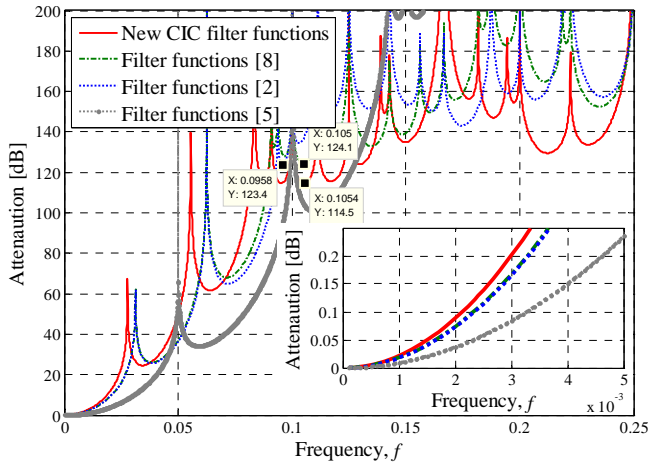


Fig. 4. Filter functions with added cosine prefilter and compensator and their normalized magnitude response characteristics in dB (solid red lines - novel filter functions for  $N = 9$ ,  $K = 7$ ,  $L = 1$ ; green line - filter functions given in [8] for  $N = 8$ ,  $L = 1$ ,  $K = 8$ ; dashed blue line - filter functions given in [2] for  $N = 8$ ,  $L = 1$ ,  $K = 8$ ; dashed gray line - filter functions given in [5] for  $N = 5$ ,  $L = 2$ ,  $K = 14$ ), and  $b = 0$ .

### C. General Conclusion

In this kind of modified CIC filter functions, the cosine prefilter is not required in order to improve the magnitude characteristic. Novel CIC filter functions exhibit high stopband attenuation of themselves. Compensator block can be used if better passband characteristic is required.

## VIII. CONCLUDING REMARKS

This paper briefly introduced the history and classical CIC filters. Also, the paper investigates and proposes a new CIC (cascade of integrator and comb filters) digital filter with

presumably advantages such as higher stopband attenuation, smaller impulse-response coefficients and better passband characteristics.

On the basis of the main idea of CIC filter, in this paper, a modified CIC linear phase FIR filters are designed. For the novel filters, filter coefficients are computed by some closed-form equations, and the frequency response characteristics are also studied. Some performance comparison results are presented. Design examples for proposed filter are given and performance comparisons are given as well. These new designs can be useful, for example, in digital communications and sigma delta converters.

The phase response doesn't change by cascading CIC filter sections of different lengths. Group delay is a constant value. These cascade connections of non-identical CIC sections are therefore very useful for modifying the magnitude frequency response characteristics of filters without changing the group delay of classical CIC filters.

## ACKNOWLEDGMENTS

This paper is supported in part by the Ministry for Education, Science and Technological Development of Serbia, under projects numbered TR32052 and TR32023.

## REFERENCES

- [1] E. B. Hogenauer, "An economical class of digital filters for decimation and interpolation", *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 29, no. 2, pp. 155-162, 1981.
- [2] B. P. Stošić, V. D. Pavlović, "On design of a novel class of selective CIC FIR filter functions with improved response", *Int. J. Electron. Commun. (AEÜ)*, vol. 68, no. 8, pp. 720-729, 2014.
- [3] B. P. Stošić, V. D. Pavlović, "Design of selective CIC filter functions", *Int. J. Electron. Commun. (AEÜ)*, vol. 68, no. 12, pp. 1231-1233, 2014.
- [4] D. N. Milić, V. D. Pavlović, "A new class of low complexity low-pass multiplierless linear-phase special CIC FIR filters", *IEEE Signal Proc. Lett.*, vol. 21, no. 12, pp. 1511-1515, 2014.
- [5] B. P. Stošić, V. D. Pavlović, "Using cascaded non-identical CIC sections to improve insertion loss", *J. Circuit Syst. Comp.*, vol. 24, no. 6, 2015.
- [6] B. P. Stošić, D. N. Milić, V. D. Pavlović, "New CIC filter architecture: Design, parametric analysis and some comparisons", *IET Electron. Res.*, vol. 61, no. 3, pp. 244-250, 2015.
- [7] B. P. Stošić, D. N. Milić, V. D. Pavlović, "Design and Properties of New CIC FIR Filter Functions", *Proceedings of the Romanian Academy, Series A: Mathematics, Physics, Technical Sciences, Information Science*, vol. 16, no. 4, pp. 539-546, 2015.
- [8] B. P. Stošić, V. D. Pavlović, "Design of new selective CIC filter functions with passband-droop compensation", *Electron. Lett.*, vol. 52, no. 2, pp. 115-117, 2016.
- [9] A. Fernandez-Vazquez, G. Jovanović Doleček, "Maximally flat CIC compensation filter: Design and multiplierless implementation", *IEEE T. Circuits - II*, vol. 59, no. 2, pp. 113-117, 2012.
- [10] G. Jovanović Doleček, A. Fernandez-Vazquez, "Novel droop-compensated comb decimation filter with improved alias rejections", *Int. J. Electron. Commun. (AEÜ)*, vol. 67, no. 5, pp. 387-396, 2013.
- [11] G. Jovanović Doleček, A. Fernandez-Vazquez, "Trigonometrical approach to design a simple wideband comb compensator", *Int. J. Electron. Commun. (AEÜ)*, vol. 68, no. 5, pp. 437-441, 2014.
- [12] G. Jovanović Doleček, M. Laddomada, "An improved class of multiplierless decimation filters: Analysis and design", *Digit. Signal Process.*, vol. 23, pp. 1773-1782, 2013.
- [13] Y. Lian, Y. C. Lim, "New prefilter structure for designing FIR filters", *Electronics Letters*, vol. 29, no. 11, pp. 1034-1036, 27 May 1993.
- [14] G. J. Doleček, J. D. Carmona, "Generalized CIC-cosine decimation filter", *2010 IEEE Symposium on Industrial Electronics & Applications (ISIEA)*, Penang, Malaysia, pp. 640-645, 2010.

# Model of the Operational Conveyor based on Current-steering output stage

Slobodan Djukic<sup>1</sup> and Milan Veskovic<sup>2</sup>

**Abstract** – This study presents the improved version of the operational conveyor based on current-steering output stage. Related to the previous study the improvement refers to the fact that the required precise current distribution to the two equal parts should be performed by the Wilson's current mirrors. The suggested operational conveyor has very good voltage and current transfer characteristics, and has constant gain-bandwidth product close to 2MHz. Also, it does not require the transistors with matched characteristics. The results of the simulation circuit with the PSPICE program are also presented in this study. The operational conveyor uses the standard electronic components and is suitable for the implementation in the integrated technics.

**Keywords** – Current-mode processing technique, current mirror, current conveyor, operational conveyor, PSPICE simulation

## I. INTRODUCTION

The concept of the current conveyor (CCI) was introduced by Sedra and Smith, back in 1970. [1], and the possibility of its application in the implementation of the full-wave precision rectifier for the small signals is presented in the literature [2-5]. However, the area of application of the current conveyor is much broader. The symbol of the negative second type current conveyor (CCII<sup>-</sup>) is shown in Figure 1, and it can be mathematically described by the following matrix equation:

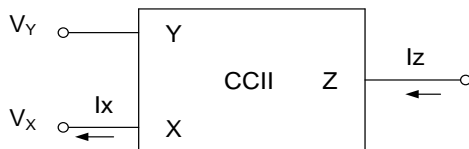


Fig. 1. Symbol of the second type current conveyor (negative)

$$\begin{bmatrix} I_Y \\ V_X \\ I_Z \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \pm 1 & 0 \end{bmatrix} \begin{bmatrix} V_Y \\ I_X \\ V_Z \end{bmatrix} \quad (1)$$

or with three linear equations:

<sup>1</sup>Slobodan Djukić is with Faculty of Technical Science Čačak, University of Kragujevac, Svetog Save 65, 32000 Čačak, Serbia E-mail: slobodan.djukic@ftn.kg.ac.rs.

<sup>2</sup>Milan Vesković is with Faculty of Technical Science Čačak, University of Kragujevac, Svetog Save 65, 32000 Čačak, Serbia E-mail: Milan.veskovic@ftn.kg.ac.rs.

$$V_x = V_y, \quad (2)$$

$$I_y = 0, \quad (3)$$

$$I_z = \pm I_x. \quad (4)$$

The voltage signal  $V_y$  is connected to the high-resistance input  $Y$  and it buffers with the unity gain at the input  $X$  and the current  $I_x$  is mapped to the output  $Z$ . The problem is that the resistance  $r_x$ , which is recognized by the inverting input of the conveyor, is not zero. This resistance is small but it does not have a constant value and it affects the error of the voltage transfer function.

This current conveyor can be mathematically described by the following matrix equation [5]:

$$\begin{bmatrix} I_Y \\ V_Y \\ I_Z \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & r_x & 0 \\ 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} V_X \\ I_X \\ V_Z \end{bmatrix} \quad (5)$$

or with three linear equations:

$$I_Y = 0 \quad (6)$$

$$V_Y = V_X + r_x I_X \quad (7)$$

$$I_Z = -I_X \quad (8)$$

The concept of operational conveyor (OC) introduced by Gift [6] completely eliminates the resistance  $r_x$ , and it consists of the voltage operational amplifier (VOA), which works with the current conveyor (CCII) in a feedback loop, as shown in figure 2.

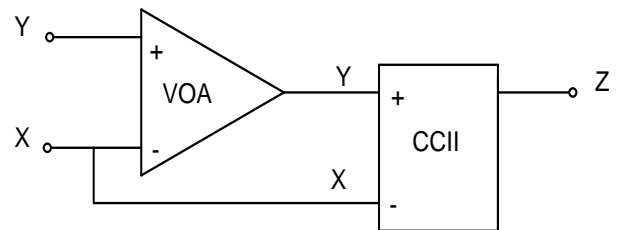


Fig. 2. Operational conveyor with CCII used as a feedback circuit

The operational conveyor can be represented as a circuit with three connectors, as shown in Figure 3.

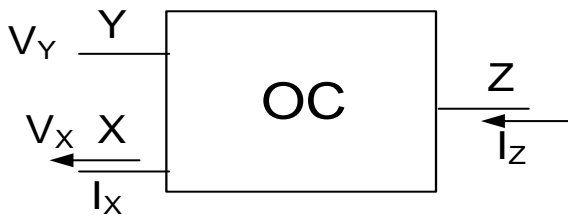


Fig. 3. Symbol of the operational conveyor

Figure 4 shows the concept of the operational conveyor, with the voltage operational amplifier and current-steering output stage as the feedback circuit.

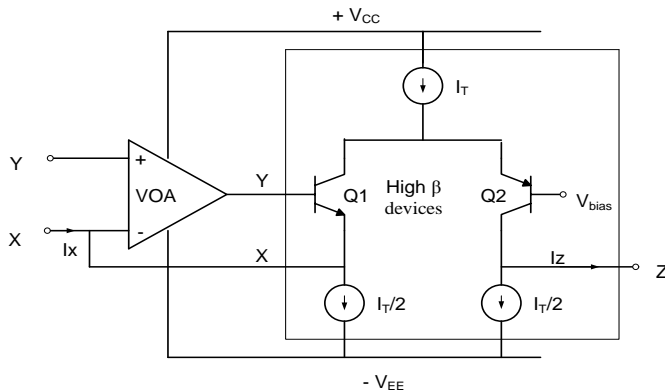


Fig. 4. Operational conveyor with CCII used as a feedback circuit – basic formulation

Good voltage transfer characteristic between the terminals Y and X is provided by the usage of the negative feedback with the voltage operational amplifier. In order to keep the error in the voltage transmission function as small as possible, it is necessary that the current  $I_T$  is as higher, so that the input resistance of the transistor  $Q_1$  be smaller comparing to the output resistance of current sources. The increasing of the current  $I_x$  at port X for some amount causes the decrease of the current of the transistor  $Q_1$  by the same amount, which then causes an increase in current through the transistor  $Q_2$ , which causes an increase at the output for the same amount. The same can be concluded for the reduction of current  $I_x$ , so it can be concluded that the relation  $I_x = -I_z$  is valid. Transistors  $Q_1$  and  $Q_2$  should have a large current gain, in order to ensure the equality of their emitter and collector currents. It is important to notice that it is necessary to provide the exact division of the current  $I_T$  into two equal parts, which is not easy, given the fact that the current  $I_T$  should be variable. Figure 4 shows that in the feedback loop of the voltage operational amplifier there is a current conveyor. The question is whether the level of the current steering stage (bounded by rectangle) is in fact the current conveyor? The answer is positive, and for the following reasons:

a) Its voltages  $V_x$  and  $V_y$  are roughly equal (they differ for the base-emitter voltage of the transistor  $Q_1$ )

b) Its current  $I_y$  value is about zero because it is the transistor  $Q_1$  base current, and it has great current amplification,

c) It has already been demonstrated that the  $I_x = I_z$

## II. DISCRETE VERSION OF THE OPERATIONAL CONVEYOR REALISATION

The complete scheme of the operational conveyor is presented in Figure 5. In relation to the literature [5], we performed the improvement in terms of the exact division of the current  $I_T$  into two equal parts using Wilson's current mirrors, but the current mirror of the unit amplifier can also be used [7]. The usage of discrete components requires their patient matching by their characteristics, although there are those ready-made components, such as CA3096AE (transistor arrays) [8]. To set the current  $I_T$ , the variable resistor  $R_2$  is used, and for the elimination of the offset voltage at the output terminal, the variable resistor  $R_3$  is used. This is also an improvement as compared to the literature [8] because there is one resistor less.

## III. PSPICE SIMULATION

In order to test the characteristics of the operational conveyor, it is presented as a voltage amplifier such as the excitation generator  $V_{in}$  is linked via the resistor  $R_{in}$  to the input X of the operational conveyor, and the Y connection to the ground, while the output Z through the load  $R_L$  is also connected to the ground. The supply voltages are  $V_{CC} = -V_{EE} = 5V$ , while the used transistors are 2n4014 (npn) and 2n4059 (pnp). Current amplification of the both transistor is approximately 180. For the value of the current  $I_T = 2mA$  the following values are used for the resistors  $R_2 = 4.6 k\Omega$  and  $R_3 = 200 k\Omega$  approximately. For the voltage operational amplifier we used TL082. The resistance  $r_x$  of such operational conveyor is below  $10 \Omega$ , which practically means that it does not exist, or that the first matrix equation which describes its behavior is valid. We have made another simulation with uA741 and get little larger value for  $r_x$ .

Figure 6 shows the frequency dependence of the output impedance of the operational conveyor for the different values for the current  $I_T$ . It is noticed that it has a constant value (the order of magnitude of  $80 k\Omega$ ) to the limit frequency which in this case is in the order of a few tens of kilohertz size. Based on Figure 6 it can be concluded that the output impedance consists of the parallel connection of resistance and capacitance that can be seen at the port Z of the operational conveyor. It would be good if the magnitude of the output impedance module had as higher value as possible, which can be achieved by reducing the value of the set current  $I_T$ , but then the scope of the possible changes of the current  $I_z$  at the output, reduces, which is not good.

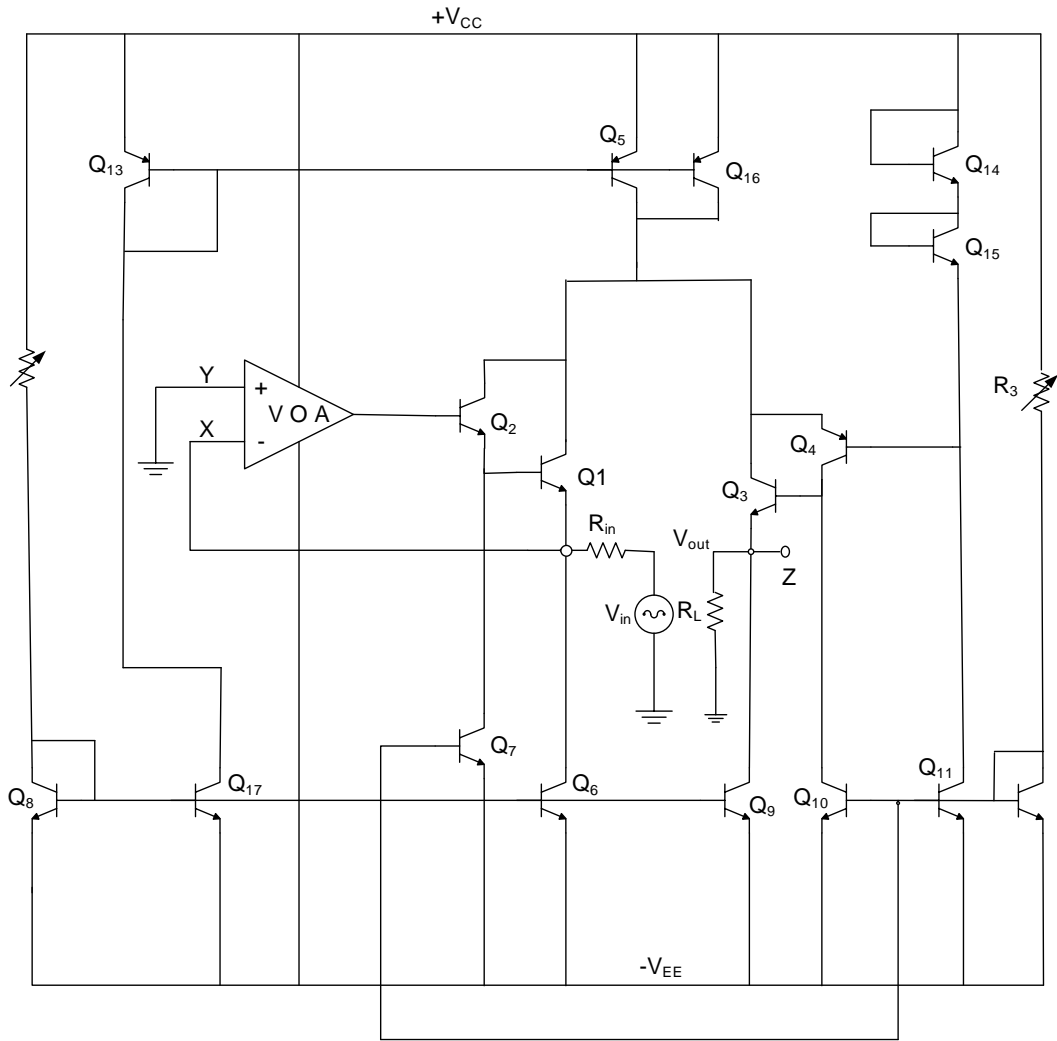


Fig 5. Discrete version of the proposed operational conveyor

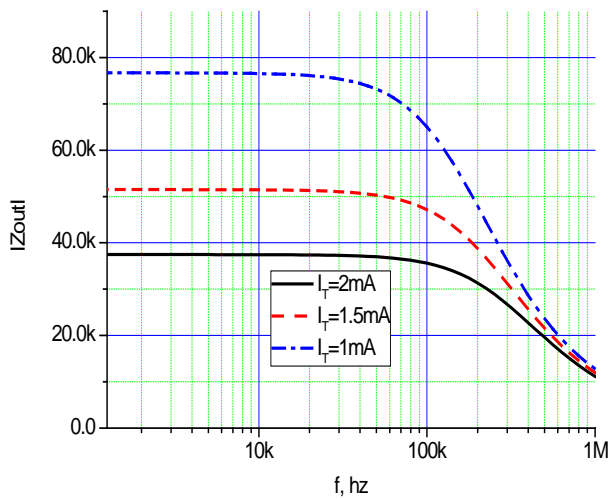


Fig. 6 Output impedance curves of the operational conveyor for different current  $I_T$ .

Figure 7 shows the amplification of the operational conveyor in the function of frequency, based on which we conclude that the amplification change does not change the

bandwidth range, and that the constancy of the product of the amplification and the bandwidth range does not apply to this operational conveyor

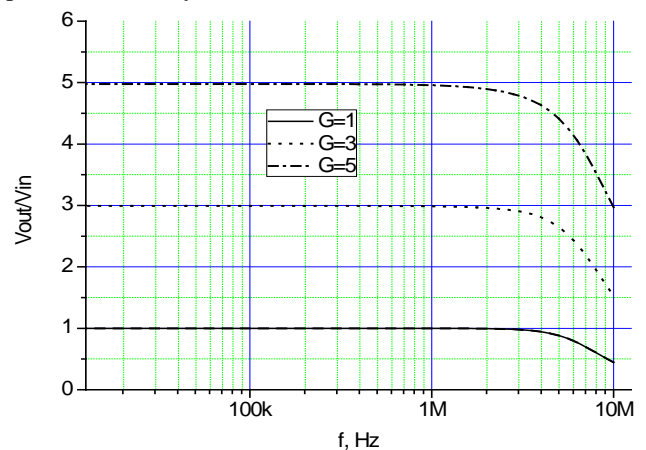


Fig. 7. Frequency response of the operational conveyor for different gain  $G$



Figure 8 and Figure 9 show the current and voltage transfer function of the conveyor. It can be seen that the transfer characteristics of the observed ranges are practically ideal

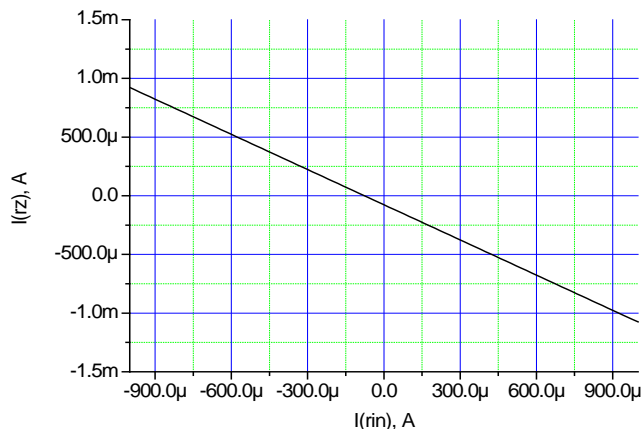


Fig. 8. Current transfer characteristic of the operational conveyor.

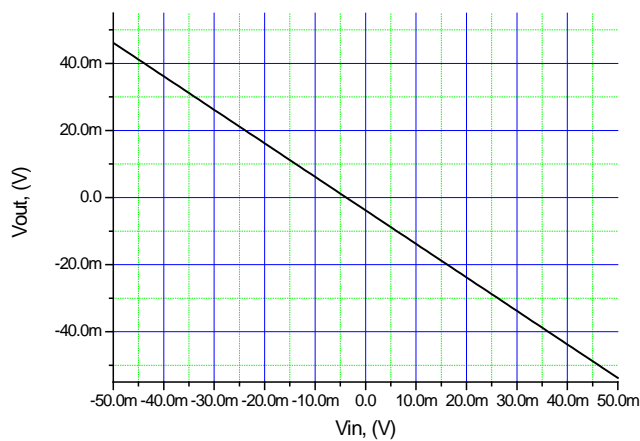


Fig. 9. Voltage transfer characteristic of the operational conveyor

#### IV. CONCLUSION

The proposed implementation of the operational conveyor based on the current steering output and current mirrors, has excellent current and voltage transfer function, because a resistance at the inverting input of the operational amplifier has a negligible value. The transient behavior of that discrete version of the operational conveyor is very well-behaved, with a well-damped response giving a constant gain-bandwidth product close to 2 MHz. Limitation of the gain-bandwidth product is primarily caused by the  $f_T$  of the pnp transistors.

#### REFERENCES

- [1] Sedra A., Smith K. C., "A second generation current conveyor and its applications", IEEE Trans. Circ. Theory, CT – 17, pp. 132 - 134, 1970.
- [2] Wilson B. "High-performance current conveyor implementation", Electronics Letters, 20, pp. 990 - 991, 1984.
- [3] Djukic R. S., Veskovic M., Vulovic A., "An improved precision full-wave rectifier for low-level signal", 9th International symposium on electronics and telecommunications (ISETC), pp 33 - 38. DOI :10.1109/ISETC. 2010.5679351., 2010
- [4] Sahu P., Sing M., Baishya A., "A Novel Versatile Precision Full-Wave Rectifier", IEEE Transactions on instrumentation and measurement, vol. 59 n.10, pp. 2742 - 2746, 2010.
- [5] Munir A. Al-Absi, "A novel highly accurate current mirror", Int. J. of Electronics, Vol. 96. No. 88, pp. 781 - 786, 2009.
- [6] Gift S.J.G., Maundy, "Versatile Precision Full-Wave Rectifiers for Instrumentation and Measurement", IEEE Trans. Instrumentation and measurement, vol. 56 n. 5, pp. 1703 - 1709, 2007.
- [7] Hart B. L., Hayatleh K. And Lidgley F. J, " A new type of unity-gain current-mirror ", Circuits and Systems, Gatineau, Quebec, Canada, IEEE North-East Workshop, june 2006, pp 89-92, 2006.
- [8] Wilson B., Al-Gahtani M., Vosper J., Deloughry R., "High Precision Current Conveyor Implementation Employing a Current-Steering Output Stage", Int. J. of Electronics, vol. 93 No. 10, pp. 653 – 662, 2006.

# The Evaluation of Speech Intelligibility in the Orthodox Church on the Basis of MOS Test Intelligibility Logatom Type CCV

Dijana Kostić<sup>1</sup>, Zoran Milivojević<sup>2</sup> and Violeta Stojanović<sup>3</sup>

**Abstract** – This paper presents the evaluation of the intelligibility of speech in the Orthodox Church on the basis of MOS test intelligibility logatom type CCV in the presence of Babble noise, for some values of the SNR. First part of the paper described the experiment and the results of the experiment are shown in graphical and tabular form. The second part is an analysis of the results and a comparison with the International Standard IEC 60268 – 16. On the basis of comparative analysis was shown that the intelligibility of speech in the Orthodox Church is poor.

**Keywords** – intelligibility of speech, the room impulse response (RIR), Babble noise, MOS, intelligibility of logatom.

## I. INTRODUCTION

From the aspect of information content the speech signal is the most important acoustic signal. It holds a huge amount of information and is therefore very important to ensure transmission quality and understanding of the messages that carries [1].

French, Steinberg and Beranek 1947 were the first who highlight the problem of transmission and speech intelligibility [2]. Kryten in 1962 introduced the articulation index, *AI*. Peutz in 1971 has developed an algorithm to predict the intelligibility of speech in auditoriums and suggested parameter which refers to the loss of articulation of consonants, *AI<sub>cons</sub>* [3]. Houtgast and Steeneken in 1980 confirmed an objective method to measure the quality of voice transmission in the room [2], and suggested the acoustic parameters of the speech transmission index, *STI* [4]. Subjective systems evaluation for transmitting voice using open and closed tests intelligibility, were first performed by Fletcher and Steinberg in 1929 [2], then Egan in 1944, Miller and Nicely in 1955, House in 1965 and Voiers in 1977 etc. Tests intelligibility was represented by words test (usually with logatoms (monosyllabic words without meaning) type CVC, VCV, CV, VC, CCVC, CVCC, (C - consonant, V-vocal)), test sentence or test of syllables. Review of tests for evaluating intelligibility of speech was given by Pols in 1991

and Steeneken in 1992.

The following factors have an influence on the speech intelligibility [1]: the width of the band, distortion in the figurative sense, the strength of speech, ambient noise and Babble noise, reverberation and influence direct and reflected sound. Babble noise is a consequence at the same time speech *N* persons (BNN) and its presence leads to the degradation of the useful acoustic signal which is a measure of the Signal-to-Noise Ratio SNR [5]. In the given physical circumstances intelligibility of speech depends on: the content of speech, speech production and precision of diction speaker, listener's concentration in speech, age, etc.

Intelligibility of speech in a language implies intelligibility logatom. The correlation between intelligibility logatom and intelligibility of speech is shown in Table I [6].

TABLE I  
THE CORRELATION BETWEEN THE INTELLIGIBILITY LOGATOMS (IL)  
AND SPEECH INTELLIGIBILITY (SI).

<i>IL</i> (%)	0÷34	34÷65	65÷85	85÷95	95÷100
<i>SI</i> (%)	bad	poor	fair	good	excellent

In a speech different phonemes (sounds) carry a different energy. Higher participation in speech (in terms of time excuse) has vocals than consonants. Also, they carry more energy than consonants. However, for intelligibility speech more important are consonants.

There are 30 phonemes (votes) [7] in the Serbian language. Phonemes by their common traits can be divided into groups: a) according to the noise level: silent (to the vocals, there are 5 and the noise (to the consonants, there are 25), b) according to the frequency of occurrence in the Serbian language phonetically balanced phoneme type CVC and the type of CCV and c) uniformly balanced logatoms all of which are equally represented in the Serbian language and forms are CVCV. The vocals are audible, while consonants may be audible and voiceless.

This paper presents the evaluation of speech intelligibility in the Orthodox Church on the basis of Mean Opinion Score (MOS) test intelligibility CCV logatom, in the presence of Babble noise, for some values of the SNR. The measurement was performed by creating a base of logatom, the impulse response of the church, the church of simulated acoustic signals and simulated acoustic signal with superimposed Babble noise the predefined relationships SNR. The effect of the acoustic environment of the church was simulated, using Matlab, using convoluting speech signal with recorded impulse responses of the church. Simulation of the effect of people who speak and disrupt the original signal is generated

<sup>1</sup> Dijana Kostić is with the College of Applied Technical Sciences of Niš, 20. Aleksandra Medvedeva, St, 18000 Niš, Serbia, e-mail: koricanac@yahoo.com

<sup>2</sup> Zoran Milivojević is with the College of Applied Technical Sciences of Niš, 20. Aleksandra Medvedeva, St, 18000 Niš, Serbia, e-mail: zoran.milivojevic@vtsnis.edu.rs

<sup>3</sup> Violeta Stojanović is with the College of Applied Technical Sciences of Niš, 20. Aleksandra Medvedeva, St, 18000 Niš, Serbia, e-mail: violeta.stojanovic@vtsnis.edu.rs

by superposition Babble noise and acoustic signals generated church. Testing was performed by the open MOS test and listeners were wrote down reproduced logatoms as they hear them. After analyzing the results and comparisons with the International Standard IEC 60268-16: 2011 brings to a conclusion of the evaluation of speech intelligibility in the church.

The organization of work is as follows: Section II explains the experiment and the results are presented. Section III presents the analysis results of the test MOS intelligibility CCV logatom basis on MOS test intelligibility CCV logatom in the presence of Babble noise, for some values of the SNR. Section IV is the conclusion.

## II. EXPERIMENTS

In the paper we made the evaluation of speech intelligibility for the Orthodox Church “Holy Martyr Procopius” in Katun (Nis, Serbia) using MOS test intelligibility logatom type CCV, in the presence of Babble noise for  $SNR = \{-10, -5, 0, 10, 20, 30\}$  dB.

Volume and the total internal surface of the Orthodox Churches are  $V = 1659.68 \text{ m}^3$  and  $S = 646.48 \text{ m}^2$ . Reverberation time  $RT = 2.06 \text{ s}$ .

For intelligibility of the MOS test signals are formed which block diagram shown in Fig. 1 where:  $\mathbf{x}$  pure speech signal,  $\mathbf{h}$  the impulse response of the church,  $\mathbf{y}$  generated acoustic signal and  $\mathbf{z}$  generated acoustic signal with superimposed noise BBN.

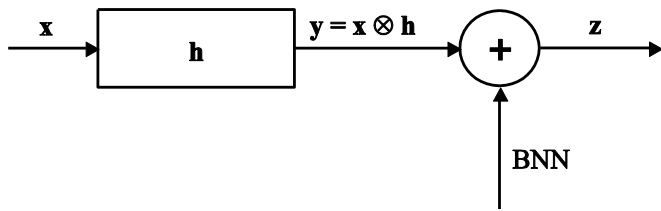


Fig. 1. Block diagram of the signal formed by MOS test intelligibility logatom type CCV

A MOS intelligibility test was conducted with 30 listeners (15 female and 15 male, age  $18 \div 25$ ), which are recorded phonetically balanced logatoms output from the player. The MOS test contains a list of 3 types of phonetically balanced logatom CCV (Table II). The results are analyzed on the basis of correct and incorrect written answers.

Intelligibility logatom,  $LI$ , in a measurement point for each MP for each listener in the presence of Babble noise with specific values SNR is calculated according to Eq. 1:

$$LI(\%) = \frac{NEWL}{NLS} \cdot 100, \quad (1)$$

where:  $NEWL$  - the number exactly written logatom and  $NLS$  - the number of words spoken logatom. Calculated the average value of each type of intelligibility logatom and all logatom for all listeners in the MP for a specific value SNR.

### A. The Basis

The base of the experiment consists of: **1)** the base of the original signal  $\mathbf{x}$ : recorded 6 speakers (3 men and 3 women, age  $18 \div 25$ ) who read test made the list of the three types phonetically balanced logatom CCV of 30 words; **2)** a base impulse response of the church: the database includes wav files that were obtained by recording the acoustic impulse response using the software package EASERA. Recordings were made at measuring point MP, which is shown in Fig.2.

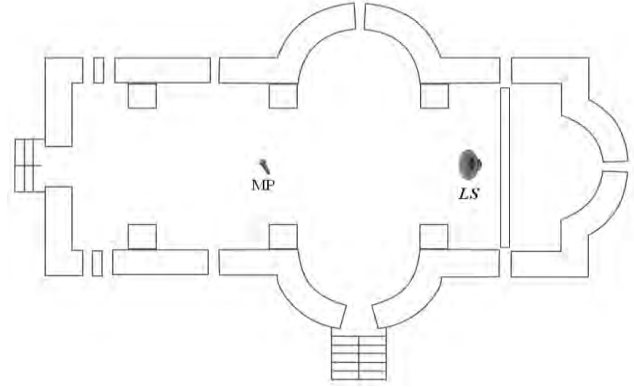


Fig. 2. The position of measuring point MP and sound source LS in the church during the recording impulse response.

Measuring point MP is 7m away from the sound source LS. 7 measurements were carried out, making a total of 7 wav files. Impulse response measurement was performed by the excitation sweep signal for a period of 6 s. The Sampling frequency is  $f_s = 44.1 \text{ kHz}$ . The measurement of the impulse response of the church was performed in accordance with ISO 3382 [8]. **3)** the base of a simulated acoustic signal: the effect of the acoustic environment of the church was simulated, using Matlab, convolution (signal  $\mathbf{y}$ ) of the speech signal from the base  $\mathbf{x}$ , with recorded impulse responses of the church to the measuring point MP. **4)** base of Babble noise BN: Babble noise was recorded with 8 speakers (BN8), 4 male and 4 female (age  $18 \div 25$ ) (signal  $\mathbf{BNN}$ ) and **5)** of the base of a simulated acoustic signal with superimposed BN8 noise for  $SNR = \{-10, -5, 0, 10, 20, 30\} \text{ dB}$ , (signal  $\mathbf{z}$ ).

### B. The Results

Table II shows the intelligibility logatom CCV type 1, type 2 and type 3. Table III shows the middle values of intelligibility logatom CCV type 1, type 2, type 3 and the middle value of the intelligibility of all types logatom observed for pure signal to the input signal  $\mathbf{x}$  and the impulse response of the church,  $\mathbf{y}$ . Figs. (3 – 5) shows the intelligibility of CCV logatom type 1  $SNR = \{-10, 10\} \text{ dB}$ , type 2 for  $SNR = \{-5, 20\} \text{ dB}$  and type 3 for  $SNR = \{0, 30\} \text{ dB}$ , respectively. Fig. 6. shows the intelligibility of all analyzed CCV logatom on these values SNR. On figures at the abscissa are ordinal numbers that are shown in Table II.

TABLE II  
INTELLIGIBILITY CCV LOGATOM.

CCV type 1										
ord. num.	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
logatom	BLE	ŠLJA	DNJE	BRI	VRE	PLJE	RNO	DRU	TRA	GDI
x (%)	66.67	77.78	100	100	100	22.23	100	100	100	88.89
logatom	ČNU	SMI	TNI	FRA	PRO	SKO	SKE	PRI	VRA	STU
y (%)	0	0	55.56	77.78	55.56	77.78	55.56	77.78	88.89	22.23
logatom	SPO	BRE	PRU	TRO	GLE	GRU	BRA	TNA	VNE	ŠKO
$z _{SNR=-10dB}$	0	0	11.11	0	0	0	11.11	0	0	0
$z _{SNR=10dB}$	0	44.44	55.56	88.89	55.56	66.67	55.56	33.33	0	11.11
CCV type 2										
ord. num.	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
logatom	SHO	DRE	KSI	ČKU	BRE	HVA	ZNA	ŠKE	KNI	TLA
x (%)	81.82	100	100	81.82	100	90.91	100	90.91	90.91	81.82
logatom	KRI	VNO	DRA	SLO	ČNO	SLI	SKI	PRU	SJI	DRI
y (%)	54.55	9.09	72.73	9.09	18.18	45.45	54.55	18.18	0	0
logatom	ŠLU	PNO	DRO	GRA	DRU	TNE	DRA	SRE	PRE	SNA
$z _{SNR=-5dB}$	0	0	54.55	18.18	45.45	0	45.45	18.18	45.45	0
$z _{SNR=20dB}$	0	0	45.45	45.45	63.64	0	63.64	54.55	81.82	63.64
CCV type 3										
ord. num.	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
logatom	TRA	ŽDO	TRI	SPI	KTE	ČNA	KLE	PNE	SVA	ZLA
x (%)	70	30	90	80	80	30	60	40	90	100
logatom	TVA	MNE	SLU	TSO	SPE	DNE	SRA	SNO	SKO	ŠLU
y (%)	10	0	0	30	50	0	20	0	20	10
logatom	STE	DNO	PRA	ŠTE	KSA	MNA	SMA	PRI	STI	SVE
$z _{SNR=0dB}$	0	0	20	0	0	0	0	20	0	20
$z _{SNR=30dB}$	70	0	40	50	0	0	0	50	40	10

TABLE III  
THE MIDDLE VALUE INTELLIGIBILITY CCV LOGATOMS.

Signal	SNR (dB)	Type of logatom			
		CCV type 1	CCV type 2	CCV type 3	CCV
x (%)	∞	85.56	91.82	67	81.46
y (%)	∞	51.11	27.27	14	30.79
z (%)	-10	1.11	/	/	
	-5	/	22.73	/	
	0	/	/	6	
	10	41.11	/	/	
	20	/	41.82	/	
	30	/	/	26	

### III. THE RESULTS ANALYSIS

1. Based on the results for the user CCV logatom after signals (Table II and Figures (3 - 6) concluded that: a) 46.67% logatom CCV presented by pure speech signal has excellent,

at 23.33% good, and fair, and 6.67% have a bad intelligibility; b) 60% logatom represented CCV the input signal with the impulse response of the church has a bad, 23.33% poor, 13.33% fair and 3.33% have a good intelligibility; c) all CCV logatoms represented by the input signal with the impulse response of the church and Babble noise with SNR = {-10, -5, 0 dB} have the highest percentage of bad intelligibility: 100%, 60% and 100%, respectively; 40% of these signals with the SNR = -5 dB are poor intelligibility; d) CCV logatoms signal presented in the impulse response of the church and Babble noise with SNR = 10 dB has 10% good and fair intelligibility a 40% poor and bad intelligibility; e) logatoms signal presented to the impulse response of the church and Babble noise with SNR = 20 dB have the highest percentage of poor intelligibility, 60%, followed by bad with 30% and fair intelligibility with 10%; f) logatoms signal presented in the impulse response of the church and Babble noise with SNR=30dB has the highest percentage of poor intelligibility, 50%, then bad, 40% and fair intelligibility only 10%.

2. Based on the middle value intelligibility logatom CCV by signals (Table III.) concludes that: a) for all CCV logatoms presented pure input signal intelligibility is fair (81.46%). The only logatoms CCV type 2, with middle value intelligibility



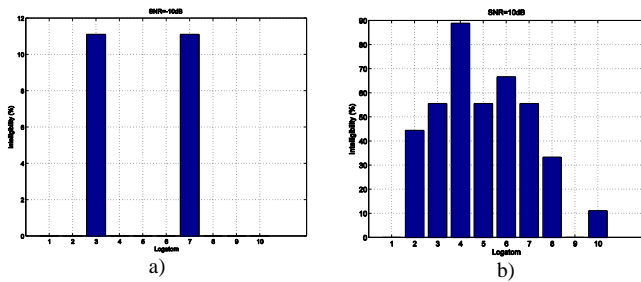


Fig. 3. Intelligibility CCV logatom type 1: a) SNR = -10 dB and b) SNR = 10 dB

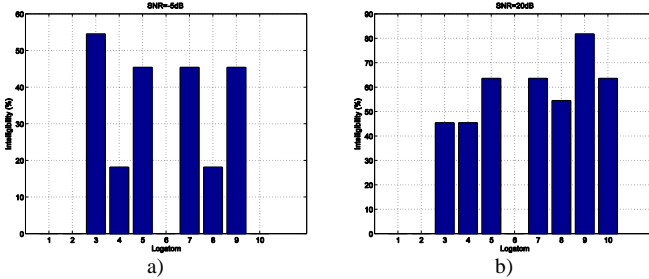


Fig. 4. Intelligibility CCV logatom type 2: a) SNR = -5 dB and b) SNR = 20 dB.

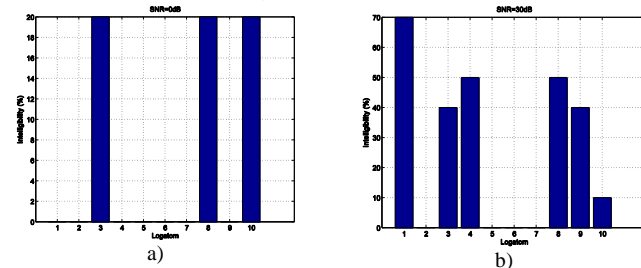


Fig. 5. Intelligibility CCV logatom type 3: a) SNR = 0 dB and b) SNR = 30 dB.

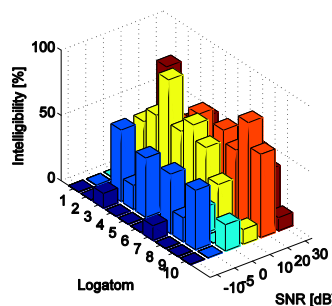


Fig. 6. Intelligibility CCV logatom for SNR = {-10, -5, 0, 10, 20, 30} dB.

that belongs to the good intelligibility while logatoms CCV type 1 have a fair intelligibility (85.56%), and logatoms CCV type 3 have a bad intelligibility (67%); b) middle value intelligibility logatom CCV type 1 (51.11%), type 2 (27.27%) and type 3 (14%) represented by the input signal with the impulse response of the church belongs to the values of bad intelligibility. This has the effect that the middle value of intelligibility of all these signals (30.79%) belongs to the range of values which classifies the signal into bad intelligibility.

3. All CCV logatoms presented with impulse response signals and Babble noise with one of the selected values SNR = {-10, -5, 0, 10, 20, 30} dB have a middle value of

intelligibility, which puts them in a bad intelligibility. The highest percentage of bad intelligibility is presented logatoms signals with SNR = 20 dB, 41.82%. The lowest percentage of bad intelligibility is presented logatoms signals with SNR = -10 dB, 1.11%.

Based on the results of the experiment it is concluded that the greatest value of the percentage of intelligibility CCV logatom presented with a pure speech signal, 46.67%, classifies speech intelligibility in the Orthodox Church as an excellent intelligibility. However, already the largest percentage of intelligibility CCV logatom presented through the signals from the impulse response of the church, 60% classify speech intelligibility in the Orthodox Church as a bad intelligibility. Bad speech intelligibility in the Orthodox Church confirmed by the percentage of intelligibility logatom presented with impulse response signals and Babble noise with values SNR = {-10, -5, 0, 10, 20, 30} dB, which was expected.

#### IV. CONCLUSION

This paper is based on a comparison of the results obtained MOS test for the evaluation of subjective intelligibility logatom CCV type in the presence of Babble noise from the Orthodox Church and the comparable value for the intelligibility logatom and intelligibility of speech given to the International Standard IEC 60268-16.

Based on the above performed analysis showed that CCV logatoms represented signals with impulse response and Babble noise with values SNR = {-10, -5, 0, 10, 20, 30} dB have a middle value of subjective intelligibility  $6\% \div 41.82\%$ . On this basis, it is concluded that the intelligibility of speech in the ambience of the Orthodox Church is *bad*. The reason for bad speech intelligibility in the ambience of the Orthodox Church is definitely the huge value of the reverberation time.

#### REFERENCES

- [1] H. Kuttruff, Room acoustics, fifth edition, Spoon Press, 2009.
- [2] H. J. M. Steeneken, "On measuring and predicting speech intelligibility", Academisch Proefschrift, Soesterberg, 1992.
- [3] Peutz, V. M. A. "Articulation Loss of Consonants as a criterion for Speech Transmission in a Room", J. AUDIO – ENG. SOC. Vol 19, p. 915 - 919, 1971.
- [4] Steeneken, H. J. M., "Apsyhical Method for measuring Speech Transmission Quality", J. ACOUST. SOC. AM., Vol 19, 1980.
- [5] N. Krishnamurthy, Student member, IEEE, and John H.L. Hansen, Fellow IEEE, "Babble Noise: Modeling, Analysis and Applications", IEEE Transactions on audio, speech and language processing, vol 17, no.7 September 2009.
- [6] International Electrotechnical Commission IEC 60268-16 – International Standard: Sound system equipment – Part 16: Objective rating of speech intelligibility by speech transmission index, Switzerland: IEC, 2011.
- [7] Lj. Subotić, D. Sredojević, I. Bjelaković, Fonetika i fonologija: ortoepska i ortografska norma standardnog srpskog jezika, Filozofski fakultet, Novi Sad, 2012.
- [8] ISO 3382: 1997, Acoustic - Measurement of the Reverberation Time of Rooms with Reference to Other Acoustical Parameters, 1997.

# Research and signal analysis with different classes of wavelets

Mariana Todorova<sup>1</sup> and Reneta Ivanova<sup>2</sup>

**Abstract** – The theory of elementary waves is a powerful alternative to the classic Fourier analysis and it assures a more flexible technique for processing signals. Wavelet analysis allows the examination of data while having the capabilities to compress, filter and analyze information. In this paper signal analysis using different wavelets is made. The obtained results are compared and appropriate conclusions are made.

**Keywords** – Signal analysis, Wavelets, Wavelet Toolbox, Wavelet Transformation

## I. INTRODUCTION

In the recent years a special kind of signals called elementary waves (wavelet functions) is actively used in modern theory and practice. The wavelets are particularly effective in performing spectrum analysis and signal compression. They are one of the newest tools for decomposition of functions or continuous time signals in frequency components and the study of each frequency component with a resolution corresponding to its scale. Basically they are considered as an alternative to the Fourier transform. The wavelet transformation is one of the most popular of the time-frequency-transformations.

Wavelet analysis has begun to play a serious role in a broad range of applications, including data and image compression, signal processing, solution of partial differential equations, statistics, etc.

Wavelet analysis [1], [2], [3], [4], [6] has significant advantages that stem from the very properties of wavelet functions. The wavelet transformation allows the components of a non-stationary signal to be analyzed. The wavelets also allow filters to be constructed for stationary and non-stationary signals.

In the present paper signal analysis using wavelets is made. Haar, Daubechies (db2, db4, db7), Coiflets (coif3, coif4), Symlets (sym4, sym5, sym7) and discrete Meyer (dmey) wavelets are applied. The values of the retained energy of the considered signal are compared and appropriate conclusions are made.

## II. WAVELETS AND WAVELET TRANSFORMATION

Wavelets are a class of a functions used to localize a given function in both space and scaling. A family of wavelets can be

constructed from a function  $\psi(t) \in L_2(R)$ . It denote the vector space of measurable functions that are absolutely and square integrable:

$$\int_{-\infty}^{\infty} |\psi(t)| dt < \infty \quad (1)$$

$$\int_{-\infty}^{\infty} |\psi(t)|^2 dt < \infty \quad (2)$$

Sometimes  $\psi(t)$  is known as a "mother wavelet" which is confined in a finite interval. "Daughter wavelets"  $\psi_{a,b}(t)$  are then formed by translation  $b$  and contraction  $a$ .

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right), a, b \in R, a \neq 0 \quad (3)$$

where  $a$  is positive and defines the scale and  $b$  is any real number and defines the shift. The pair  $(a, b)$  defines a point in the right half plane  $R_+ \times R$ .

There are a variety of wavelets. In this paper some of the most popular wavelets are used.

The **Haar** wavelet is the simplest possible wavelet. The mother wavelet function can be described as

$$\psi(t) = \begin{cases} 1, & 0 \leq t < 1/2 \\ -1, & 1/2 \leq t < 1 \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

Its scaling function  $\phi(t)$  is given in Eq.5.

$$\phi(t) = \begin{cases} 0, & 0 \leq t < 1 \\ 1, & \text{otherwise} \end{cases} \quad (5)$$

The **dbN** wavelets are the Daubechies' extremal phase wavelets (N refers to the number of vanishing moments).

The symlet wavelets (**symN** wavelets) are also known as Daubechies' least-asymmetric wavelets. The symlets are more symmetric than the extremal phase wavelets. In symN, N is the number of vanishing moments.

Coiflet scaling functions also exhibit vanishing moments. In **coifN**, N is the number of vanishing moments for both the wavelet and scaling functions.

The Meyer wavelet (**dmey**) is an orthogonal wavelet.

The wavelet transformation is often compared with the Fourier transformation, in which signals are represented as a sum of sinusoids. The main difference is that wavelets are localized in both time and frequency whereas the standard Fourier transformation is only localized in frequency.

Wavelet transforms are classified into discrete and continuous wavelet transforms.

A continuous wavelet transform (CWT) is used to divide a continuous-time function into wavelets. The CWT of a function  $f(t) \in L_2(R)$  is defined as

$$CWT_f(a, b) = \int_{-\infty}^{\infty} \psi_{a,b}(t) \cdot f(t) \cdot dt \quad (6)$$

<sup>1</sup>Mariana Todorova is with the Faculty of Informatics and Automatics at the Technical University of Varna, 1 Studentska Str., Varna 9005, Bulgaria, E-mail: mgtodorova@tu-varna.bg.

<sup>2</sup>Reneta Ivanova is PhD student with the Faculty of Informatics and Automatics at the Technical University of Varna, 1 Studentska Str., Varna 9005, Bulgaria, E-mail: renetaivanova1@gmail.com.

### III. SIGNAL COMPRESSION USING WAVELETS

In this section compression of a signal representing the sum of two sinusoids (Fig. 1) is made. Haar, Daubechies, Coiflets, Symlets and discrete Meyer wavelets are applied. Graphical user interface in Matlab is used for this purpose. It provides point-and-click control of software applications, eliminating the need to learn a language or type commands in order to run the application. The obtained results by using Haar wavelets are given on Figs. 1-3.

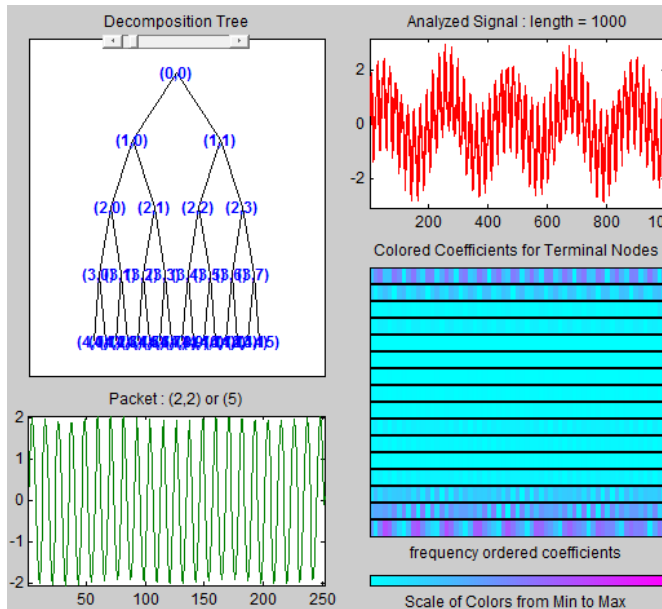


Fig. 1. Decomposition of the signal (Haar wavelets)

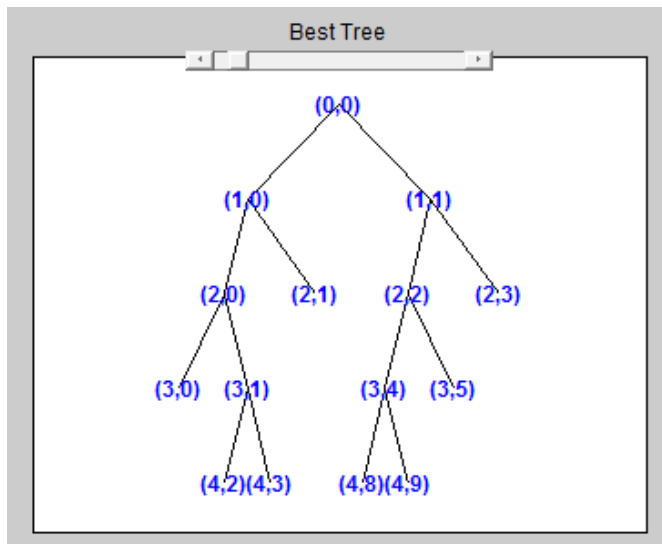


Fig. 2. Best Tree (Haar wavelets)

The original signal and compressed signal using Daubechies wavelets db2, bd4 and db7 are given on Figs.4, 5 and 6.

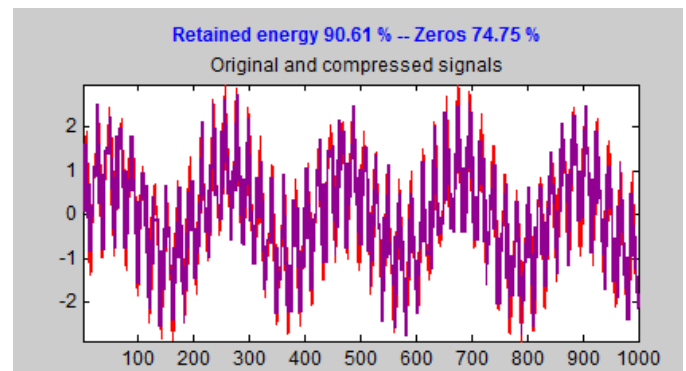


Fig. 3. Original and compressed signals using Haar wavelets

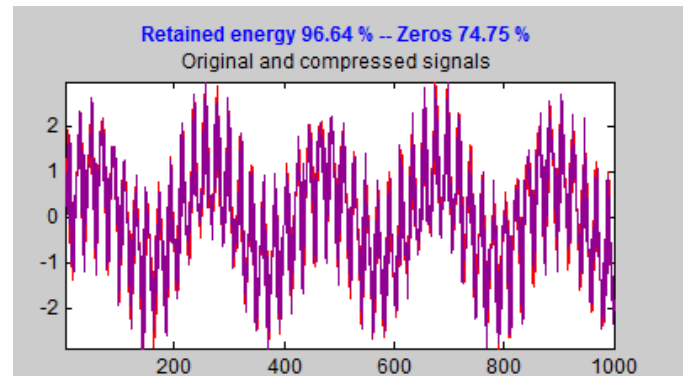


Fig. 4. Original and compressed signals using db2

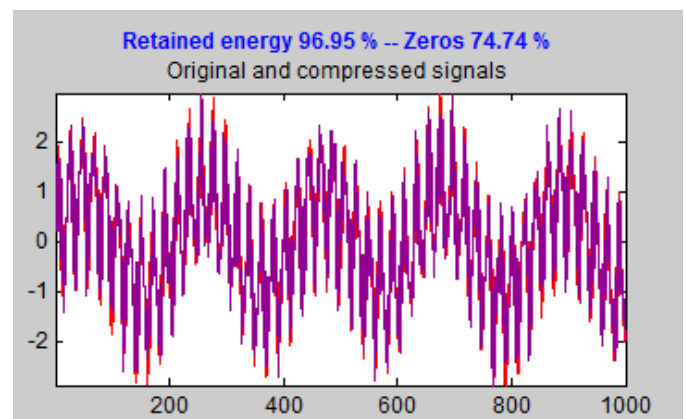


Fig. 5. Original and compressed signals using db4

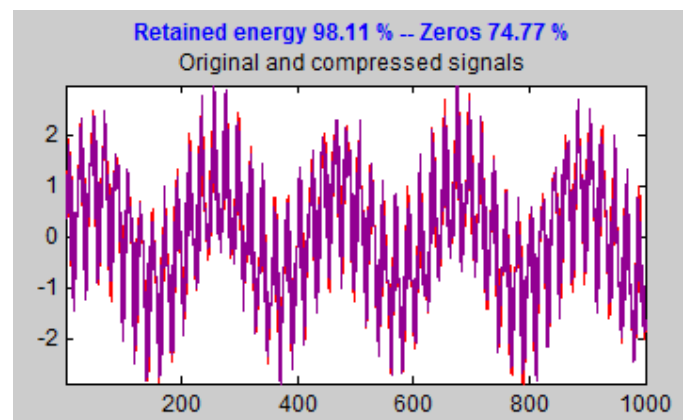


Fig. 6. Original and compressed signals using db7

The obtained results by using Coiflet wavelets are given on Figs.7 and 8.

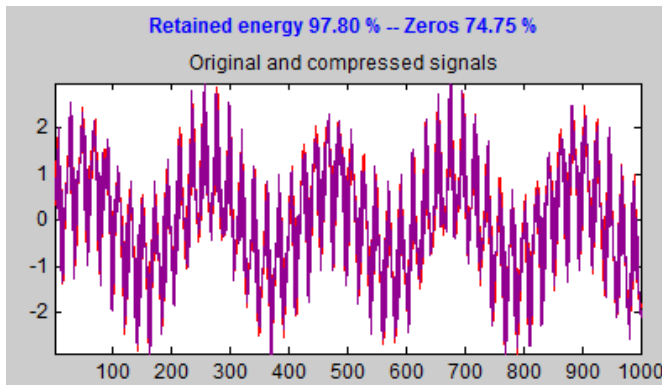


Fig. 7. Original and compressed signals using coif3

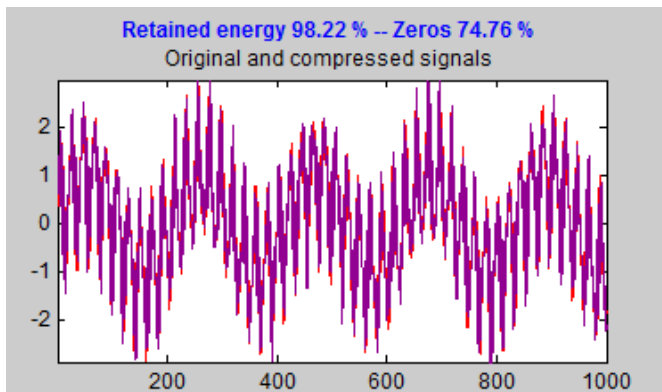


Fig. 8. Original and compressed signals using coif4

The results by using Symlet wavelets are given on Figs.9-11.

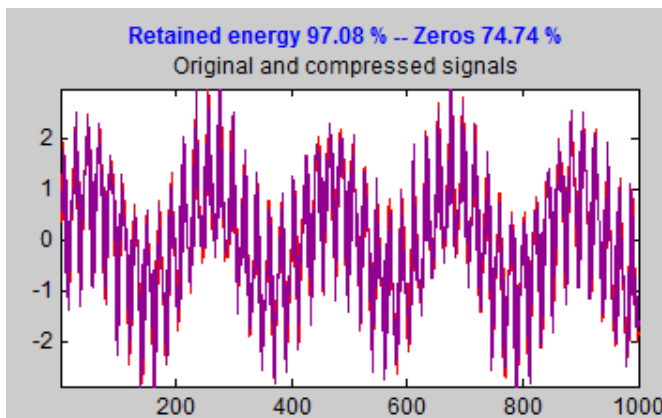


Fig. 9. Original and compressed signals using sym4

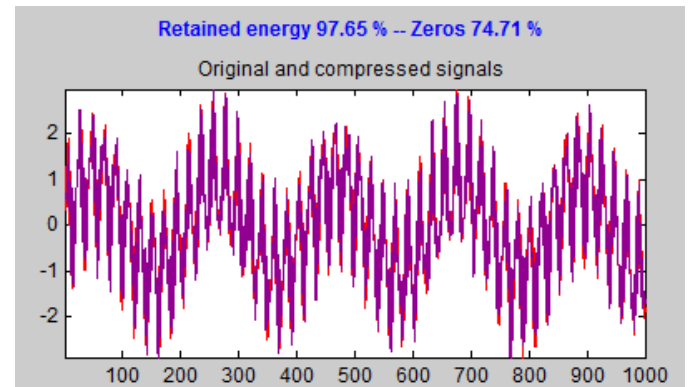


Fig. 10. Original and compressed signals using sym5

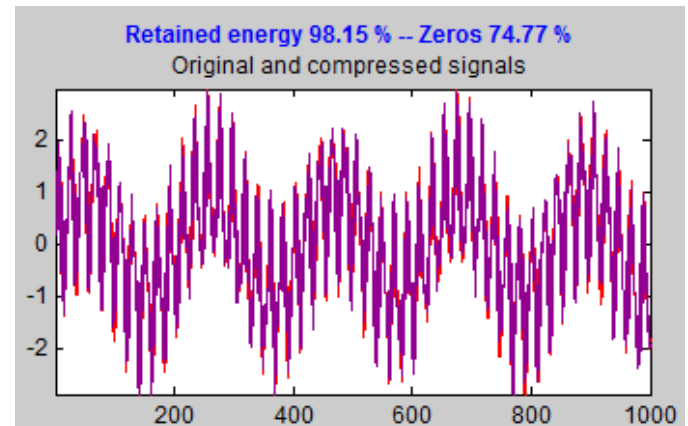


Fig. 11. Original and compressed signals using sym7

The obtained results by using discrete Meyer wavelets are given on Figs.12, 13 and 14.

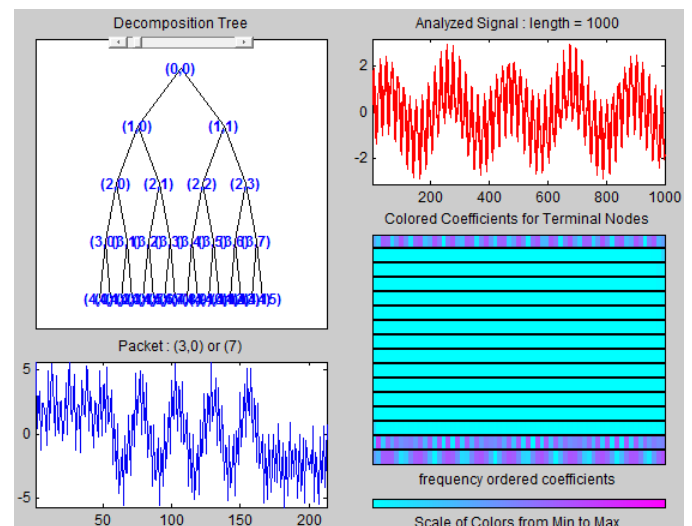


Fig. 12. Decomposition of the signal (dmey)



#### IV. CONCLUSION

From the analysis of the so considered signal with different types of wavelet functions can be seen that the compression with the same parameters gives different results. When setting a threshold by which to achieve the same number of zeros of the wavelet coefficients (74.7%), it is preferable to use a discrete Meyer function. The saved energy using discrete Meyer function is 99.61%. The most accurate results are obtained using a Haar function. In Fig. 3 are visible relatively large differences between the original and the compressed signal.

The advantage of the discrete Meyer wavelet over the others can be seen in figure 14, which shows excellent overlaying of the compressed signal on the original.

#### REFERENCES

- [1] Adhemar Bultheel, Wavelets with applications in signal and image processing, pp. 181, 2006.
- [2] Charles K. Chui, series editor, Wavelet analysis and its applications, Academic press, San Diego, CA 92101-4495, USA, 1992.
- [3] Sifuzzaman M. , M.R. Islam, M.Z. Ali, Application of Wavelet Transform and its Advantages Compared to Fourier Transform, Journal of Physical Sciences, Vol. 13, 121-134 ISSN: 0972-8791, 2009.
- [4] Sheng, Y., Wavelet Transform. The Transforms and Applications Handbook: Second Edition. Ed. Alexander D. Poularikas, CRC Press, ISBN 0-8493-8595-4, 2000.
- [5] Todorova M, R. Ivanova, Research on wavelet function applications in system identification, Computer science and technologies, vol. 2, pp. 44-55, ISSN 1312-3335, Bulgaria, 2015.
- [6] Walnut, David F., An Introduction to Wavelet Analysis, Springer, ISBN:9781461265672, 2004.
- [7] Wavelet Toolbox User's Guide, The MathWorks, Inc., 2016.

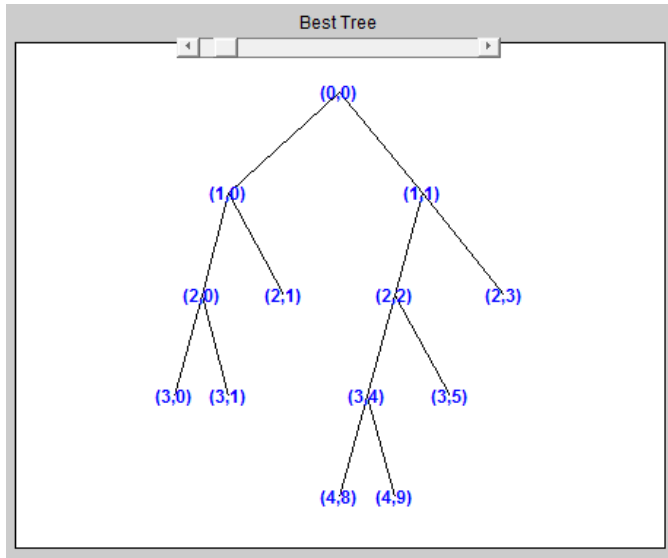


Fig. 13. Best Tree (dmey)

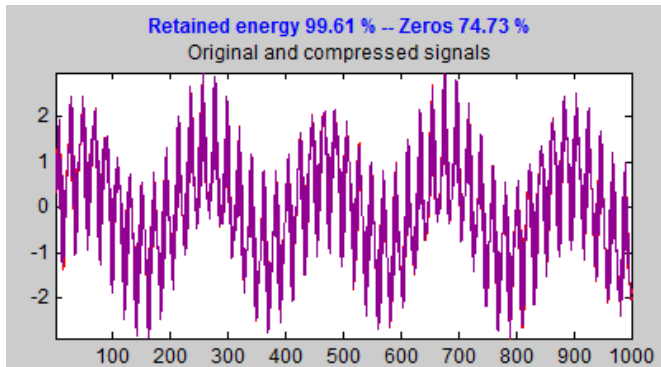


Fig. 14. Original and compressed signals using dmey

The obtained values of the retained energy of the signal using different wavelets are compared in Table I.

Table I  
Retained energy

Wavelets	Global threshold	Number of zeros %	Retained energy %
Haar	1.002	74.75	90.61
db2	1.101	74.75	96.64
db4	0.9614	74.74	96.95
db7	0.7381	74.77	98.11
Coif3	0.7835	74.75	97.80
Coif4	0.6691	74.76	98.22
Sym4	0.9257	74.74	97.08
Sym5	0.847	74.71	97.65
sym7	0.7186	74.77	98.15
dmey	0.5301	74.73	99.61

# Optimization of the parameter of 2P Keys interpolation kernel according to the criterion of similarity to sinc function in the spectral domain

Nataša Savić<sup>1</sup>, Zoran Milivojević<sup>1</sup>, Zoran Veličković<sup>1</sup>, Darko Brodić<sup>2</sup>

**Abstract** – This paper provides the optimization of the parameters of two-parameter Keys kernel. Optimization was performed in the spectral domain. The optimization criterion is the condition that the amplitude characteristic is a good approximation of the amplitude characteristic of the ideal interpolation kernel of the form  $\sin x/x$ . The second part of the paper shows an experimentally determined optimal value of the kernel parameters for an audio signal.

**Keywords** – Convolution, Interpolation kernel, Optimal value parameter.

## I. INTRODUCTION

The measurement results are mainly expressed on a discrete set of points. Based on this data intermediate values are estimated. This process is called interpolation. The cubic convolution interpolation is usually used for working in real time [1-8]. It is suitable because it uses cubic interpolation kernels and presents a compromise between the speed of performance and the numerical precision. Parametric kernels are a very significant class of cubic interpolation kernels. They are suitable because by making a choice of parameter values it is possible to control the efficacy of the use of kernels for solving various problems. A greater number of algorithms were developed to be used for optimizing kernel parameters. Optimization is performed in the time and spectral domain. [4] suggests the application of parametric kernel for image reconstruction, and there is an algorithm for determining optimum parameter values,  $\alpha_{opt}$ . The kernel was later termed Keys one-parameter (1P) kernel. The optimum value of Keys (1P) kernel in application to image processing is  $\alpha_{opt}=-0.5$  [5] provides an algorithm for determining  $\alpha_{opt}$  for

<sup>1</sup>Nataša Savić is with the College of Applied Technical Sciences Niš, Aleksandra Medvedeva 20, Niš 18000, Serbia, E-mail: [natasa.savic@vtsnis.edu.rs](mailto:natasa.savic@vtsnis.edu.rs).

<sup>1</sup>Zoran Milivojević is with College of Applied Technical Sciences Niš, Aleksandra Medvedeva 20, Niš 18000, Serbia, E-mail: [zoran.milivojevic@vtsnis.edu.rs](mailto:zoran.milivojevic@vtsnis.edu.rs)

<sup>1</sup>Zoran Veličković is with the College of Applied Technical Sciences Niš, Aleksandra Medvedeva 20, Niš 18000, Serbia, E-mail: [zoran.velickovic@vtsnis.edu.rs](mailto:zoran.velickovic@vtsnis.edu.rs).

<sup>2</sup>Darko Brodić is with the University of Belgrade, Technical Faculty in Bor, Serbia, E-mail: [dbrodic@tf.bor.ac.rs](mailto:dbrodic@tf.bor.ac.rs)

Keys 1P kernel in the spectral domain through the application of Taylor expansion. The paper [6] shows how to determine  $\alpha_{opt}$  based on the slope of the amplitude characteristic on the border of the low-pass and high pass range. The paper demonstrates [8] that by using the expansion of the amplitude characteristic in the Taylor series based on the criterion of reducing the wiggles of the amplitude characteristic in the low-pass range we can determine the optimum parameters of the Keys 2P kernel  $\alpha_{opt}=-44/81$  i  $\beta_{opt}=7/81$ .

This paper also presents the optimization of the parameters of 2P Keys kernel. The optimization was carried out in the spectral domain. The criterion of the optimization is the similarity of the amplitude characteristics. Optimum parameter values were obtained by minimizing the deviations of the spectral characteristic in the low-pass and high pass range compared to the spectral characteristic of the ideal interpolation kernel in the form  $\sin x/x$ . By the application of 2P Keys kernel an interpolation of an audio signal was performed (tones  $G_1-G_7$  piano August Forster) and the optimum values of the parameters  $\alpha_{opt}$  and  $\beta_{opt}$  were obtained. Thereafter, a statistical analysis of the results was performed. The results are presented in graphs and tables.

The paper is structured as follows: section II shows an analytical form of Keys 2P interpolation kernels. Optimization of the parameters of Keys 2P kernel based on the criterion of similarity between the amplitude characteristic to the characteristic of the sinc function is presented in the section III. The experimental results and the analysis of the results are shown in section IV. Section V contains the conclusion.

## II. KEYS TWO PARAMETER INTERPOLATION KERNEL

This section first describes the analytical form of 2P Keys kernel suggested in [7].

The two-parameter kernel in the form:

$$r(x) = \begin{cases} (\alpha + \beta + 2)|x|^3 - (\alpha + \beta + 3)|x|^2 + 1; & 0 < |x| \leq 1, \\ \alpha|x|^3 - (5\alpha - \beta)|x|^2 + (8\alpha - 3\beta)|x| - (4\alpha - 2\beta); & 1 < |x| \leq 2, \\ \beta|x|^3 - 8\beta|x|^2 + 21\beta|x| - 18\beta; & 2 < |x| \leq 3, \\ 0; & 3 > |x| \end{cases} \quad (1)$$

where  $\alpha$  and  $\beta$  are the kernel parameters is well-known in the literature as Keys 2P kernel. For  $\beta=0$  we obtain 1P Keys kernel (2) and it can be written in the form of the sum of the components:

$$r(x) = r_0(x) + \alpha r_1(x) + \beta r_2(x). \quad (2)$$

where:

$$r_0(x) = \begin{cases} 2|x|^3 - 3|x|^2 + 1, & |x| \leq 1 \\ 0, & |x| > 1 \end{cases} \quad (3)$$

$$r_1(x) = \begin{cases} |x|^3 - |x|^2, & |x| \leq 1 \\ |x|^3 - 5|x|^2 + 8|x| - 4, & 1 < |x| \leq 2, \\ 0, & |x| > 2 \end{cases} \quad (4)$$

and

$$r_2(x) = \begin{cases} |x|^3 - |x|^2, & 0 < |x| \leq 1 \\ |x|^2 - 3|x| + 2, & 1 < |x| \leq 2 \\ |x|^3 - 8|x|^2 + 21|x| - 18, & 2 < |x| \leq 3 \\ 0, & 3 < |x| \end{cases} \quad (5)$$

#### B. Amplitude characteristic of the kernel

By applying the Fourier transformation (FT) on the kernel  $r(x)$  the amplitude characteristic of the kernel is obtained:

$$H(f) = FT(r(x)) = H_0(f) + \alpha H_1(f) + \beta H_2(f) = H(f, \alpha, \beta), \quad (6)$$

where:

$$H_0(f) = FT(r_0(x)) = \int_{-\infty}^{\infty} r_0(x) e^{-2\pi x f i} dx, \quad (7)$$

$$H_1(f) = FT(r_1(x)) = \int_{-\infty}^{\infty} r_1(x) e^{-2\pi x f i} dx, \quad (8)$$

and

$$H_2(f) = FT(r_2(x)) = \int_{-\infty}^{\infty} r_2(x) e^{-2\pi x f i} dx, \quad (9)$$

of the spectral component of the kernel. By using (3), (4) and (5) in (7), (8) and (9) respectively after the application of partial integration, spectral components of the kernel can be written as follows:

$$H_0(f) = \frac{6 \sin^2(\pi f) - 3\pi f \sin(2\pi f)}{2\pi^4 f^4}, \quad (10)$$

$$H_1(f) = \frac{3 \sin^2(2\pi f) - 4\pi f \sin(2\pi f) - \sin(4\pi f)}{2\pi^4 f^4}, \quad (11)$$

and,

$$H_2(f) = \frac{1 - 2 \sin^2(\pi f)}{2\pi^4 f^4} \cdot (6 \sin^2(2\pi f) - 6 \sin^2(\pi f) - 6\pi f \sin(2\pi f) - 2\pi f \sin(4\pi f) + 2\pi^2 f^2) \quad (12)$$

By using (10), (11) and (12) in (6) we get the analytical form for amplitude characteristic of Keys 2P kernel which depends on the  $\alpha$  and  $\beta$  parameters. The choice of optimum parameters  $\alpha_{opt}$  and  $\beta_{opt}$  is described in the following section.

### III. OPTIMIZATION OF KERNEL PARAMETERS

In [4] the optimization of parameter of 1P Keys kernel was performed in the time domain. The algorithm for determining the optimal parameter by minimizing the interpolation error is presented. [5] shows the optimization of 1P Keys kernel in the spectral domain. The optimum parameters of Keys 2P kernel are determined in [8] based on the criterion of the decrease of wiggles of the amplitude characteristic in the low-pass range.

#### A. Total mean square error

This paper presents the optimization of the 2P Keys kernel parameters in the spectral domain. The main idea is that the spectral characteristic of the kernel should be a good approximation of the ideal rectangular function  $H_B(f)$  which has the value 1 in the interval [0-0.5] and value 0 in the interval [0.5-1], i.e. characteristics of the kernel in the form  $\sin x/x$ . Total mean square error is:

$$E_T = 2 \left( \int_0^{0.5} |1 - H(f)|^2 df + \int_{0.5}^1 |H(f)|^2 df \right), \quad (13)$$

After dividing the segment [0-1] into M subsegments, the discrete form of the total mean square error (MSET) is:

$$MSE = \frac{1}{M} \sum_{k=0}^{M-1} |1 - H(f_k)|^2 + \frac{1}{M} \sum_{k=\frac{M-1}{2}+1}^M |0 - H(f_k)|^2. \quad (14)$$

In consideration to (7) the mean square error depends on the parameters  $\alpha$  and  $\beta$ . Optimum value of the parameters was obtained based on the position of the minimum value MSE ( $\alpha$ ,  $\beta$ ).

#### B. The algorithm for the estimation of parameters

The algorithm for the estimation of optimum parameters  $\alpha_{opt}$  and  $\beta_{opt}$  of Keys 2P kernel in the spectral domain is shown in the fig.1. The input parameters are the spectral components  $H_0$ ,  $H_1$  and  $H_2$  (equations 10, 11 and 12), the range of  $\alpha$  parameter ( $\alpha_{min}$ ,  $\alpha_{max}$ ), iterative step  $\Delta\alpha$ , the range of  $\beta$  parameter ( $\beta_{min}$ ,  $\beta_{max}$ ), iterative step  $\Delta\beta$ , the range of normalized frequency  $f$  ( $f_{min}$ ,  $f_{max}$ ), iterative step  $\Delta f$ . The output values are  $\alpha_{opt}$ ,  $\beta_{opt}$  and the minimum value MSET<sub>min</sub>.

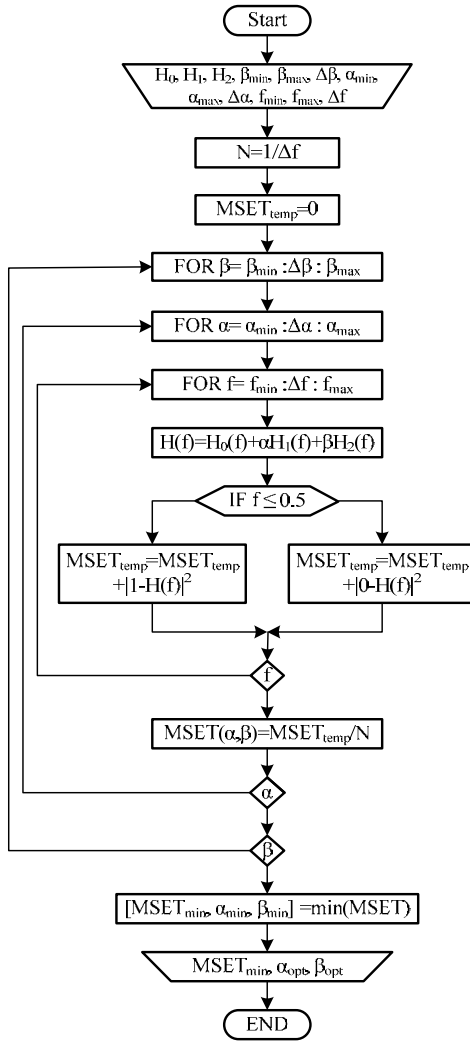


Fig. 1. Algorithm for the estimation of optimum parameters.

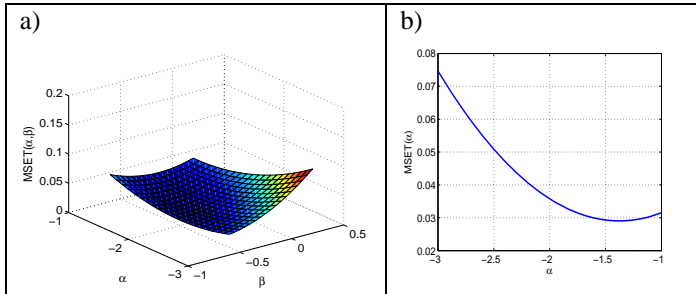


Fig. 2. a) Total mean square error between the amplitude characteristic of 2P Keys kernel, H, and the ideal Hb spectral characteristic depending on parameters  $\alpha$ ,  $\beta$  b). Total mean square error between amplitude characteristic 1P Keys kernel, H, and the ideal Hb spectral characteristic depending on  $\alpha$  parameters.

TABLE II  
PARAMETERS OF THE FUNCTION OF GAUSSIAN DISTRIBUTION

$f_s$ [Hz]	$\alpha_{opt}$		$\beta_{opt}$	
	$\sigma^2$	$\mu$	$\sigma^2$	$\mu$
44100	0.0120	-0.6929	$5.38 \cdot 10^{-4}$	0.0146
22050	0.0165	-0.7214	0.0012	0.0255
8000	0.1662	-0.6571	0.5330	0.5144

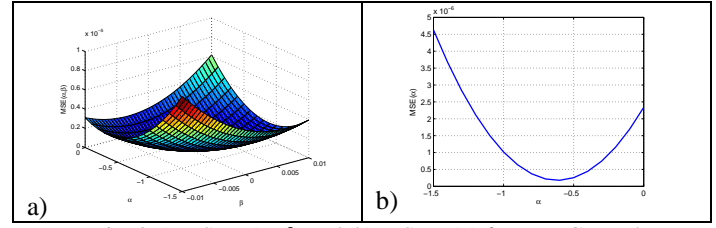


Fig. 3. a) MSET ( $\alpha$ ,  $\beta$ ) and b) MSET ( $\alpha$ ) for tone  $G_3$  on the following sample frequency  $f_s=44.1$ kHz

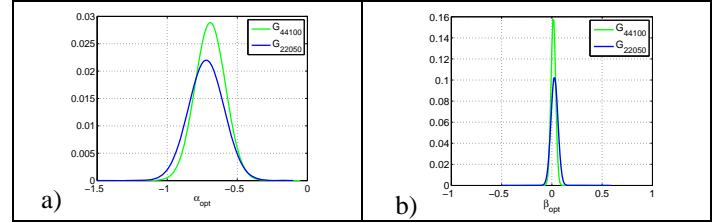


Fig. 4. Function of Gaussian distribution of optimum values a) of  $\alpha$  parameters b) of  $\beta$  parameters for sample frequencies  $f_s=44.1$ kHz and  $f_s=22.05$ kHz.

#### IV. EXPERIMENTAL RESULTS AND ANALYSIS

##### A. Experiment

The optimum values of the parameters  $\alpha$  and  $\beta$  in the case of interpolation of the audio signal were obtained experimentally. The mean square error was determined in the process of interpolation  $i+1$  sample based on  $\{i-2, i, i+2, i+4\}$ . This process was repeated for the whole signal starting from  $i=2$  to  $i=M-2$  where  $M$  is the length of the array. A statistics analysis of the results was performed by using the Gaussian distribution.

##### B. The basis

The basis is made of audio signals obtained by recording of G tones ( $G_1-G_7$ ) performed on the August Forster piano using the sample frequencies  $f_s=\{44.1, 22.05, 8\}$ kHz.

##### C. Results

The experiments obtained optimum values of parameters  $\alpha$  and  $\beta$  Keys 2P kernel by the use of algorithm presented in figure 1. for the input parameters ( $\alpha_{min}=-3, \alpha_{max}=-1, \Delta\alpha=0.1, \beta_{min}=-0.6, \beta_{max}=0.2, \Delta\beta=0.05$  and  $f_{min}=0, f_{max}=0.2, \Delta f=0.01$ ). By the application of fig.1. for the aforementioned input parameters optimum values of the parameter  $\alpha_{opt}=-1.9$  and  $\beta_{opt}=-0.4$  were obtained and  $MSET=0.0198$ . Fig.2 a) shows the mean square error of 2P Keys kernel depending on the parameters  $\alpha$  and  $\beta$ , b) shows the mean square error of 1P Keys kernel depending on the parameters  $\alpha$  (result from [8]). Minimum values of MSE in the interpolation of the tones  $G_1-G_7$  for sample frequencies  $f_s=\{44.1, 22.05, 8\}$ kHz are presented in the table 1. Fig. 3 a) shows the total mean square error (MST) for tone  $G_3$  with the application of 2P Keys kernel, and b) shows the MST for tone  $G_3$  with the application



of 1P Keys kernel. The parameters of the function of Gaussian distribution a)  $\sigma^2$  (variance) and b)  $\mu$  (mean value) were obtained based on the data from table 1 and shown in table 2. The functions of Gaussian distribution of optimum values of the parameters  $\alpha$  and  $\beta$  for sample frequencies  $f_s=44.1\text{kHz}$  and  $f_s=22.05\text{kHz}$  are shown in fig. 4. A comparative analysis was performed with the results of the application of 2P Keys kernel where the optimum parameters were  $\alpha_{opt}=-44/81$  and  $\beta_{opt}=7/81$  based on the criterion of the decrease of wiggles of amplitude characteristic [8].

TABLE I  
EXPERIMENTAL VALUES OF THE PARAMETERS

Tone	$f_s[\text{Hz}]$	$\alpha_{opt}$	$\beta_{opt}$	MSE
G <sub>1</sub>	44100	-0.5500	0	$2.2906 \cdot 10^{-8}$
	22050	-0.6500	0.0010	$8.1374 \cdot 10^{-7}$
	8000	-0.7000	0.0110	$3.8772 \cdot 10^{-4}$
G <sub>2</sub>	44100	-0.5500	0.0010	$7.8752 \cdot 10^{-8}$
	22050	-0.7000	0.0060	$5.0339 \cdot 10^{-6}$
	8000	-0.8000	0.0500	$9.3589 \cdot 10^{-5}$
G <sub>3</sub>	44100	-0.6000	0.0005	$1.7162 \cdot 10^{-7}$
	22050	-0.6000	0.0020	$1.3603 \cdot 10^{-5}$
	8000	-0.8000	0.0500	$4.6006 \cdot 10^{-5}$
G <sub>4</sub>	44100	-0.7000	0.0050	$8.5367 \cdot 10^{-7}$
	22050	-0.7000	0.0145	$7.8177 \cdot 10^{-5}$
	8000	-1.0000	0.1500	$1.6737 \cdot 10^{-4}$
G <sub>5</sub>	44100	-0.8000	0.0075	$7.3630 \cdot 10^{-7}$
	22050	-0.7000	0.0170	$3.9116 \cdot 10^{-5}$
	8000	-1.1000	0.4000	$2.8594 \cdot 10^{-5}$
G <sub>6</sub>	44100	-0.8000	0.0245	$4.0872 \cdot 10^{-6}$
	22050	-0.7000	0.0380	$1.4387 \cdot 10^{-4}$
	8000	-0.2000	0.9500	$4.3180 \cdot 10^{-4}$
G <sub>7</sub>	44100	-0.8000	0.0635	$1.9657 \cdot 10^{-6}$
	22050	-1	0.1000	$8.3373 \cdot 10^{-5}$
	8000	0	1.9900	0.0119

#### A. Result analysis

Based on the experimental results shown in table 1, tbl. 2 and the optimum values of the parameters  $\alpha_{opt}=-1.9$  and  $\beta_{opt}=-0.4$  obtained by the application of the algorithm and optimum parameters  $\alpha_{opt}=-44/81$  and  $\beta_{opt}=7/81$  from [8] it can be concluded that:

- in the case of  $f_s=44100\text{Hz}$  the range of the optimum values of the parameter are  $\alpha$ ,  $\alpha_{opt} \in [-0.8 \div -0.55]$  and the mean value is  $\overline{\alpha_{opt}} = -0.6929$ , while  $\beta_{opt} \in [0 \div 0.0635]$  and  $\overline{\beta_{opt}} = 0.0146$ ;
- $\alpha_{opt} \in [-1.0 \div -0.6]$ , mean value  $\overline{\alpha_{opt}} = -0.7214$  and  $\overline{\beta_{opt}} = 0.0255$  for  $\beta_{opt} [0.001 \div 0.1]$  for  $f_s=22050\text{Hz}$ .
- in case when  $f_s=8000\text{Hz}$ ,  $\alpha_{opt} \in [-1.1 \div 0]$  and  $\beta_{opt} \in [0.01 \div 1.99]$ , while are  $\overline{\alpha_{opt}} = -0.6571$  and  $\overline{\beta_{opt}} = 0.5144$ .
- the error of the estimation of the  $\alpha$  parameter is smallest at sample frequency 22050Hz

$$\Delta_\alpha = |\alpha_{opt} - \overline{\alpha_{opt}}| = |-1.9 - (-0.7214)| = 1.1786,$$

$$\text{while for } \alpha_{opt}=-44/81 \cdot \Delta_{\alpha T} = |-0.5432 - (-0.7214)| = 0.1782$$

e) the error of the estimation of the  $\beta$  parameter is smallest at sample frequency 22050Hz

$$\Delta_\beta = |\beta_{opt} - \overline{\beta_{opt}}| = |-0.4 - 0.0255| = 0.4255,$$

$$\text{for the } \beta_{opt}=7/81 \text{ this error is } \Delta_{\beta T} = |0.08642 - 0.0255| = 0.0609$$

## V. CONCLUSION

This paper shows an algorithm for the optimization of the parameters of Keys 2P kernel in the spectral domain. The optimization of the parameters was performed so that the amplitude characteristic of the kernel is a good approximation to the ideal amplitude characteristic. By applying the algorithm suggested in this paper the following was obtained  $\alpha_{opt}=-1.9$  and  $\beta_{opt}=-0.4$ . An analysis of the interpolation of an audio signal (tones G<sub>1</sub>-G<sub>7</sub> piano August Forster) was performed and optimum values of the parameters for each signal for the sample frequencies  $f_s=\{44.1, 22.05, 8\}\text{kHz}$  were obtained. Optimum values of the parameters  $\alpha$  and  $\beta$  are in the range  $\alpha_{opt} \in [-1.0 \div -0.6]$  and  $\beta_{opt} \in [0 \div 1.99]$ . The smallest error in the estimation of the parameters is for sample frequencies  $f_s=22.05\text{kHz}$ . Considering the results from [8], by comparing the errors of parameter estimation it is concluded that the kernel with the parameters obtained by the method of decreasing the wiggles of the amplitude characteristic is more precise.

## REFERENCES

- [1] E. Meijering, M. Unser, "A Note on Cubic Convolution Interpolation", IEEE Transactions on Image Processing, vol. 12, no. 4, pp. 447-479, April 2003.
- [2] H.S. Pang, S.J. Baek, K.M. Sung, "Improved Fundamental Frequency Estimation Using Parametric Cubic Convolution", IEICE Trans. Fundamentals, vol. E83-A, no. 12, pp. 2747-2750, Dec. 2000.
- [3] S. E. Reicherbach, "Two-Dimensional Cubic Convolution ", IEEE Trans. Image Processing, vol. 12, no. 8, pp. 857-865, Aug. 2003.
- [4] R. G. Keys, "Cubic convolution interpolation for digital image processing", IEEE Trans. Acoust. Speech, & Signal Processing, vol. ASSP-29, pp. 1153-1160, Dec. 1981.
- [5] K. S. Park, R. A. Schowengerdt, "Image reconstruction by parametric cubic convolution", Computer Vision, Graphics & Image Process., vol. 23, pp. 258-272, 1983
- [6] E. Meijering, K. Zuiderveld, M. Viergever, "Image Reconstruction by Convolution with Simetrical Piecewise nth-Order Polynomial Kernels", IEEE Transactions on Image Processing, vol. 8, no. 2, pp. 192-201, Feb. 1999..
- [7] . R. Hanssen, R. Bamler, "Evaluation of Interpolation Kernels for SAR Interferometry", IEEE Transactions on Geoscience and Remote Sensing, vol. 37, no.1, pp. 318-321, Jan. 1999
- [8] Z. Miliivojević, N. Savić, D. Brodić, P. Rajković, "Optimizacija parametara Kejsvog dvoparametarskog konvolucionog jezgra u spektralnom domenu", XV Internacional Scientific-Professional Symposium INFOTEH-Jahorina 2016.

# Digital Image Processing



# Kernel based Extreme Learning Machines for Image Classification

Stevica S. Cvetković<sup>1</sup>, Miloš B. Stojanović<sup>2</sup>, Saša V. Nikolić<sup>3</sup>, Goran Z. Stančić<sup>4</sup>

**Abstract** – This paper investigates possibilities for application of Kernel based Extreme Learning Machines (K-ELM) to the problem of multiclass image classification. It is combined with Local Binary Pattern (LBP) image descriptor, to reach highly accurate results. LBP is widely used global image descriptor characterized by compactness and robustness to illumination and resolution changes. Classification is done using recently introduced K-ELM method. Experimental evaluation on a standard benchmark dataset consisting of thousand images classified in ten categories, has shown high accuracy of results comparing to other benchmark models.

**Keywords** – Image classification, Neural networks, Kernel based extreme learning machines, Local binary patterns

## I. INTRODUCTION

Image classification based on visual content is a crucial problem in computer vision research. The goal of an image classification system is to assign a category label with the most similar visual content, to the given query image. Visual similarity between images is commonly measured using robust and compact image descriptors (features).

There is a large set of visual descriptors available in the literature [1, 13]. The choice of the descriptor essentially affects the overall performance of the classification system. Local Binary Pattern (LBP) is one of the most widely used descriptor due to robustness to resolution and lighting changes, low computational complexity, and compact representation [2, 3, 4, 7]. The second crucial part of the system is machine learning technique to be applied for classification of descriptors. Support Vector Machine (SVM) is the most widely used machine learning technique for image classification purpose [5, 6].

In this study we investigate application of Kernel based Extreme Learning Machines (K-ELM) [8, 9, 10, 11, 12] for image classification, as an alternative to the commonly used SVM technique. The training of the SVM is based on solving

quadratic programming problem, which is usually time consuming when the number of training examples is large. Beside that SVMs are originally proposed for binary classification, while for the multi-class classification one-against-all (OAA) or one-against-one (OAO) approaches must be used in SVM implementation. On the other hand, K-ELM shows much better generalization performances for multiclass classification cases, and has better scalability and much faster training speed, compared with SVM [9].

In the rest of the paper we first give an overview of K-ELM classification method for multi-class image classification. Then we describe the process of LBP descriptor extraction. Finally, experimental evaluation and conclusion are presented.

## II. KERNEL BASED EXTREME LEARNING MACHINES (K-ELM) FOR MULTICLASS CLASSIFICATION

Let us define  $N$  training examples as  $(\mathbf{x}_j, \mathbf{y}_j)$  where  $\mathbf{x}_j = [x_{j1}, x_{j2}, \dots, x_{jn}]^T \in \mathbf{R}^n$  denotes  $j$ -th training instance of dimension  $n$  and  $\mathbf{y}_j = [y_{j1}, y_{j2}, \dots, y_{jm}]^T \in \mathbf{R}^m$  represents  $j$ -th training label of dimension  $m$ , where  $m$  is the number of classes. LBP image descriptor, which will be described in the next section, will further be denoted as  $\mathbf{x}_j$ . As  $\mathbf{y}_j$ , we will denote  $m$  dimensional vector of binary class labels with value “1” denoting membership to the class. SLFN with activation function  $h(x)$  and  $L$  hidden neurons could be defined as:

$$\sum_{i=1}^L \beta_i h(\mathbf{w}_i \cdot \mathbf{x}_j + b_i) = \mathbf{f}_j, j = 1, \dots, N \quad (1)$$

where  $\mathbf{w}_i = [w_{i1}, w_{i2}, \dots, w_{in}]^T$  denotes the vector of weights which connects the  $i^{\text{th}}$  hidden neuron and all input neurons,  $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T$  is the weight vector which connects  $i^{\text{th}}$  hidden neuron and all output neurons, and  $b_i$  is the bias of the  $i^{\text{th}}$  hidden neuron. By ELM theory [8],  $\mathbf{w}_i$  and  $b_i$  can be assigned in advance randomly and independently, without a priori knowledge of the input data. The ELM network structure is presented in Figure 1.

SLFN in (1) should satisfy  $\sum_{i=1}^L \|\mathbf{f}_i - \mathbf{y}_i\| = 0$ , i.e., there exist  $\beta_i$ ,  $\mathbf{w}_i$  and  $b_i$  such that:

$$\sum_{i=1}^L \beta_i h(\mathbf{w}_i \cdot \mathbf{x}_j + b_i) = \mathbf{y}_j, j = 1, \dots, N \quad (2)$$

<sup>1</sup>Stevica S. Cvetković is with the University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, e-mail: stevica.cvetkovic@elfak.ni.ac.rs

<sup>2</sup>Miloš B. Stojanović is with the College of Applied Technical Sciences Niš, Aleksandra Medvedeva 20, Niš 18000, Serbia, e-mail: milos.stojanovic@vtsnis.edu.rs

<sup>3</sup>Saša V. Nikolić is with the University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, e-mail: sasa.nikolic@elfak.ni.ac.rs

<sup>4</sup>Goran Z. Stančić is with the University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, e-mail: goran.stancic@elfak.ni.ac.rs



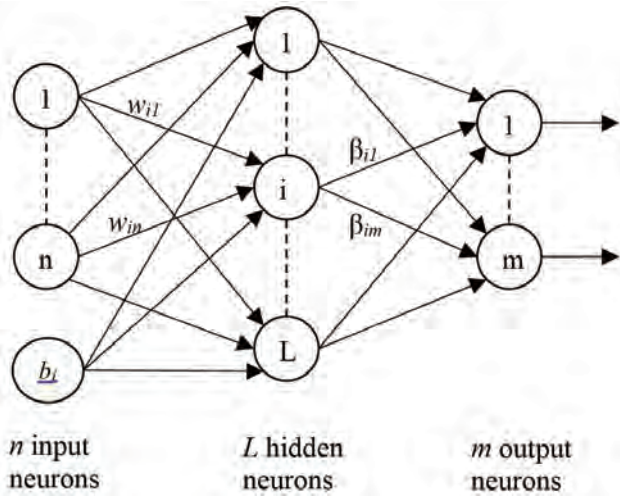


Fig. 1. Structure of an ELM network.

The equivalent compact matrix form of (2) can be written as

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{Y} \quad (3)$$

where  $\mathbf{H}$  in (3) represents the hidden layer output matrix of the neural network; the  $i^{\text{th}}$  column of  $\mathbf{H}$  represents the  $i^{\text{th}}$  hidden neuron's output vector in regard to inputs  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ .

$$\mathbf{H} = \begin{bmatrix} h(\mathbf{w}_1 \cdot \mathbf{x}_1 + b_1) & \cdots & h(\mathbf{w}_L \cdot \mathbf{x}_1 + b_L) \\ \vdots & \cdots & \vdots \\ h(\mathbf{w}_1 \cdot \mathbf{x}_N + b_1) & \cdots & h(\mathbf{w}_L \cdot \mathbf{x}_N + b_L) \end{bmatrix}_{N \times L} \quad (4)$$

and

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times m} \quad \text{and} \quad \mathbf{Y} = \begin{bmatrix} \mathbf{y}_1^T \\ \vdots \\ \mathbf{y}_N^T \end{bmatrix}_{N \times m} \quad (5)$$

Although the output weights can be analytically determined by finding the unique smallest norm least-squares solution of the linear system (3) in order to improve the performance the constrained optimization problem can be formed for ELM multiclass classifier with multiple outputs, as shown in [9]:

$$\text{Minimize : } L_p = \frac{1}{2} \|\boldsymbol{\beta}\|^2 + C \frac{1}{2} \sum_{i=1}^N \|\boldsymbol{\xi}_i\|^2 \quad (6)$$

$$\text{Subject to : } h(\mathbf{x}_j)\boldsymbol{\beta} = \mathbf{y}_j^T - \boldsymbol{\xi}_j^T, j = 1, \dots, N$$

where  $\boldsymbol{\xi}_j = [\xi_{j1}, \dots, \xi_{jm}]^T$  is the training vector of the  $m$  output nodes with respect to the training sample  $x_i$ , while  $C$  represents tradeoff parameter between model complexity and

allowed errors  $\xi_j$  during training. Based on *Karush - Kuhn - Tucker (KKT)* theorem, the optimization problem defined in (6) is equivalent of solving the dual optimization problem:

$$L_D = \frac{1}{2} \|\boldsymbol{\beta}\|^2 + C \frac{1}{2} \sum_{j=1}^N \|\boldsymbol{\xi}_j\|^2 - \sum_{j=1}^N \sum_{i=1}^m \alpha_{ji} (h(\mathbf{x}_j)\boldsymbol{\beta}_i - \mathbf{y}_{ji} + \xi_{ji}) \quad (7)$$

where  $\alpha_j = [\alpha_{j1}, \dots, \alpha_{jm}]^T$  are Lagrange multipliers.

After solving (7) based on *KKT* conditions, which can be found in detail in [9], the following solution is obtained:

$$\boldsymbol{\beta} = \left( \frac{\mathbf{I}}{C} + \mathbf{H}^T \mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{Y} \quad (8)$$

and the decision function of ELM classifier is:

$$f(\mathbf{x}) = h(\mathbf{x})\boldsymbol{\beta} = h(\mathbf{x})\mathbf{H}^T \left( \frac{\mathbf{I}}{C} + \mathbf{H}^T \mathbf{H} \right)^{-1} \mathbf{Y} \quad (9)$$

If feature mapping  $h(\mathbf{x})$  is unknown, we can apply Mercer's condition on ELM. We can define kernel matrix for ELM as:

$$\begin{aligned} \boldsymbol{\Omega}_{ELM} &= \mathbf{H}\mathbf{H}^T \\ \Omega_{ELM i,j} &= h(\mathbf{x}_i)h(\mathbf{x}_j) = K(\mathbf{x}_i, \mathbf{x}_j) \end{aligned} \quad (10)$$

In KELM  $\mathbf{H} = [h(\mathbf{x}_1)^T \dots h(\mathbf{x}_N)^T]^T$  represents hidden layer output matrix which maps data  $\mathbf{x}_i$  from the input space to the hidden layer feature space and it is irrelevant to target values  $y_i$  and number of output nodes  $m$ . The kernel matrix  $\boldsymbol{\Omega}_{ELM} = \mathbf{H}\mathbf{H}^T$  is related only to input data  $\mathbf{x}_i$  and number of training samples  $N$ , for regression, binary classification and multi class classification.

Then, the output function of ELM classifier (9) can be written compactly as:

$$f(\mathbf{x}) = \left( \begin{bmatrix} K(\mathbf{x}, \mathbf{x}_1) \\ \vdots \\ K(\mathbf{x}, \mathbf{x}_N) \end{bmatrix}^T \left( \frac{\mathbf{I}}{C} + \boldsymbol{\Omega}_{ELM} \right)^{-1} \mathbf{Y} \right) \quad (11)$$

In this case the feature mapping  $h(\mathbf{x})$  does not need to be defined by users, as well as the dimensionality of feature

space  $L$  (number of hidden nodes), just its kernel  $K(\mathbf{u}, \mathbf{v})$ . In our experiments RBF kernel is used, defined as:

$$K(\mathbf{u}, \mathbf{v}) = \exp(-\gamma \|\mathbf{u} - \mathbf{v}\|^2) \quad (12)$$

where  $\gamma$  represents parameter of Gaussian kernel. It can be noted from (11) and (12) that optimal combination of parameters  $C$  and  $\gamma$  have to be obtained in order to achieve good generalization performance.

### III. LOCAL BINARY PATTERNS (LBP)

Local Binary Pattern (LBP) is a popular image descriptor that captures local appearance around a pixel. LBP descriptor of the complete image is formed as a histogram of quantized LBP values computed for every pixel of the image. It was introduced in [4] for the texture classification problem, and extended to general neighborhood sizes and rotation invariance in [2]. Since then, LBP has been extended and applied to variety of applications [3].

For a given image  $I$ , the local LBP descriptor centered on pixel  $I(x, y)$  is an array of 8 bits, with one bit encoding each of the pixels in the  $3 \times 3$  neighborhood (Fig 2.). Each neighbor bit is set to 0 or 1, depending on whether the intensity of the corresponding pixel is greater than the intensity of the central pixel. To form the binary array, neighbors are scanned starting from the one to the right, at position  $I(x+1, y)$ , in anti-clockwise order.

a) Pixel intensities	b) Thresholded difference	c) LBP
167   221   221	0   1   1	<b>01100010</b>
147 <b>217</b> 198	0     0	
132   230   212	0   1   0	

Fig. 2. Example of a LBP extraction process for central pixel of intensity 217.

If  $3 \times 3$  neighborhood is used, there are 256 possible basic LBP codes. Using an extension from [2], this can be further reduced into a smaller number of patterns (58), which forms in a rotation-invariant descriptor. The extension is inspired by the fact that some binary patterns occur more frequently than others.

To describe the complete image, the quantized LBP patterns are grouped into histograms. The image could be divided into blocks, with a histogram computed for every block and concatenated to form the final descriptor. In our method we used only one image block, i.e. a global histogram is computed for the complete image.

To include image details at multiple scales, we extracted LBP histograms over the original image and several times resized image. Resizing is done to the half width and height of the original image using bicubic interpolation method.

Color image information is exploited by first converting an image into  $YCbCr$  color space and using all three color channels for LBP extraction. Final descriptor is formed by concatenation of the LBP histograms extracted at 3 scales (original + 2 downsampled) and 3 color channels. The computed image descriptor contains  $3 \times 3 \times 58 = 522$  dimensions.

### IV. EXPERIMENTAL EVALUATION

Test of the proposed method is performed using publicly available Corel1000 dataset [7]. The dataset consists of 1000 images classified into following 10 categories: *Africa people, Beach, Buildings, Buses, Dinosaurs, Elephants, Flowers, Horses, Mountains and Food*. An example image for every category is presented in Fig. 2. The dataset is characterized by large intra-category variations, which makes this dataset close to the real world image classification scenario.

We implemented the method in MATLAB and used it to measure the classification accuracy and time performance. To achieve correctness of results, tests were repeated for 50 times over random partitions of every category. We experimented with different number of training images per class. Classification accuracy results are presented in Table 1.

Table 1. Comparison of classification accuracy on Corel1000 dataset depending on the number of images per class. Standard deviations are given in brackets.

	% of training images per class		
	80%	50%	20%
K-ELM	90.86 ( $\pm 1.32$ )	89.58 ( $\pm 1.24$ )	85.49 ( $\pm 1.07$ )
SVM (kernel) [14]	90.79 ( $\pm 1.92$ )	88.26 ( $\pm 1.62$ )	84.04 ( $\pm 0.87$ )
SVM (linear) [14]	89.62 ( $\pm 1.27$ )	87.65 ( $\pm 1.03$ )	83.41 ( $\pm 1.25$ )

We further measured average training and testing time of the method on an *Intel Core i7 3.5GHz* computer. Training time of the complete training set was less than 1 second, while classification of a test image is done instantly ( $< 0.1\text{ms}$ ). These results demonstrate high performances in terms of training and test speed on the test dataset.

In order to compare results of the ELM with other common classification techniques, we measured accuracy of the Linear

SVM and kernelized RBF SVM [14], on the same dataset. Linear SVM parameter  $C$  was set to value 0.1. RBF SVM optimal parameters were determined by grid-search and 5-fold cross-validation, where  $C$  was examined in range  $[2^{-4}, \dots, 2^{10}]$ , and  $\gamma$  in range  $[2^{-10}, \dots, 2^4]$ . On the other hand, parameters of K-ELM were set to fixed values  $C=10$  and  $\gamma=2^{10}$ . It can be noted from experimental results that in terms of accuracy ELM constantly outperforms both Linear SVM and kernelized RBF SVM, without additional computational costs.

## V. CONCLUSION

In this study we presented results of our research in the field of automatic image classification using Kernel based ELM classifier (K-ELM) combined with LBP image descriptor. K-ELM classifier could be used as an effective alternative to the commonly used SVM methods. We reached classification accuracy of over 90%, on a test dataset containing 1000 images in 10 categories, what is high quality result on this dataset. It can be concluded that combination of K-ELM classifier with the LBP image descriptor is reasonable choice for image classification applications. In the further research, we plan to investigate integration of other image descriptors combined with K-ELM classifier. Particularly we will focus our research on fusion of color and texture descriptors.

## REFERENCES

- [1] Xin Zhang, Yee-Hong Yang, Zhiguang Han, Hui Wang, and Chao Gao, "Object class detection: A survey," *ACM Computing Surveys*, 46, 1, Article 10, July 2013.
- [2] T. Ojala, M. Pietikäinen and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with Local Binary Patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7), pp. 971-987, 2002.
- [3] M. Pietikäinen and G. Zhao, "Two decades of local binary patterns: A survey," In: E Bingham, S Kaski, J Laaksonen & J Lampinen (eds) *Advances in Independent Component Analysis and Learning Machines*, Elsevier, 2015.
- [4] T. Ojala, M. Pietikäinen, and D. Harwood, "A Comparative Study of Texture Measures with Classification Based on Feature Distributions", *Pattern Recognition*, vol. 29, pp. 51-59, 1996.
- [5] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories," *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 2169–2178, 2006.
- [6] C.-C. Chang and C.-J. Lin, "LIBSVM: A Library for Support Vector Machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 27:1–27:27, May 2011.
- [7] James Z. Wang, Jia Li, Gio Wiederhold, "SIMPLicity: Semantics-Sensitive Integrated Matching for Picture Libraries," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 9, pp. 947-963, 2001.
- [8] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, 2006.
- [9] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme Learning Machine for Regression and Multiclass Classification," *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics*, vol. 42, no. 2, pp. 513-529, 2012.
- [10] L. L. C. Kasun, H. Zhou, G.-B. Huang, and C. M. Vong, "Representational Learning with Extreme Learning Machine for Big Data," *IEEE Intelligent Systems*, vol. 28, no. 6, pp. 31-34, December 2013.
- [11] G. Huang, G.-B. Huang, S. Song, and K. You, "Trends in Extreme Learning Machines: A Review," *Neural Networks*, vol. 61, no. 1, pp. 32-48, 2015.
- [12] Stevica Cvetković, Miloš B. Stojanović, Saša V. Nikolić, "Multi-channel descriptors and ensemble of Extreme Learning Machines for classification of remote sensing images," *Signal Processing: Image Communication*, vol. 39, 2015, pp. 111-120.
- [13] R. Szeliski, *Computer Vision: Algorithms and Applications*, Springer, 2011.
- [14] C.-C. Chang and C.-J. Lin, "LIBSVM: A Library for Support Vector Machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 27:1–27:27, May 2011.

# New interactive interface for visualization of 3D meshes

Nicole Christoff<sup>1,2</sup>, Hadrien Greef<sup>3</sup>, Sébastien Germond-Mazet<sup>3</sup>, Jean-Luc Mari<sup>2</sup>,  
Laurent Jorda<sup>4</sup> and Agata Manolova<sup>1</sup>

**Abstract** – We propose a novel interface that visualizes a 3D mesh form the Polygon File Format (PLY), which represent a Martian terrain data. It juxtaposes the crater location, which coordinates of the center and diameter are specified in another input file. This program is designed to check and manually correct the results, obtained by the craters detection algorithm (CDA). This validation software is developed for use by astronomers. The focus is on the ergonomics of the interface and features.

**Keywords** – 3D mesh, interface, visualization, craters

## I. INTRODUCTION

The collision of celestial bodies, such as asteroids or meteorites, gives as result craters on their surface. The study of different morphological features of craters, and counting their numbers, is the only method to estimate the age of the celestial body. Therefore, it is a problem of great scientific importance in astronomy. The complexity of the problem arises when the impact crater zones are very heterogeneous due to the distribution and size of the craters.

To be able to solve partially this issue, different methodologies of geometric image analysis are proposed. They are applied to 2D or 3D data for automatic crater detection. Indeed, the manual identification of craters on the planet surface is an effective and precise method. The volume of data that needs to be processed grows with the development of new imaging technologies. That pushes scientists to look for a fully automatic method.

The combination of the two techniques (manual and automated) is a good solution of the problem. It is based on a rapid automatic pre-detection of craters that serves as a base to the human validation. We provide an ergonomic interface for visualization of 3D craters, which can correct the automatic detection.

The rest of the paper is organized as follows: in Section II, the existing approaches related to crater detection and interface are briefly described. In Section III, we describe the environmental data, then we develop the core of our interface in Section IV. The Section V is dedicated to the theoretical solutions for the visualization of 3D meshes. We present our results in Section VI. Conclusions and direction of future research are given in Section VII.

## II. RELATED WORKS

Many CDAs have been developed. They are based on 2D image data or 3D data, often in the form of Digital Elevation Model (DEM). The different approaches developed for 2D CDA are: the circle fitting [1], the exhibition of highlight and shadow region matching [2] or analysis textures [3]. Several approaches combine many of these methods [4].

The advantages of using detection from DEM data are the high resolution and the lack of hypothesis for the reflection of the surface, as well as the integration of additional 3D information. Previous works in this area are based on the detection of edges [5], identification of contours as a local circular depression [6] or the circle fitting Hough transform [7].

The creation of an ergonomic interface and 3D visualization of data, has been sparsely discussed in scientific works, often focusing on results and theoretical issues.

There are some works, which incorporate a graphical interface and they make the connection with CDA [8] and the existing functionality of Geographic Information System (GIS) software. For example, ArcGIS and QGIS and various external searching modules [9], [10], based on the same software, each time in 2D. As far as we know, there is no reference regarding the 3D visualization of meshes, combined with the displaying of detecting craters by CDA.

## III. ENVIRONMENT AND DATA

### A. Environment

The C++ programming language is used for the visualization of the interface and the data structures, used for meshes: Polygon File Format (PLY) and Comma Separated Values (CSV). Qt 4.8.6 in its version of 2013 is a cross-platform application framework that is widely used for developing application software [13].

The Visualization ToolKit (VTK) is used for the acquisition of a 3D visualization library for manipulating meshes and .PLY files. It is a tool, directly linked with Qt and supports advanced graphics features. VTK is optimized to manage heavy amounts of data and complex displays. Its graphic operating resources are developed to run our software on fairly modest configurations, even with large files.

### B. Data

The 3D mesh, in .PLY format, represents an area of Mars. It is considered for the detection, already triangulated from the

<sup>1</sup> Technical University of Sofia, Visual System Information Lab, Bulgaria.

<sup>2</sup> Aix-Marseille Université, CNRS, LSIS UMR 7296, France.

<sup>3</sup> Aix-Marseille Université, France.

<sup>4</sup> Aix-Marseille Université, CNRS, LAM UMR 7326, France.



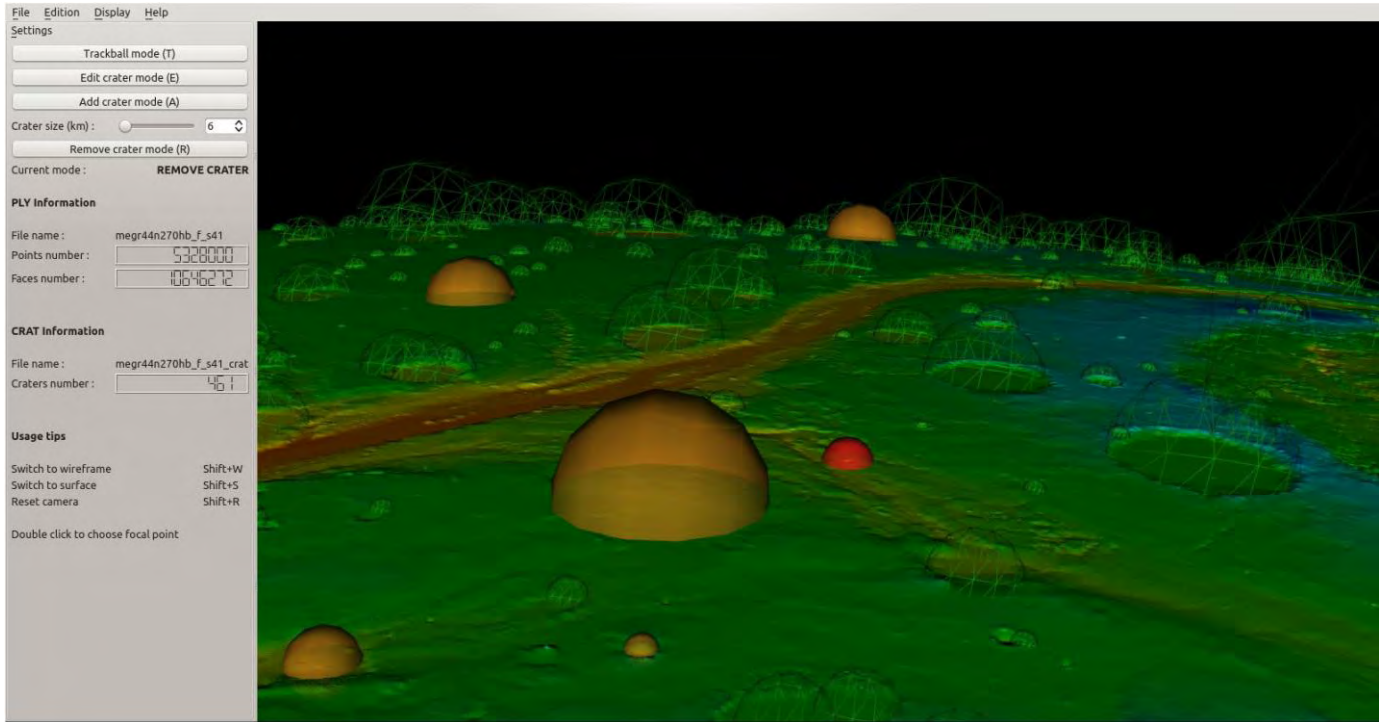


Fig. 1. The interface

DEM. In our case, it is a large data: over 5 million points (vertex) and 10 million faces. For comparison, the smaller ones are about 3 million points and 6 million faces. The area is a sample of 0.46 km / mesh. The input file is detected with a .CRAT file (CRATER extension, especially created for this work). This is a CSV file, with the following structure: coordinates of X, Y, Z of the center and the diameter of craters, detected by a CDA.

#### IV. DEVELOPMENT OF THE INTERFACE

##### A. Menus and windows

For the functional and ergonomic interface, we use the complex Qt classes - QMainWindow. The global window of the software contains a top menu, side toolbars, footer menu and the viewing window VTK. We made many inheriting of Qt classes to define our own widgets (Fig. 2). The main file generates a MainWindow, containing a CentralWindow, derived from VTKWidget class. The latter inherits the characteristics of QVTKWidget class. These heritages allow us to override certain functions of the original classes, while maintaining the interactions between all these objects. In the main window, we define the areas of the program: (1) a horizontal toolbar (Fig. 1. at the top of the window.), which contain the main interactions, opening files, saving, display options, etc. ; (2) a lateral toolbar (Fig. 3) irremovable for configuring the mouse modes of action (add a crater trackball mode, etc.) and display a basic information about uploaded files; (3) the central window, in which the mesh is displayed and where the user can click to move or edit 3D craters (Fig. 1 right).

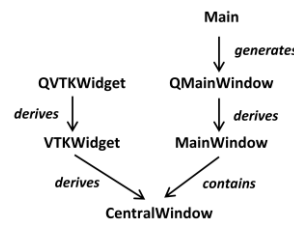


Fig. 2. Bloc diagram

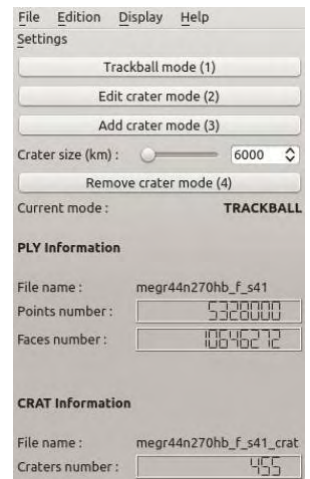


Fig. 3. The lateral toolbar

All functionality of the menu and the toolbar are connected to signals and slots (a language construct introduced in Qt for communication between objects [14]), to pass them on the program's global data (the list of craters) or to change the Interactor mouse (see Section IV-c). We manage pop-ups: (1) for searching of files that will be used and (2) to link these windows to the main application. The theoretical solutions of all the features are explained in Section V.

##### B. Displaying the grid

The solution had to be considered for a fluid display of the mesh, which can be moved, enlarged and covered, in all directions, by the user. To load the file, a decimation algorithm was



used to produce a lower level of detail. It is sufficient to continue to distinguish the biggest reliefs. In fact, it is destined to be displayed instead of the detailed model, every time the camera moves and zooms. In this way, to move the entire mesh remains a fluid process. Once the model is positioned, it is the detailed version that appears. This technique, used in GIS software [10] does not utilize many graphics resources and makes visualizing in real-time the mesh of millions of polygons possible. The decimation technique is presented in Section IV. The number of polygons, obtained in the decimation step, cannot be known in advance. In our case, it depends on the desired refreshing rate, produced during the movement and is thus likely to vary, depending on system configuration. The performances do not vary. Fig.4 illustrates both Level of Detail (LOD).

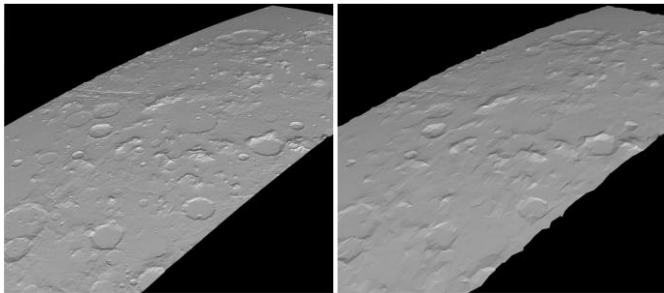


Fig. 4. Comparison between original 10 million polygons meshes (left) and decimated one of ~ 1000 polygons (right).

On Fig. 5 is presented a gray material, specular component of the dominant light. It is rather difficult to spot instinctively the potential craters. We implemented a staining method, which is called geographic. We have done a color mapping at each point, which depend on the value, ranges of distances from the center of Mars. More a point is far from the center of Mars (and therefore the high altitude), more the color should be cold (in a classic pattern red  $\rightarrow$  green  $\rightarrow$  blue). We get a very convenient local coloring, because of the abrupt altitude depressions, the result of a color change, as shown in Fig. 5.

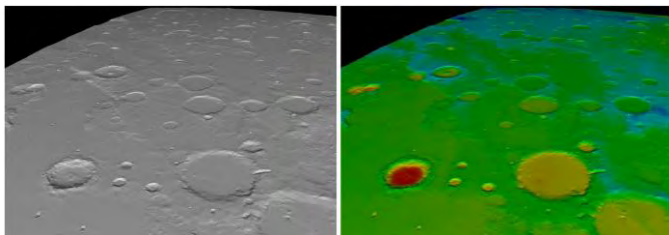


Fig. 5. Comparison between the original (left) and the geographic staining (right).

This staining was obtained by the computing the radius of the planet center, located at X, Y, Z (0,0,0) in the landmark planet. It is a color scale between the minimal and the maximal height and the correct color range is assigned to each point. There is a significant improvement for the identification of small craters (a few kilometers) on the mesh. If we are very close to the surface, it was a difficult case to distinguish craters without coloring. That becomes an additional tool to facilitate the manual detection.

### C. User interactions and the .CRAT file

We present an ergonomic operating interface, which seems to be intuitive to users, who are accustomed to software as QGIS [10]. We implemented four classes. Each of them corresponding to one of the four modes: (1) TRACKBALL uses to move the mesh to rotate and zoom in / out, and resumes the VTK usual trackball by adding features, such as changing the associated focal point with a double click. (2) ADD CRATER allows clicking on the mesh to include therein a crater. It is represented as a sphere. Holding down the button, the size of the crater can be adjusted and it starts from the original size adjustment, with the slider lateral toolbar. (3) EDIT CRATER can "raise" a crater to its correct position and adjust its diameter. (4) REMOVE CRATER allows removing a crater. If the user makes a mistake or if he wants to invalidate a false positive, it is proposed a drag function. If we lose the click outside the bounding sphere of the crater, it is not deleted. During these actions, the overall craters list is updated in real time and is recordable, at any time, in .CRAT format. This file format is used as input or output information data for the CSV files. We can load the detection file, make all kinds of changes and save the modified file. This makes much faster the detection of craters, retaining the human critical look at the results. Finally, we illustrate the different modes and which colors are associated with the craters, as shown in Fig. 1 (central window). With green mesh are covered craters, detected by the CDA, orange mesh is added by hand and a red mesh indicates a crater ready to be removed.

## V. THEORETICAL SOLUTIONS

### A. Decimation of surfaces

To streamline the interface, we discussed the idea of a hierarchy of levels of detail. It is called by the current performance of the center window. Thus, we use a quadratic mesh decimation that preserves the overall topology of the form by drastically reducing the number of polygons. It is based on the classification of points [12] according to their complexity: (1) by calculating the distance to the tangent to the surface plane and (2) by comparing the decimated surfaces and the not decimated, with a quadratic computing (improved decimation simple). It is in conjunction with the class, which manages the different LODs and allows setting the desired performance.

### B. Picking

Another important feature is the ability to click on a 3D point and to select the mesh, corresponding to the clicked place, or picking. The mouse coordinates are tied with the 2D (on the screen) to their 3D projection on the existing surface and recovers the point, which is the nearest place, considered as a mesh. The VTK has a picking manager, which works by ray tracing. Between the camera and the clicked point, we are launching a radius straight until the launching of the mesh or the bounding sphere of a crater. Then, we made a linear interpolation and we found a contact with the point, which recovers a "real" point. This method of picking is very effective and accurate. It allows

moving a crater, "sticky" to the surface below, rather than moving the camera in the plane or on an axis of the mark. It ensures that the action of moving of a crater remains faithful to other 3D craters.

### C. Fitting / Orientation

Finally, two problems have been solved: (1) to correct an offset between the mesh and the real crater and (2) to fit the mesh covering the crater to the size of the real crater. The offset is a known error, due to the improper manual submission by geologists [11]. We propose a correction of inaccurate detections craters (see Section IV-C). To guide our disc in the space, we used the least squares optimization method. For surface fitting, we have the coordinates of the center of Mars, as information on the craters file. We calculate the direction of the Martian axe, passing through its center. We use it as a normal to orient our mesh in successive rotations. This method is more elegant in its simplicity.

## VI. RESULTS

Once we have a .CRAT file, listing all the supposed detected craters by the CDA, the corresponding mesh is loaded into our visualization program. We can see the 3D detected craters. It is easy to identify the offset and to correct in some few clicks. In addition, we can add a forgotten crater.

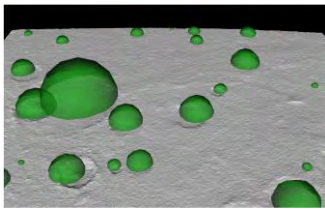


Fig. 6. Visualization of craters, which are detected by a CDA.

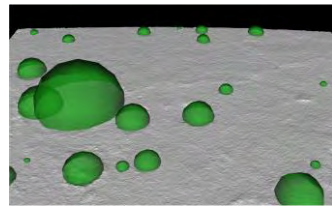


Fig. 7. Visualization of craters, which are detected after manual corrections.

It is difficult to represent exhaustively the results, because the detection rate is important. We can present it as a significant contribution to the detection algorithms. We illustrate the offset correction on the file, between the output of an algorithm CDA and its validation by the user on a small area of Mars. Corrections are applied in a few seconds (Fig. 6 and Fig. 7).

After the saving a .CRAT file, we obtain a theoretically perfect detection (if it is performed by an expert). The simplicity of the validation provides high-performance detection in a short time, by combining a starting rate and an effective manual correction.

## VII. CONCLUSION

We present an interface for validation, which allows users to browse intuitively 3D meshes. It induces a better perception of reliefs than 2D. It can move and adjust the craters. For this

work, we were focused on the development phase on the fundamental points of the program, namely the picking and user interactions. For future work, we think to improve the performance, with the management of more than two levels of detail.

## ACKNOWLEDGEMENT

This paper was supported by Contract No 162Π/Π0033-07 of Technical University-Sofia, Research Sector. Research project: «Segmentation and modeling of geometrical characteristics of 3D objects» - 2016.

## REFERENCES

- [1] L. Luo, X. Wang, W. Ji and C. Li, "Automated detection of lunar craters based on Chang'E-1 CCD data", The Fourth International Congress on Image and Signal Processing, vol. 2, Shanghai, 2011
- [2] E. R. Urbach, "Classification of objects consisting of multiple segments with application to crater detection", Proceedings of Eighth International Symposium on Mathematical Morphology, pp. 81–82, Rio de Janeiro, Brazil, 2007.
- [3] T. Barata, E.I. Alves, J. Saraiva and P. Pina, "Automatic recognition of impact craters on the surface of Mars", A. Campiolo, M. Kamel, (Eds.), Image Analysis and Recognition, Lecture Notes in Computer Science, vol. 3212. Springer, Berlin/Heidelberg, pp. 489–496, 2004
- [4] W. Ding, T.F. Stepinski, Y. Mu, "Subkilometer crater discovery with boosting and transfer learning", ACM Trans. Intell. Syst. Technol. (TIST) vol. 2, no.4, p.39, 2011
- [5] G. Salamuniccar and S. Loncaric, "Method for crater detection from Martian digital topography data using gradient value/orientation, morphometry, vote analysis, slip tuning, and calibration.", IEEE Trans. Geosci. Remote Sens. vol. 48, pp. 2317–2329, 2010
- [6] T.F. Stepinski, M.P. Mendenhall and B.D. Bue, "Machine cataloging of impact craters on Mars", Icarus vol. 203, no. 1, pp. 77–87, 2009
- [7] L. Luo, L. Mu, X. Wang, C. Li, W. Ji, J. Zhao and H. Cai, "Global detection of large lunar craters based on the CE-1 digital elevation model", Front. Earth Sci. vol. 7, no. 4, pp. 456–464, 2013
- [8] F. Baum and M. Zanetti, "Streamlined generalization tool for planetary surface mapping using ArcGIS ModelBuilder software on multispectral datasets", 46th Lunar and Planetary Science Conference, 2015
- [9] T. Kneissl, S. van Gasselt and G. Neukum, "New software tool for map-projection-independent crater size-frequency determination in ArcGIS", 41st Lunar and Planetary Science Conference, 2010
- [10] T. Ojala, M. Pietikinen and D. Harwood, "A comparative study of texture measures with classification based on feature distributions", Pattern Recognition, vol. 29, pp.51–59, 1996
- [11] W. J. Schroeder, J. A. Zarge and W. E. Lorensen, "Decimation of Triangle Meshes", Conference Proceedings of SIGGRAPH, pp. 65-70, 1992
- [12] W. J. Schroeder, J. A. Zarge and W. E. Lorensen, "Decimation of Triangle Meshes", Conference Proceedings of SIGGRAPH, pp. 65-70, 1992
- [13] [https://wiki.qt.io/Qt\\_for\\_Beginners](https://wiki.qt.io/Qt_for_Beginners)
- [14] [https://en.wikipedia.org/wiki/Signals\\_and\\_slots](https://en.wikipedia.org/wiki/Signals_and_slots)

# Comparison of Complex Hadamard Transform with other Unitary Transforms

Rumen P. Mironov

**Abstract** – A comparison of the developed complex Hadamard transform with the well-known unitary (orthogonal) transforms of Karhunen-Loeve, Fourier, Walsh-Hadamard and discrete cosine transform is presented. The comparison is made on the base of minimization of mean-squared error of reconstructed transform coefficients for two test images.

**Keywords** – Digital Image Processing, Complex Hadamard Transform, Orthogonal Transforms.

## I. INTRODUCTION

The use of unitary (orthogonal) transformations is inextricably linked with the development of modern digital image processing methods. Discrete unitary transforms described in [1], [2] and [3], have found applications in many areas of  $N$ -dimensional signal processing, spectral analysis, pattern recognition, digital coding, computational mathematics and etc. Stated simply, these transform coefficients that are small may be excluded from processing operations, such as filtering, compression and etc., without much loss in processing accuracy.

In this article a comparison of the developed complex Hadamard transform [4] with most used unitary transforms of Karhunen-Loeve (KLT), Fourier (DFT), Walsh-Hadamard (WHT) and discrete cosine transform (DCT) is presented. There are various studies and comparisons of the orthogonal transforms [5], [6], [7], in which they discussed their properties, advantages and disadvantages from a statistical point of view. In this article a comparison of transformations in terms of the developed method [8] for optimal performance of the coefficients at the block coding for two test images was made.

The comparison is made through simulation of the developed algorithms for five unitary transforms – FFT, DCT, WHT, KLT and CHT on Matlab environment for two test images “Lena” and “Fruits” and the results are given in the experimental part.

## II. MATHEMATICAL DESCRIPTION

The forward and the inverse 2D discrete unitary transform of sub-image  $g(x,y)$  of size  $N \times N$  can be expressed as the

following equations:

$$\begin{aligned} S(u, v) &= \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} g(x, y) \cdot r(x, y, u, v) \\ g(x, y) &= \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} S(u, v) \cdot t(x, y, u, v) \end{aligned} \quad (1)$$

In this equations  $g(x,y)$  is the input image with spatial variables  $(x,y)$ ,  $S(u,v)$  is forward transform with transform variables  $(u,v)$ ,  $r(x,y,u,v)$  and  $t(x,y,u,v)$  are called the forward and inverse transformation kernels, respectively. Because the inverse kernel  $t(x,y,u,v)$  in (1) depends only on the indices  $(x,y,u,v)$  and not on the values of  $g(x,y)$  and  $S(u,v)$ , it can be viewed as defining a set of basis functions or basis images [3].

This interpretation becomes clearer if the equation is modified in matrix form:

$$\mathbf{G}_{xy} = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} S(u, v) \mathbf{T}_{uv}, \quad (2)$$

where:  $\mathbf{G}_{xy}$  is  $N \times N$  matrix containing the pixels of  $g(x,y)$ , the matrices  $\mathbf{T}_{uv}$  are the basis images and  $S(u,v)$  are the spectral coefficients.

The forward transformation kernel is said to be separable if:

$$r(x, y, u, v) = r_1(x, u) \cdot r_2(y, v), \quad (3)$$

and the kernel is said to be symmetric if  $r_1(x, u)$  functionally equal to  $r_2(y, v)$ , so that:

$$r(x, y, u, v) = r_1(x, u) \cdot r_1(y, v). \quad (4)$$

Identical comments apply to the inverse kernel by replacing function  $r(x,y,u,v)$  with  $t(x,y,u,v)$  in the equations (1).

As a sample the 2D Fourier transform has the following forward and inverse kernels:

$$\begin{aligned} r(x, y, u, v) &= e^{-j2\pi(ux/M + vy/N)} \\ t(x, y, u, v) &= \frac{1}{M \cdot N} e^{j2\pi(ux/M + vy/N)}, \end{aligned} \quad (5)$$

where  $j = \sqrt{-1}$ , so these kernels are complex.

A computationally simpler Walsh-Hadamard transform is derived from the following functionally identical kernels [3]:

$$r(x, y, u, v) = t(x, y, u, v) = \frac{1}{n} (-1)^{\sum_{i=0}^{m-1} [b_i(x)p_i(u) + b_i(y)p_i(v)]}, \quad (6)$$

where  $n=2^m$ . The summation in the exponent is performed in modulo 2 arithmetic and  $b_i(k)$  is the  $i^{\text{th}}$  bit in the binary representation of  $z$ . The  $p_i(l)$  is the conversion of  $b_i(k)$  with code of Grey.

<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)

The most used discrete cosine transformation is obtained by following equal kernels:

$$r(x, y, u, v) = t(x, y, u, v) = a(u) \cdot a(v) \cdot \cos\left(\frac{(2x+1)u\pi}{2n}\right) \cos\left(\frac{(2y+1)v\pi}{2n}\right), \quad (7)$$

where:

$$a(u) = \begin{cases} \sqrt{\frac{1}{n}} & \text{for } u = 0, \\ \sqrt{\frac{2}{n}} & \text{for } u = 1, 2, \dots, n-1, \end{cases}$$

and similarly for  $a(v)$ .

The identical description for the developed complex Hadamard transformation kernels are made in [9] and are used for comparison.

The kernels of optimal Karhunen-Loeve transformation are calculated for each test image by calculation the correlation function and the preparation of the eigenvalues and eigenvectors as is given in [1]. To simplify the calculations the input learning vectors are taken from the current test image consecutively according to image linear scanning.

From the equations (1) to (4) the compared transformations can be generalized for two-dimensional signals (images) in the following way:

$$\begin{cases} [Y] = [T_N] [X] [T_N]^T \\ [X] = \frac{1}{N^2} [T_N]^T [Y] [T_N] \end{cases}, \quad (8)$$

where:  $[X]$  is matrix of the input image with size  $N \times N$ ,  $[T]$  is a matrix form of each transform kernel and the result is a spatial spectrum matrix  $[Y]$  with the same size.

The symmetry of  $[T]$  matrix coefficients allows 2D transforms to be accomplished in two steps. The first one is one-dimensional transform for every row of the image and the second one is one-dimensional transform for the columns. This difference of transformation makes easier the calculations and the symmetry guarantees that the correlations between image elements in horizontal and vertical direction will influence in the same way the determination of transformed elements. The same considerations can be made for two steps calculation of the inverse 2D transforms.

### III. Experimental Results

The comparison is made on Matlab environment for five 2D unitary transforms – Karhunen-Loeve, Discrete Fourier Transform, Discrete Cosine Transform, Discrete Walsh-Hadamard Transform and Complex Hadamard Transform. The obtained experimental results for the test grayscale images “Lenna” and “Fruits” shown on the Fig.1b and Fig.1a with size 512x512 pixels and 8 bits per pixel.

The simulations were made for each image by transformation with kernel with size 8x8. The transform

coefficients are reduced by the using of presented in [8] block truncation coding algorithm.

The obtained experimental results for the test images “Lenna” and “Fruits” (512x512, 8 bits) with sub-image kernel 8x8 are given on Table 1 and Table 2 respectively.

TABLE 1.

Reduction type	MSE	NMSE	SNR, dB	PSNR, dB
48 Reduced Coefficients, 16 saved				
Zero DFT	114.488	$1.47 \times 10^{-5}$	48.3180	27.5772
Mean DFT	103.124	$1.32 \times 10^{-5}$	48.7720	28.0312
Zero DCT	-0.3842	$-4.9 \times 10^{-8}$	73.0599	52.3191
Mean DCT	-0.2903	$-3.7 \times 10^{-8}$	74.2765	53.5357
Zero WHT	19.0409	$2.45 \times 10^{-6}$	56.1087	35.3679
Mean WHT	16.1269	$2.07 \times 10^{-6}$	56.8301	36.0893
Zero CHT	16.2735	$2.09 \times 10^{-6}$	56.7908	36.0500
Mean CHT	15.0354	$1.93 \times 10^{-6}$	57.1345	36.3936
Zero KLT	299.676	$3.86 \times 10^{-5}$	44.1391	23.3983
Mean KLT	249.875	$3.21 \times 10^{-5}$	44.9283	24.1875
55 Reduced Coefficients, 9 saved				
Zero DFT	126.300	$1.62 \times 10^{-5}$	47.8915	27.1507
Mean DFT	113.403	$1.45 \times 10^{-5}$	48.3593	27.6185
Zero DCT	-0.2014	$-2.6 \times 10^{-8}$	75.8649	55.1241
Mean DCT	-0.0054	$-7.0 \times 10^{-10}$	91.5437	70.8029
Zero WHT	48.4430	$6.23 \times 10^{-6}$	52.0533	31.3125
Mean WHT	45.8332	$5.89 \times 10^{-6}$	52.2938	31.5530
Zero CHT	46.2743	$5.95 \times 10^{-6}$	52.2522	31.5114
Mean CHT	42.5687	$5.48 \times 10^{-6}$	52.6147	31.8739
Zero KLT	313.557	$4.03 \times 10^{-5}$	43.9424	23.2016
Mean KLT	260.008	$3.34 \times 10^{-5}$	44.7557	24.0149

TABLE 2.

Reduction type	MSE	NMSE	SNR, dB	PSNR, dB
48 Reduced Coefficients, 16 saved				
Zero DFT	20.2979	$3.53 \times 10^{-6}$	54.5211	35.0903
Mean DFT	17.7615	$3.08 \times 10^{-6}$	55.1008	35.6699
Zero DCT	-2.0464	$-3.5 \times 10^{-7}$	64.4855	45.0547
Mean DCT	-2.0174	$-3.5 \times 10^{-7}$	64.5477	45.1168
Zero WHT	3.7753	$6.56 \times 10^{-7}$	61.8261	42.3952
Mean WHT	0.8382	$1.45 \times 10^{-7}$	68.3620	48.9312
Zero CHT	0.8032	$1.39 \times 10^{-7}$	68.5470	49.1162
Mean CHT	0.3326	$5.78 \times 10^{-8}$	72.3756	52.9448
Zero KLT	61.9106	$1.07 \times 10^{-5}$	49.6779	30.2472
Mean KLT	55.0704	$9.58 \times 10^{-6}$	50.18.64	30.7556
55 Reduced Coefficients, 9 saved				
Zero DFT	22.203	$3.86 \times 10^{-6}$	54.1313	34.7005
Mean DFT	20.814	$3.62 \times 10^{-6}$	54.4119	34.9811
Zero DCT	-3.8473	$-6.7 \times 10^{-7}$	61.7440	42.3132
Mean DCT	-4.0750	$-7.1 \times 10^{-7}$	61.4943	42.0635
Zero WHT	5.2066	$9.06 \times 10^{-7}$	60.4300	40.9992
Mean WHT	3.7104	$6.45 \times 10^{-7}$	61.9014	42.4705
Zero CHT	7.1197	$1.24 \times 10^{-6}$	59.0710	39.6402
Mean CHT	6.1880	$1.07 \times 10^{-6}$	59.6801	40.2493
Zero KLT	62.011	$1.07 \times 10^{-5}$	49.6709	30.2400
Mean KLT	54.837	$9.53 \times 10^{-6}$	50.2048	30.7740



In the first part of each table the results for 48 reduced coefficients (the upper-left block with size 4x4 are saved), approximated with zero and mean values are shown, and in the second part the same experiments for 55 reduced coefficients (the upper-left block with size 3x3 - saved) are shown. The calculated values for the mean-square error (MSE), normalized mean-square error (NMSE), signal to noise ratio (SNR) and peak signal to noise ratio (PSNR) are shown.



Fig.1a. Test image "LENNA" (512x512 pixels and 256 gray levels).



Fig.1b. Test image "FRUITS" (512x512 pixels and 256 gray levels).

The output images after inverse transform for 48 reduced coefficients are showed on Fig.2a and 2b respectively.

#### IV. CONCLUSION

The general principles of comparison of 2D unitary transforms by the using of block transform coding of their coefficients of high order are given. The basic properties of CHT are discussed in previous publications. The obtained simulation results are practically identical for the CHT and real HT and show that both can be used in similar applications.

The best results are obtained for optimal Karhunen-Loeve transform and for the most used in compression standards

discrete cosine transform. The results for the CHT are better then the DFT for small size of kernels.

The main advantages of the developed algorithm for CHT are:

- faster calculation compared to other transformations;
- similar results with integer valued HT;
- using the CHT instead most complicated Fourier transform and keep the possibilities for working with complex spectrum.

The developed Complex Hadamard Transform can be used in digital signal processing for spectral analysis, pattern recognition, digital watermarking, coding and transmission of one-dimensional and two-dimensional signals.

#### V. ACKNOWLEDGEMENT

The author thanks the National Fund for Scientific Research of the Bulgarian Ministry of Education and Science for the financial support by the contract I-02/1.

#### VI. REFERENCES

- [1] N. Ahmed, K. R. Rao. *Orthogonal Transforms for Digital Signal Processing*, Springer-Verlag Berlin, Heidelberg, 1975.
- [2] W. K. Pratt. *Digital Image Processing*, 4<sup>th</sup> Ed., John Wiley & Sons. Inc., Hoboken, New Jersey, 2007.
- [3] R. C. Gonzalez, R. E. Woods. *Digital Image Processing*, Third Ed., Pearson Education Inc., 2008.
- [4] R. Mironov, R. Kountchev. "Analysis of Complex Hadamard Transform Properties", *XLI International Scientific Conference on Information, Communication and Energy Systems and Technologies, ICEST 2006*, 26 June – 1st July, Sofia, Bulgaria, 2006, pp.173-176.
- [5] S. Rahardja, B. Falkowski. "Comparative Study of Discrete Orthogonal Transforms in Adaptive Signal Processing", *IEICE Trans. Fundamentals*, Vol. E82-A, No. 8, August 1999, pp.1386-1390.
- [6] L. A. Zalmaznion. "Fourier, Walsh, and Haar transforms and their application in control, communication and other fields", in Science. Moscow, U.S.S.R.: Science Publisher, 1989.
- [7] O. Hunt, R. Mukundan. "A Comparison of Discrete Orthogonal Basis Functions for Image Compression", *Proc. of Conference on Image and Vision Computing*, New Zealand (IVCNZ) 2004, pp 53-58.
- [8] R. Mironov, R. Kountchev. "Spectrum Optimization of Truncated Complex Hadamard Transform", *XLVI Intern. Scientific Conf. on Information, Communication and Energy Systems and Technologies (ICEST'11)*, Serbia, Niš, June 29-July 1, 2011, Proc. Vol.1, pp.23-26.
- [9] R. Mironov, R. Kountchev. "Analysis of Complex Hadamard Transform Properties", *XLI International Scientific Conference on Information, Communication and Energy Systems and Technologies, ICEST 2006*, 26 June – 1st July, Sofia, Bulgaria, 2006, pp.173-176.
- [10] B. Falkowski, S. Rahardja. "Complex Hadamard Transforms: Properties, Relations and Architecture", *IEICE Trans. Fundamentals*. Vol. E87-A, No.8, August 2004.



b) Output image “Fruits”

a) Output image “Lenna”

Fig.2. Output images after reduction with 48 coefficients for FFT, DCT, WHT, CHT and KLT

# Some Finger Vein Databases Evaluation

Ivo Draganov<sup>1</sup>, Darko Brodić<sup>2</sup>

**Abstract** –In this paper a comparative evaluation is presented of some popular finger vein databases freely available for scientific research. The main aspects of the evaluation concern the quality of the images relating to their contrast, sharpness, present noise and distortions. Also quantitatively estimated values among images of one and the same database are presented. They are based on various measures which give an initial idea of the distribution of the visual representation of the veins for further recognition. Useful directions are made for image enhancement based on pre-processing prior to segmentation preceding the identification stage.

**Keywords** –Finger Vein Recognition, Database, Contrast, Sharpness, Noise

## I. INTRODUCTION

Finger vein recognition becomes more popular in the recent years as additional modality available in biometric systems [1-7] that could be used separately or in combination with other modalities. Vein patterns captured in two-dimensional image representation by illuminating human finger either from above relying on transmissive principle or from below based on diffusive scattering and reflectance with near infrared light provide a number of advantages compared to other biometric features [2]. First of all they are considered unique for every human even between twins [3]. Since finger veins are located under the skin there is less probability for visual change over long period of time and in parallel to that the current outer condition of the finger is of no importance to the capturing process, e.g. whether it's dry, wet or contaminated [4]. The vein patterns are permanent with aging. Like other features they allow to check the liveness of the sample, i.e. it's an effective measure against forgery [6]. The authentication is eased by the contactless way of acquisition. It is also thought to be zero-valued as for the Failure to enrolled rate (FER) and along with its other benefits qualified as highly secure and reliable.

Since a set of promising algorithms for finger vein recognition were proposed in the last decade [1-4] and still there are currently introduced enhancements to them an unified approach is needed for estimation of their overall performance and applicability to the practice. Considered as appropriate for incorporating in electronic passports and additionally assuring robust personal identification together with facial and fingerprint recognition [7] especially in the variable conditions in developing countries [5] a suitable database of finger veins needs to be selected as a starting point

for the foundation of such unified testing framework. It should have a variety of abundant properties such as wide range of covered ages for the volunteers, equally spread from youngsters to elderlies with equal presence between men and women. Different nationalities preferably from all continents are desirable to participate. Also more captured fingers from both hands would assure better selection of the proper ones for application and better evaluation of the tested algorithms will be made although the index and middle fingers are thought to be the most promising followed by the ring finger and much less attention is directed to the thumb and little finger. A few captures should be made of all selected fingers for testing with positional variations and partly such for the illumination conditions aiming to resemble real-case scenarios. Proper indexing preferably in the form of a related database and/or following specific xml schema is another requirement for efficient access while testing. All these features along with the objective measures for image quality are going to be evaluated below in this paper.

In Section II description is given for each of 5 evaluated databases, in Section III – quantitative analysis of the quality of the images from them and some considerations based on subjective inspection, and then in Section IV a conclusion is made.

## II. EVALUATED DATABASES

SDUMLA-HMT database [8] is presented in 2010 by the Group of Machine Learning and Applications from Shandong University (SDUMLA) as a Homologous Multi-modal Traits (HMT) set. It consists of face images incorporating 7 view angles, finger vein images including 6 fingers per individual, iris images from a single sensor and fingerprint images obtained with the use of 5 different sensors. All this multimodal data is collected from 106 individuals. It is considered by its collectors as the first open database of this kind. The capturing device used for the finger vein images was developed by the Joint Lab for Intelligent Computing and Intelligent Systems of Wuhan University. Each participating individual in the collection process provided an image of his/her index, middle and ring fingers from both hands for totally 6 times. Thus 3816 images were gathered in bmp format with resolution of 320x240 pixels resulting in 0.85 GB database size.

The Hong Kong Polytechnic University Finger Image Database [9] has been gathered since April 2009 aiming large scale bundling of finger vein images publicly available for research. Along with the finger vein images finger surface texture images were also acquired from both male and female volunteers – a process which ended in March 2010 followed by database release in September 2010. A contactless capturing device was used producing 6264 images from 156 individuals in bmp format. Around 93 % of the subjects were below 30 years of age at the time of the collection. It was

<sup>1</sup>Ivo Draganov is with the Faculty of Telecommunications, Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: [ivodraganov@tu-sofia.bg](mailto:ivodraganov@tu-sofia.bg).

<sup>2</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](mailto:dbrodic@tf.bor.ac.rs)



done in 2 sessions separated from 2 to 6 months apart in each of which 6 images of the finger vein patterns of the index and middle fingers of the left hand and 6 finger texture images or 24 images totally per subject.

Idiap Research Institute in Martigny with the assistance of Haute Ecole Spécialisée de Suisse Occidentale in Sion, Switzerland, developed the VERA finger vein database [10]. It consists of 440 images from 110 subjects. An open finger vein sensor [11] was used for the purpose. Both index fingers were scanned of each subject in two sessions following in 5 minutes one after the other. Forty women and seventy men between 18 and 60 years old volunteered in the process. Images are in png format with resolution 665 x 250 pixels with an average size of 80 kB per file.

Another effort for free finger vein database for academic use was made by the Tsinghua University by putting together Finger Vein and Finger Dorsal Texture images (THU-FVFD) [12]. All images were taken from 610 subjects in two sessions separated by a period from a few dozens of seconds up to a week where four finger vein and four finger dorsal texture images were captured. In the publicly available version of the database only one of these four images of each type is offered since the difference with the rest three is considered negligible. The resolution is 720 x 576 pixels and consequently a Region of Interest (ROI) was introduced with a size of 200 x 100 pixels covering the veins' area for each picture.

Incorporating information for the finger veins along with finger geometry it becomes possible to verify unimodal or bimodal (fusion of vein data with outer shape of the fingers) biometrics systems as it is done in the Finger Vein Universiti Sains Malaysia (FV-USM) database [13]. A ROI was also introduced in it for the veins to be used in algorithms' testing and comparing. Totally of 123 volunteers participated in the data collection of whom 83 were males and 40 females, mainly staff and students from the Universiti Sains Malaysia. Their age ranges from 20 to 52 years. Four fingers were captured – left index, left middle, right index and right middle which led to obtaining of 492 classes. Six snapshots were made for each finger in one session and after more than two weeks period the capturing was repeated and thus 5904 images were gathered as a result, with 2952 images from each session. Resolution is 640 x 480 pixels with 256 shades of gray per pixel.

### III. DATABASE IMAGE PROPERTIES

In order to evaluate quantitatively the images from the databases described in Section II here the approach described in [14] is employed where the following parameters are found on average for each one:

$$\bar{I} = \frac{1}{P.M.N} \sum_{p=0}^{P-1} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} I(i, j), \quad (1)$$

where  $\bar{I}$  is the average intensity of the images;  $P$  – their number in the current database;  $M$  and  $N$  – the dimensions of the images from it in number of pixels. The intensity  $I(i, j)$  for

the current spatial position  $(i, j)$  where  $i = \overline{0, M-1}$  and  $j = \overline{0, N-1}$  is changing between 0 and 255, i.e. in 8 bpp scale representing single tone variability.

Secondly, the average Root-Mean-Square (RMS) contrast  $\bar{C}$  for a whole database is found based on the definition:

$$\bar{C} = \frac{1}{P} \sum_{p=0}^{P-1} \sqrt{\frac{1}{M.N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [I(i, j) - \bar{I}]^2}. \quad (2)$$

Lastly, the average image entropy  $\bar{E}$ , again for a whole database, is found using the following expression:

$$\bar{E} = \frac{1}{P} \sum_{p=0}^{P-1} \left[ - \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} h(I(i, j)) \log_2(h(I(i, j))) \right], \quad (3)$$

where  $h$  is the probability for particular intensity  $I$  to occur inside the image.

Sample images from the tested databases are given in Fig. 1.

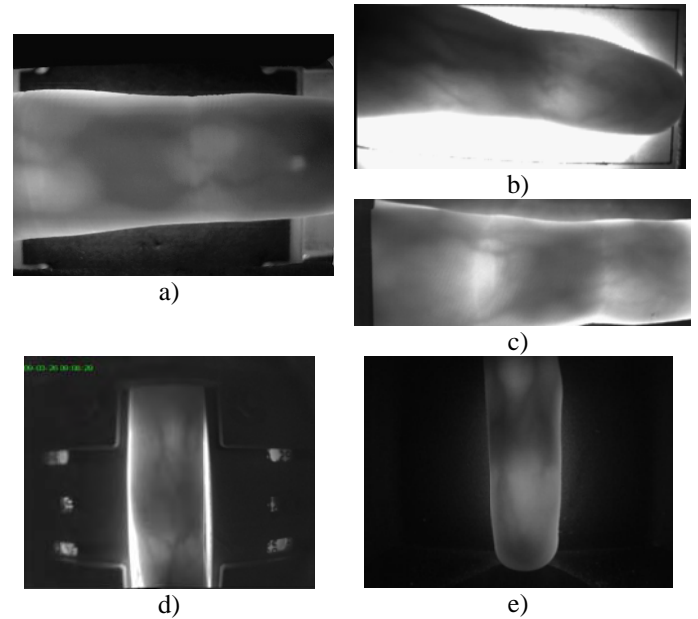


Fig. 1. Sample finger vein images from: a) SDUMLA-HMT, b) HKPUFID, c) VERA, d) THU-FVFD, e) FV-USM databases

The results from calculating (1)-(3) for all 5 databases described above are given in Table I. The average intensity for the SDUMLA-HMT and HKPUFID are the same as those estimated in [14] but the average contrast and entropy differ since the calculations are performed for the images as a whole here rather than using mosaicking windows.

First of all an impression makes the significant difference for the mean intensity between FV-USM from below as minimum and HKPUFID as maximum more than 7 times. The other three databases are scattered in between at more than 2 and 4 times higher than the darkest set. Obviously the



intensity of the LEDs used and the sensitivity of the receiver influence considerably the quality of the images in general. Darker images not necessarily mean worse quality. As the average contrast shows no more than just 3 levels is the difference with the highest value for the SDUMLA-HMT in comparison to FV-USM and less with the others. From Fig. 1.e it is visible that there are no surrounding flashes from diffuse scattered light around the finger contour for FV-USM and the contrast itself is decent as a subjective estimate. Not that is the case for the images from Fig. 1.b, c, and d – considerable side illuminations are present which possibly could affect the recognition process further without additional pre-processing. Parts of the support equipment are also visible in the Fig. 1.a, b and d which is desirable to be removed before further processing possibly at the stage of finger contour detection.

Similar is the case for the average image entropy (Table I) – there is less than 2 levels deviation among evaluated databases – from 5.48 to 7.37. Nevertheless of the presence of uneven background due to scattered light, sustaining elements or other factors the homogeneity of the image structure is affected mostly by the blood vessels pattern and the smooth intensity change introduced by finger bones and tissue. It is a positive factor which indicates the possibility for higher recognition rates when combined with proper normalization procedures.

TABLE I  
EVALUATED FINGER VEIN DATABASES PARAMETERS

Parameter	SDUMLA-HMT	HKPUFID	VERA	THU-FVFDT	FV-USM
Average Intensity	55.51	191.33	119.51	53.49	26.78
Average Contrast	10.72	10.06	9.61	8.09	7.67
Average Entropy	6.85	6.22	7.37	5.66	5.48

Detailed examination of the structure of the finger region inside tested images and the surrounding area pose additional considerations about the quality of veins' structure (Fig. 2). For the first two cases – excerpts from the SDUMLA-HMT (Fig. 2.a) and HKPUFID (Fig. 2.b) there are significant portions of the finger area that possess almost even distribution of intensity for both darker and lighter areas. The presence of these isles as they could be called lead to smaller local contrast which in turn makes the veins' pattern less distinctive over the tissue surroundings. This effect is considerably suppressed if not absent at all for the VERA, THU-FVFDT and FV-USM databases (Fig. 2.c, d and e). Although for the latter two the average entropy is smallest the intensity change around blood vessels has its repetitive nature underlying better their path. Various reasons may cause such differences among examined databases – from unsatisfactory set saturation levels (calibration) of sensitivity for the sensors accompanied by unwanted reflections from adjacent structures, through unsatisfactory initial image transformation, e.g. inappropriate intensity modification curve, or particular properties of the selected file format (mostly predetermined

by the coder at use) for permanent storage as well as other factors. All of them should be carefully examined at the stage of design and implementation of the experimental setup for gathering the database and only as exception post-processing after the final files had been written would be recommended for further enhancement of image quality.

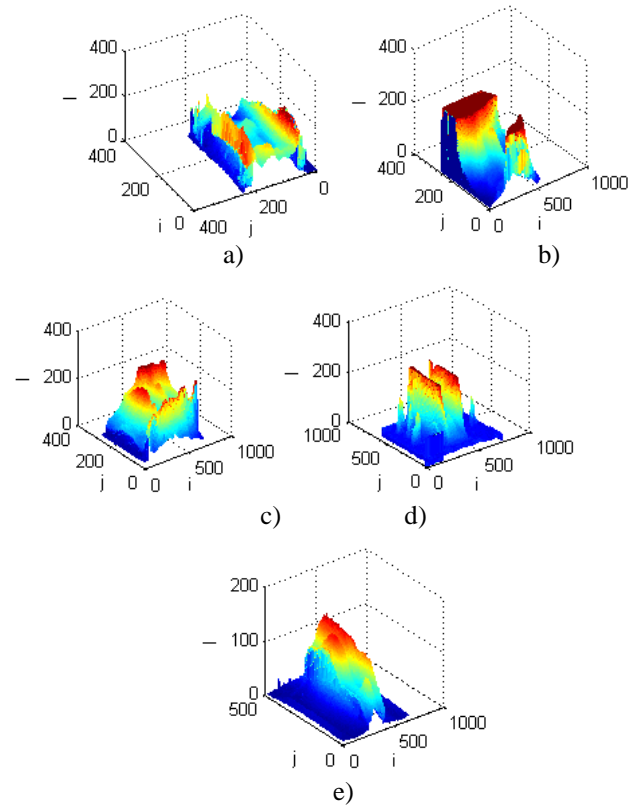


Fig. 2. Intensity distributions of sample finger vein images from: a) SDUMLA-HMT, b) HKPUFID, c) VERA, d) THU-FVFDT, e) FV-USM databases

Image compression especially block-based one such as the JPEG algorithm may additionally increase the effect of lowering the local contrast around veins. Since image dimensions are relatively small – of about 100x200 pixels or so for the Region of Interest (RoI) where the vessels are located and because of their repetitive structure with smaller number of intersections and small overall contrast it would be recommended to use some other compressor preserving more naturally continuous tones along separate directions such as the JPEG-2000 for example. Indexed-based formats, e.g. GIF, are also less perspective for such application. Ofcourse, the optimal solution disregarding content size is to store data in a raw stream or in a popular uncompressed format such as the BMP as in some of the databases described here.

Noises are other aspect of the quality assesment of the finger vein database images. They are realted with the magnitude fluctuation of the near-infrared light flow used for capturing the image and the level of thermogenerated currents inside the sensor. Due to the absence of raw data obtained with another type of illuminator and detector it is impossible to have quantitative estimation for their level and change over

time with modifications of the experimental setup. Visually a partial evaluation could be done, e.g. from Fig. 1 it is observed that with the lowering of the contrast (case d) and e)) the noise level is perceptibly higher (than that of the case for a) – c)). The selection of optical gathering system with appropriate filtering properties is another aspect to be considered when designing the complete finger vein sensor.

#### IV. CONCLUSION

In this paper evaluation is presented for 5 popular finger vein databases. It is based on objective comparison considering the average brightness, contrast and entropy of the images from each database as well as on visual inspection of the vein patterns captured, present noise, distortions and other factors concerning the overall perception of the content closely related to the recognition accuracy when identification stage takes place later. There is no particular parameter on which one can rely for simple selection of one database over the other for higher degree of confidence at testing newly developed recognition techniques. Moreover it is desirable to have images with implied imperfections such as higher noise level, distortions, diffused light, etc. in order to try unrepresentative capturing conditions which may be valid for some type of sensors from the practice. Useful directions for improving the capturing process are given in addition to the analysis of the factors leading to poor quality finger vein images which may be particularly effective for sensors design and their continuous exploitation.

#### ACKNOWLEDGEMENT

This work was supported by the National Scientific Fund at the Ministry of Education and Science, Republic of Bulgaria, within the project DFNI I02/1 "Intelligent man-machine interface for assistive medical systems in improving the independent living of motor disabled users".

#### REFERENCES

- [1] L. Yang, G. Yang, Y. Yin, L. Zhou, A survey of finger vein recognition, Chapter from Biometric Recognition, Z. Sun et al. (Eds.), Lecture Notes in Computer Science Series, Vol. 8833, Springer, Proceedings of the 9<sup>th</sup> Chinese Conference, CCBP 2014, Shenyang, China, pp. 234-243, November 7-9, 2014.
- [2] G. Gowsica1, L. Latha, A survey on biometric authentication techniques using finger vein, International Journal of Science and Research (IJSR), Vol. 3, Issue 11, pp. 1227-1231, November 2014.
- [3] M. Jadhav, P. RavaleNerkar, Survey on finger vein biometric authentication system, International Journal of Computer Applications, No. 3, pp. 14-17, 2015.
- [4] S. Vinothkumar, R. Sanmugasundaram, D. Divya, S. PadmaSarath, A survey of finger vein extraction techniques and its applications, International Journal of Engineering Research & Technology (IJERT), Vol. 3, Issue 2, pp. 2160-2165, February 2014.
- [5] Use of finger vein authentication for population-based surveys in developing countries, Report, Hitachi Review, Vol. 62, No. 8, pp. 456-462, 2013.
- [6] R. Raghavendra, M. Avinash, S. Marcel, C. Busch, Finger vein liveness detection using motion magnification, in Proceedings of the IEEE 7<sup>th</sup> International Conference on Biometrics Theory, Applications and Systems (BTAS), pp. 1-7, 2015.
- [7] S. Sree, N. Radha, A survey on fusion techniques for multimodal biometric identification, International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCCE), Vol. 2, Issue 12, pp. 7493-7497, December 2014.
- [8] Y. Yin, L. Liu, and X. Sun. SDUMLA-HMT: A multimodal biometric database. The 6th Chinese Conference on Biometric Recognition (CCBR 2011), LNCS 7098, pp. 260-268, Beijing, China, 2011.
- [9] A. Kumar and Y. Zhou, Human identification using finger images, IEEE Trans. Image Processing, vol. 21, pp. 2228-2244, April 2012.
- [10] P. Tome, M. Vanoni, and S. Marcel, On the vulnerability of finger vein recognition to spoofing, IEEE International Conference of the Biometrics Special Interest Group (BIOSIG), Darmstadt, Germany, Sept. 10-12, pp. 1-10, 2014.
- [11] B. Ton, Vascular pattern of the finger: biometric of the future? Sensor design, data collection and performance verification, Master's thesis, University of Twente, July 2012.
- [12] W. Yang, X. Yu, Q. Liao, Personal authentication using finger vein pattern and finger-dorsa texture fusion, Proceedings of the 17<sup>th</sup> ACM international conference on Multimedia, ACM, pp. 905-908, 2009.
- [13] M. Asaari, S. Suandi, B. Rosdi, Fusion of band limited phase only correlation and width centroid contour distance for finger based biometrics, Expert Systems with Applications, Vol. 41, Issue 7, pp. 3367-3382, ISSN 0957-4174, <http://dx.doi.org/10.1016/j.eswa.2013.11.033>, June 1<sup>st</sup>, 2014.
- [14] Y. Lu, S. J. Xie, S. Yoon, Zh. Wang, D. S. Park, An available database for the research of finger vein recognition, In Proc. of the 6<sup>th</sup> International Congress on Image and Signal Processing (CISP 2013), Hangzhou, China, pp. 410-415, Dec. 16-18, 2013.

# Tracking Dynamics Objects on Binocular Vision System

Rosen Spirov<sup>1</sup>, Neli Grancharova<sup>2</sup> and Svilen Soyanov<sup>3</sup>

**Abstract** – This paper presents the system to detect corresponding fragments of stereo pair and finally setting the parameters of dynamic objects, implemented in FPGA. A software-based algorithm was independently developed and examined in MATLAB to evaluate its performance and verify its effectiveness.

**Keywords** – Image Processing, Filters, FPGA, StereoVideo.

## I. INTRODUCTION

The goal of this idea was to create an FPGA system to detect and track dynamics objects in real time, as in human vision.

Today in the scientific literature there are numerous models depicting anatomy, operation and interpretation of visual information in the brain as shown in Figure.1 [1].

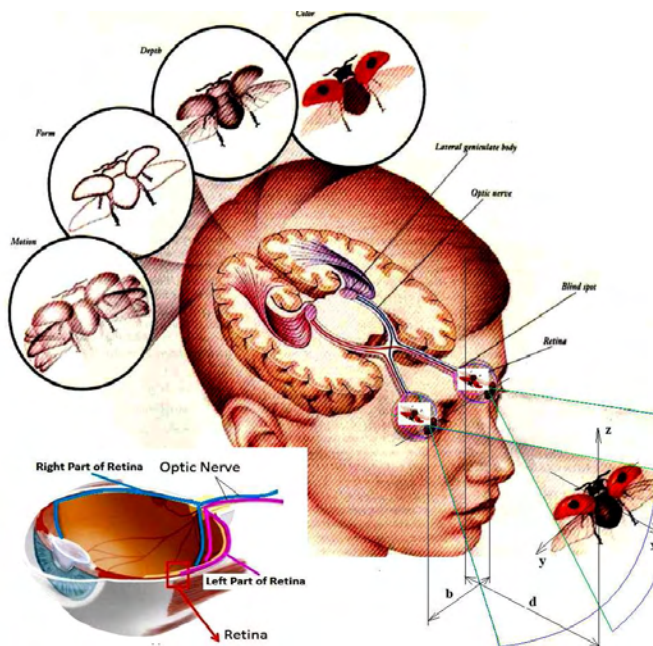


Fig.1. The visual information in the human brain

The medicine established laws of human perception, the characteristics of the optic tract, ways of forming mental images and their interpretation. Human perception is a function of a number of subconscious: processes for

determining distances, edges, shapes, sizes, colors, placements; adaptation to light, heat, radiation; determination of changes plasticity, interference, other indicators for the objects and the environment.

In the human eye, the axons of the ganglion cells are attached to the bundle of optic nerve. The entire overview that can capture the eye of man is individual visual space. For each eye is left and right visual space. The visual information from the retina reach the area of the brain called the LGN. This is represented in Figure 2 [1].

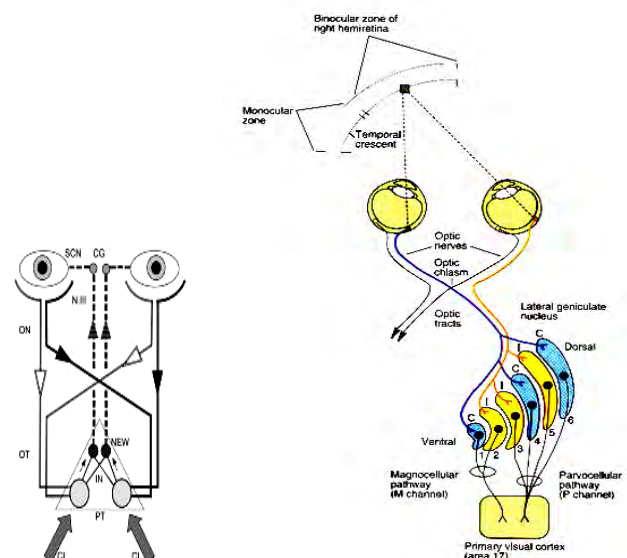


Fig.2. The eyes, visual tract and LGN area

Each of LGN consists of 6 levels- 2 inside levels that contain large cells -*magnocellular* and 4 outside levels with small cells - *parvocellular*. Between the layers there level - *interlaminar*, made up of other small cells - *koniocellular*. These layers receive input information from different types of ganglion cells. The process of entering information from each eye alternately flows to different types of layers. Visual tract continues to cortex lobe in the brain. The first cortex synapses of each neuron carries visual information in the middle part of the occipital lobe part of the brain so on...in the *Brodman-card*. This area is located medially and a knurled surface with the external folds of the cortex. *Cyto-architecture* is perfectly symmetrical and dots - formed an increase in volume by surface corrugations of the cortex. Similarly, it is built and optic cortex share, as is shown in the Figure 3[1].

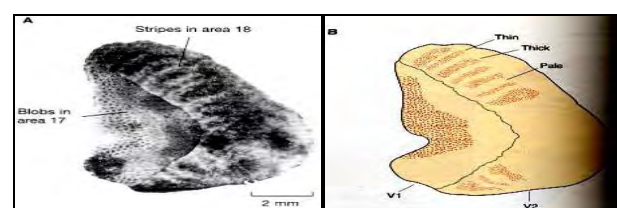


Fig. 3. The cortex share in the brain

<sup>1</sup>Rosen Spirov is with the Faculty of Electronics at Technical University of Varna, 1 Studentska Str, Varna 9010, Bulgaria, E-mail: rosexel@abv.bg.

<sup>2</sup>Neli Grancharova is with the Department of Termotechnics at Technical University of Varna, 1 Studentska Str, Varna 9010, Bulgaria, E-mail: nelly2000@abv.bg.

<sup>3</sup>Svilen Stoyanov is with the Technological College of Dobrich at Technical University of Varna, 1 Studentska Str, Varna 9010, Bulgaria, E-mail: svilenh@abv.bg.



This is consistent with the hypothesis, that there is carried out the pre-processing of visual space. Following anatomy work [1] on that partition can be represented with diagram:

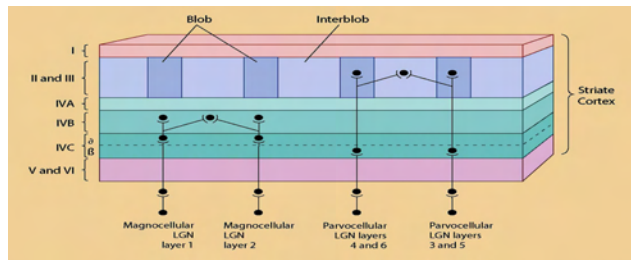


Fig. 4 . Operation of the cortex share

The separation of M and P ways are made in the cortex. Axons from both regions are terminated at layer 4 of the folds of the cortex, if (a) the terminal areas of axons in this layer have a branch each other- M - road, and (b) P - road drags second synapse information conveyed in the outer layers 2 and 3. The neurons equally selective ability are grouped in columns, according to the dominant eye and preferred orientation [1]. The orientation of the columns regularly presichat of blobs, as in figure 5.

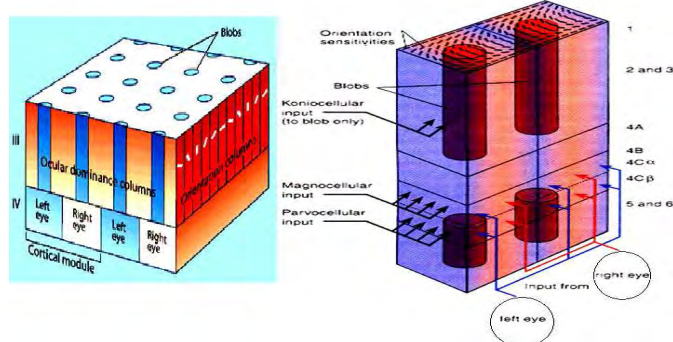


Fig.5. The processing in cortex share

Like the function of the human eye binocular video systems recreate a certain extent the human vision [3]. The own TV tracker detect and track dynamic objects and measure the parameters of the relative motion of the object set to TV system on Altera FPGA. They received images from both sensors differ in the location of objects on their screens and are determined by various observation points.

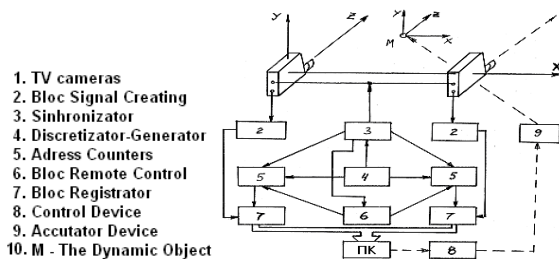


Fig. 6. TV trackers with automatically targeted

## II. THE DESIGN AND THE ALGORITHM

The mathematical model is based on an algorithm to determine the parameters of dynamic objects in series stereo

images [2], shown in figure 7. Based on the above algorithm is proposed structural diagram of the system for determining the parameters of dynamic objects, illustrated on figure 8.

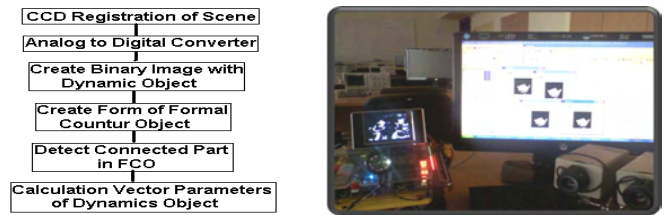


Fig.7. The algorithm and the binocular visual FPGA system

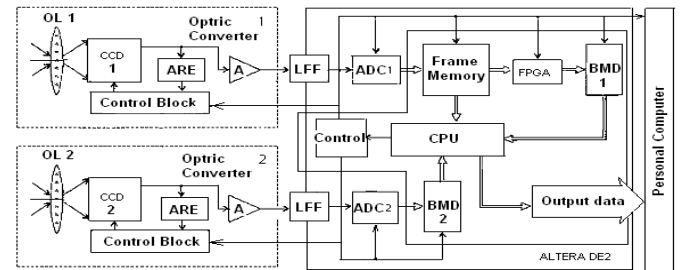


Fig. 8. The structural diagram

The description includes two optical converters (OC) with CCD [3]. The main blocks of OEП are: lens OL, CCD optical converter, control unit ARE and amplifier block (A). The output video signals of OC in digital form passed by low-pass filter (LFF), after that they enter the inputs of the 8- bits ADC. From output of ADC 2 digital signal is stored in a buffer memory device (BMD2) and from output of ADC1 digital signal is sent and saved in the block of frame memory. This block remembers N frames from the series images of the work scene. To achieve high-speed data processing is necessary to perform parallel calculations. The proposed implementation includes the structural organization of the block frame memory to design, as shown in Figure 9.

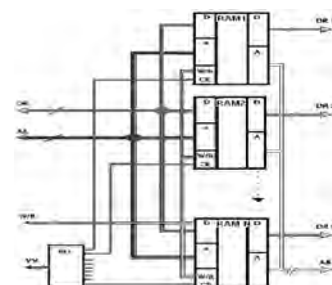


Fig. 9 The frame memory

Block of frame memory is a separate stack size N and type of organization FIFO, as the number of cells in the stack is equal to the number of frame. As a memory module used Dual-Port RAM (BMD). Both R and L ports allow independent applications to read and write to any cell array.

To achieve of parallel calculation is necessary every frame of image series to be stored in a separate of memory (BMD). The volumes of necessary memory blocks (BMD) using 8 bits quantization of size 640x480 pixels equivalence of 307.2 K bytes. Corresponding to input addresses AL, data DL, as well as the inputs W / R, the port L, are combined together.



Through port L recording is performed by the data outputs of the ADC and reading them from the microprocessor to determine related areas of the picture. The inputs to permit CE port L are connected to the outputs of the shift register, the outputs of which correspond to the number of used RAM (BMD). The management device (Control) controlled the work of register. Register carried move level logical unit to complete the recording of one frame of image. Thus, it is possible to work through the gate L only one module of RAM. Through a transferring registry is organized a stack FIFO. Through R port is carried out read data for forming the binary image, which contains a dynamic image areas  $S(k)$  by using FPGA. The corresponding input address AR, and also inputs W / R port R are merged together and entering into the control unit. Using Dual-Port RAM (BMD) realized bought of reading and record of information, without the need for waiting for the complete filling of the working RAM. This significantly shortens processing time. The device A formed signals of AL and AR, differing only by one bit, which in practice also simplifies the formation of the addresses. Thus, the output of the memory block of the frame memory is represented by the  $N-1$  - dimensional vector with values of pixels of the background image  $X_{\min} = [x_{\min}^1, x_{\min}^2, \dots, x_{\min}^{N-1}]$ , which are formed on the basis of the  $N-1$  past values  $\{x_{\min}^k, k = 0, 1, \dots\}$  for pixel in the process and current value of pixels in the image  $x_{\min}^k$ . For every component of the vector  $X_{mn}$  corresponding data bytes  $DR1-DR N-1$  and the current value of the pixel corresponds byte data  $DR N$  from worked memory (O3V). The task of checking the condition  $\min|X_{\min} - x_{\min}^k| > \sigma^{\text{доп}}$ , most often implemented using a FPGA [61]. In this case, for each component of the obtained vector with differences  $X_{\min} - x_{\min}^k$  it is necessary to check the condition  $X > C$ , where  $X$  and  $C$  is 8 bits binary numbers  $x7...x0$  and  $c7...c0$ , thus  $x0$  and  $c0$  is a low bits. From the output signal  $Q=1$  is removed by  $X > C$  or  $Q=0$  by  $X \leq C$ . The output data for coding presents in terms of Boolean algebra. The condition  $X > C$  for three orders can be set as follows:

$$Q = x_2 \bar{c}_2 + x_1 \bar{c}_1 (x_2 \oplus c_2) + x_0 \bar{c}_0 (x_2 \oplus c_2) (x_1 \oplus c_1) \quad (1)$$

The record (3.4) can be seen in the following form. The number  $X$  is larger than the number  $C$ , if one of three conditions:

- $x_2 \bar{c}_2 = 1$ , high bit of number  $X$  is equal to 1, and high bit of number  $C$  is equal to 0.
- $x_1 \bar{c}_1 (x_2 \oplus c_2) = 1$ , or the high bits is equals – its sum in modulo-2 of the inversion is equal to the 1, in that  $x_j = 1$ ,  $a c_j = 0$ .
- $x_0 \bar{c}_0 (x_2 \oplus c_2) (x_1 \oplus c_1) = 1$ , or the 2 high bits with codes  $X$  and  $C$  tally  $x_0 = 1$ ,  $c_0 = 0$ . The condition (3.4) extends to any number of bits  $n$ :

$$Q = x_{n-1} \bar{c}_{n-1} + x_{n-2} \bar{c}_{n-2} (x_{n-1} \oplus c_{n-1}) + x_{n-3} \bar{c}_{n-3} (x_{n-1} \oplus c_{n-1}) (x_{n-2} \oplus c_{n-2}) + \dots + x_0 \bar{c}_0 (x_{n-1} \oplus c_{n-1}) (x_{n-2} \oplus c_{n-2}) \dots (x_1 \oplus c_1) \quad (2)$$

From equation (3.5), it follows that when  $c_i = 1$  ( $i = 0, 1, 2, \dots, n-1$ ), then in which as a multiplier enters  $C_i$  becomes zero. Thus, when set with the right side of eq.(2) removes all multiplications corresponding to the units of bits with binary representation  $C0 \dots C7$ . In particular, when  $C = 1$  from the right side of eq. (2) removes all multiplications and variable  $Q$  becomes identically zero (condition  $X > C$  in this case is impossible because  $C$  is the maximum representable number). Conversely, however, when  $C = 0$  eq. (2) after simplification will take the form:

$$Q = x_{n-1} + x_{n-2} + x_{n-3} + \dots + x_0 \quad (3)$$

From eq. 3.6 it is following that  $Q=0$  only by  $x_{n-1} = x_{n-2} = x_{n-3} = \dots = x_0 = 0$ , but in another cases  $Q=1$ , or  $X > C=0$ . Thus, the number of the logic works in coding the FPGA is equal to the number of 0 in the binary number  $C = \delta^{\text{доп}}$

For parallel data processing in an FPGA must be implemented  $N-1$  structures described by the eq. (2), the outputs of which are connected to the inputs of the AND gate (AND). For this purpose simplify the task of checking the condition

$\min|X_{\min} - x_{\min}^k| > \sigma^{\text{доп}}$  and in the next step is reduced to task for matching the values of the components of the vectors  $X_{mn}$  to etalons  $X_{mn}^k$ , for reporting limits of disperse  $\delta^{\text{доп}}$  to compare be six senior diluted bytes of data  $DR$ . Such an organization is presented as follows:  
 $Q = (DR1_5 \oplus DR2_5 \oplus \dots \oplus DRN_5) \oplus (DR1_4 \oplus DR2_4 \oplus \dots \oplus DRN_4) \oplus \dots \oplus (DR1_0 \oplus DR2_0 \oplus \dots \oplus DRN_0)$  (4)

This equation is easily implemented with logic elements such as the functional organization of the simplified structure will look like as shown in figure 10.

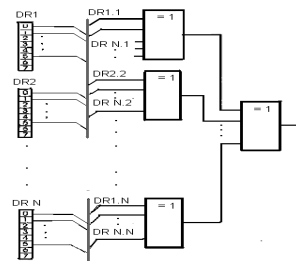


Fig. 9 The simplified structure

The FPGA handles data coming from the output of the block frame memory and stores them in buffer memory devices, the volume of which is 37.5 Kbyte. Microprocessor used to build FKO, make calculations by eq. (1) and eq. (2) of their parameters, such as processing the data. In addition to make calculations of matrix correction of Kalman filter [4]. This module is implemented with processors for embedded NIOS II on Altera FPGA. Operating time of the device  $T_y$  is defined as follows:

$$T_y = T_{\text{АЦП}} + T_n + T_{\text{FPGA}} + T_1 + T_2, \quad (5)$$

Where  $T_{\text{АЦП}}$  is the time of retention of the ADC;  $T_n$  - a delay of waiting at reading data from blocks of staffing memory is retention time processing with FPGA; is the time to create FCO in БЗУ1 including pre-filtration, filling gaps and

segmentation; is the time to detect their corresponding fragments of stereo pair and finally setting the parameters of dynamic objects. The time limit for the operation of the device in the presence of five sites in the frame constitutes 2.2ms, allowing processing of 25 frames per second.

### III. EXPERIMENTAL RESULTS

When a large number of observation accuracy of the estimate becomes sufficient to form, in practice, terminals for evaluation of the performance of compatibility, consistency and efficiency [5]. It should be noted that the width of the range of measurement of the brightness of the pixel determines the statistical difference between the minimum brightness levels of the background and a dynamic object set in the algorithm.

The reliability  $\gamma$  is assumed to be chosen greater than 0.95. Then the event, which range  $(-\hat{x}_{mn}^k, +\hat{x}_{mn}^k)$  roof parameter

$x_{mn}^k$ , will be faithful. After determining the:  $\Phi(x_\gamma) = \frac{\gamma}{2}$ ,

should find the corresponding tabular value tables for the function of Laplace. Thus, in  $\gamma = 0.99$ , to  $\Phi(x_\gamma) = 0.495$  and

in the next:  $x_\gamma = 2.58$ , and  $\delta = 2.58 \frac{\sigma_\gamma}{\sqrt{N}}$ .

Then the confidence interval will have limits

$$: (\bar{x}_B - 2.58 \frac{\sigma_\gamma}{\sqrt{N}}, \bar{x}_B + 2.58 \frac{\sigma_\gamma}{\sqrt{N}}) \quad (6)$$

Therefore, with probability 0.99 we can be sure that the interval in eq. (6) covers parameters  $m_x$ , or in other words,

with probability 0.99 magnitudes  $\bar{x}_B$  gives a value for the parameter  $n_x$  accuracy:

$\delta = 2.58 \frac{\sigma_\gamma}{\sqrt{N}}$ . From the experimental data when  $N = 50$  is

obtained  $\max(\sigma_\gamma) = 14$  and  $\delta = 2.58 \frac{14}{\sqrt{50}} \approx 5.1$ . Rounded to

obtain  $\sigma = 6$ , so it is the minimum difference between the levels of background brightness and dynamic object in the algorithm are not less than 6.

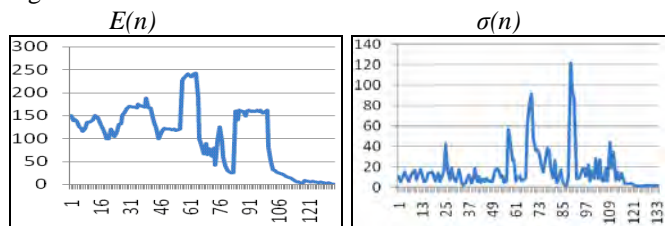


Fig.10. Values reported minimum and maximum luminance for each pixel  $E(n)$  and the values of dispersion  $\sigma(n)$  for each pixel in row.

Special attention should be paid to the pixels in the image located at the boundary of high contrast areas of the stage, the peculiarity of which lies in the sharp rise in values of dispersion. Such an increase of the variance explained by the error of quantization, i.e., quantization border pixel is set to tout for one, then the other area of the image. The minimum and maximum values of accounts brightness for each pixel

along a row of image, processed statistically in the middle of MS Excel, containing contrast areas that are visually shown in Figure 11 shows the values of the variance  $\sigma_x$ .

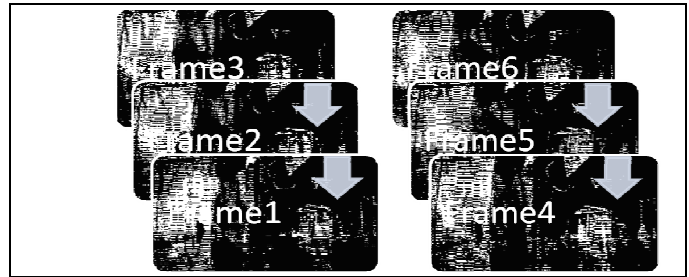


Fig. 11 The series stereo images containing dynamic object Together, graphics dispersions of samples for every pixel lines in total have kind of figure 12.

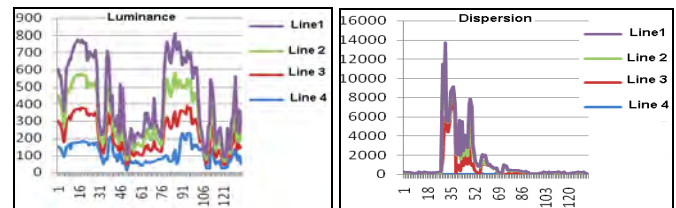


Fig.12 The luminance  $E(n)$  and dispersion  $\sigma(n)$  superimposed on each pixel for each row line of successive frames

On the right side are prominent peaks corresponding pixel in the contrast image area, but in contrast to the dynamic object are disposed in the same locality, and their position does not change in time, which can be taken into account in the subsequent separation of a dynamic object.

### IV. CONCLUSION

The Image Processing Toolbox provided in MATLAB allowed the process of developing and testing the algorithm to be more efficient. Object detection and tracking has been an active research area for a long time because it is the initial important step in many different applications, such as video surveillance, face recognition, image enhancement, video coding, and energy conservation.

### REFERENCES

- [1] Diaz J., E. Ros, F. Pelayo, "Fpga-based real-time optical-flow system," Circuits and Systems for Video Technology, IEEE Transactions on, vol. 16, Feb. 2006
- [2] Furi A., Hang H.M., An efficient block-matching algorithm for motion compensated coding, Proc. JSASSP, pp.1063-1066, 2007.
- [3] Хуанга М. Обработка изображений и цифровая фильтрация, Под ред. Т:Мир 2009.
- [4] Spirov R., Practical Object Tracking System on FPGA, 5-th Intern. Conference on Communications, Electromagnetic and Medical Application - CEMA2012, Athens, Greece, ISSN 1314-2100.
- [5] Advanced Microcontroller Final Projects, or online at: <http://people.ece.cornell.edu/land/courses/ece5760/FinalProjects>

# An Approach for Renal Vessels Evaluation in US Images

Veska Georgieva<sup>1</sup>, Stephan Vassilev<sup>2</sup>

**Abstract** – A Doppler ultrasound helps doctors by recording sound waves reflecting of fluids like blood and gives a visual or audible representation of its movement through veins, arteries and blood vessels. The resulting image is a colour overlay and it also indicates direction and blood speed. The paper presents an approach for image segmentation and renal vessels evaluation. Noise reduction of the original image is proposed as first stage of processing for obtaining a clearly contours of the objects. It is following by colour segmentation on the base of k-mean clustering and calculation of some statistical parameters in regard to evaluation some kidney diseases. The proposed algorithm is implemented in the MATLAB environment. Some results are given to demonstrate the application of this approach in clinical diagnosis.

**Keywords** – Doppler ultrasound images, renal vessels, colour segmentation, k-mean clustering.

## I. INTRODUCTION

The color Doppler ultrasound (US) makes use of a special technology to visualize the blood flowing in arteries and veins. Here the sound waves are converted to color pictures on a screen so that flow can be seen through the blood vessels (color Doppler) or on a graph showing changes in velocity. Both arteries and veins are encoded by this method. The arteries are encoded red, as the flow is coursing toward the probe to the periphery of the kidney. The veins are encoded blue. The Fig.1 presents an anatomy of kidney [1].

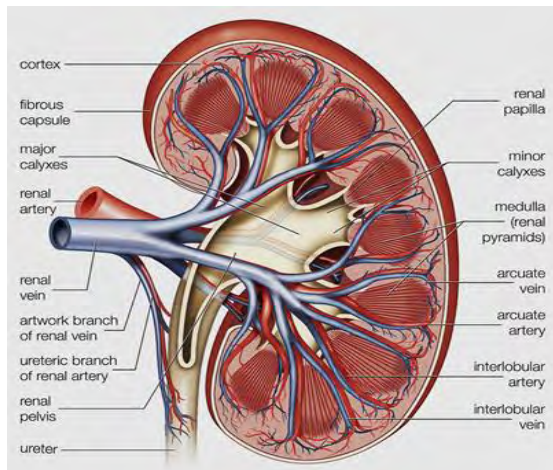


Fig. 1. Anatomy of the kidney

<sup>1</sup> Veska Georgieva is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: [vesg@tu-sofia.bg](mailto:vesg@tu-sofia.bg)

<sup>2</sup>Stephan Vassilev is with Faculty of German Engineering Education and Industrial Management, TU-Sofia, Kl. Ohridsky str.8, Sofia, Bulgaria, E-Mail: [stephan.vassilev@gmail.com](mailto:stephan.vassilev@gmail.com)

Doppler US has the advantages of being noninvasive and inexpensive. However, considerable controversy exists with regard to the role of Doppler US in screening for renal artery stenosis (RAS) or other diseases of the renal vessels [2, 3]. Two approaches are used to detect RAS with Doppler US [4]:

- direct visualization of the renal arteries;
- analysis of intrarenal Doppler waveforms.

The method of direct visualization is used in this paper, which is preferred to renal vessels evaluation. The entire length of the renal artery (RA) can be seen in this view, although when vessel tortuosity is present, the renal artery may be seen in segments. Sometimes the images are less than optimal. If there is uncertainty, images should not be acquired. However, if the renal artery can be identified coming in from the rest of the noise, then it can be sampled.

We propose to preprocess the original image in regard to noise reduction of speckle noise and segment the processed image on the base of k-mean clustering. The color segmented image can present separately the renal vessels in the obtained cluster images. This approach can be very helpful by obtaining a better evaluation of the statistical parameters of arteries and veins. The results are expressed as numbers of diameters, mean diameter, standard deviation of diameters, min and max diameters, segment length, the ratio diameter/length and tortuosity. This statistic can help the doctor to determine the probability of having RSA [5].

The remainder of this paper is organized as follows: section 2 describes the basic steps in the proposed algorithm; section 3 presents the experimental results to show the effectiveness of the method and; section 4 presents our conclusion and some future works.

## II. BASIC STAGES IN ALGORITHM FOR IMAGE PROCESSING

In the paper is proposed an effective algorithm for renal vessels evaluation in US images. It consists of following basic stages:

- Speckle noise reduction
- Segmentation of renal vessels
- Calculation of some statistical parameters of the segmented vessels
- Evaluation of the obtained results

The block diagram of the algorithm is presented in Fig.2.

Image pre-processing is performed in order to reduce the speckle noise, enhance the visual quality and for obtaining a clearly contours of the vessels.

The multiplicative nature of the speckle noise is converted to an additive one through log transformation and the problem of despeckling is thus reduced to the estimation of US image in the presence of additive noise:

$$\ln I(x, y) = \ln g(x, y) + \ln n(x, y) \quad (1)$$



where,  $I$  and  $g$  are the observed noisy image and noise-free image, respectively,  $n$  is the noise variable modeled as a stationary unity means random variable independent of  $g$ .

We have used for noise reduction a modified homomorphic filter based on wavelet packet decomposition issued from a given orthogonal wavelets and adaptive threshold of the transformed US image. The noise reduction is made in YUV color space for more effectiveness of the filtration. [6].

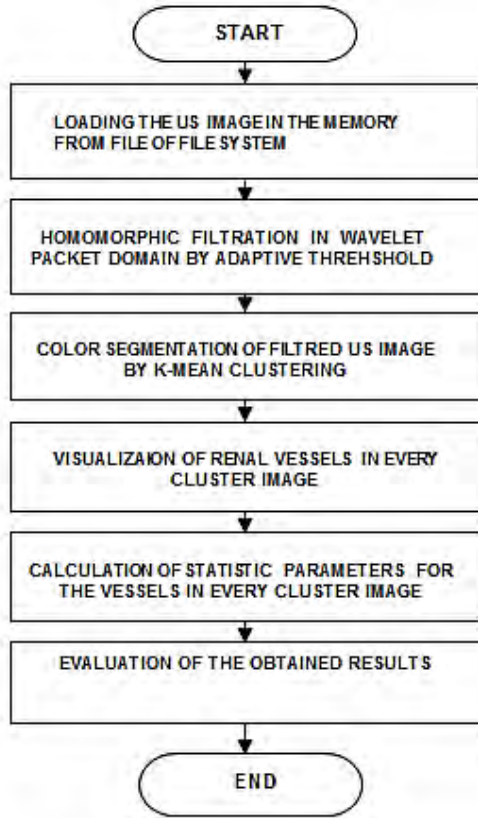


Fig. 1. Block diagram of the basic algorithm

The wavelet transform now becomes a linear operation. In this case the application of the WPT to the noisy image  $ln I(x,y)$  gives:

$$W[ln I(x,y)] = W[ln g(x,y)] + W[ln n(x,y)] \quad (2)$$

where  $W$  is the WPT operator. So the Eq. (2) can be rewritten as:

$$Y = X + N \quad (3)$$

where,  $Y$ ,  $X$ , and  $N$  are the random variables representing wavelet coefficients, respectively, of the noisy data, noise-free data, and the noise in Equation (1).

The signal components in the detail sub-bands after wavelet decompositions have heavier tails and are sharply peaked at zero. They are described by the zero-mean generalized Gaussian distribution (GGD) [7]. For modelling an image with speckle noise is used the generalized Gaussian distribution and generalized gamma distribution, which is more general to approximate the speckle statistics in the sub-band representation of US images.

To obtain a precisely detection of renal vessels and to analyze separately the arteries and veins in filtrated image a color classification based on k-mean clustering is used. This method aims to partition  $n$  observations into  $k$  clusters in which each observation belongs to the cluster with the nearest mean. Color-based Segmentation using k-mean clustering segments colors in an automated fashion using the  $L^*a^*b^*$  color space and k-means clustering method. K-means clustering treats each object as having a location in space. It finds partitions such that objects within each cluster are as close to each other as possible. K-means requires that the number of clusters to be partitioned should be specified and also a distance metric to quantify how close two objects are to each other [8].

Let the feature vectors derived from  $l$  clustered data be  $X = \{x_i, i=1,2,...,l\}$ . The generalized algorithm initiates  $k$  cluster centroids  $C = \{c_j, j=1,2,...,k\}$  by randomly selecting  $k$  feature vectors are grouped into  $k$  clusters using a selected distance measure such as Euclidean distance so that

$$d = \|x_i - c_j\| \quad (4)$$

The next step is to precompute the cluster centroids based on their group members and then regroup the feature vector according to the new cluster centroids. The clustering procedure stops only when all cluster centroids tend to converge. This approach can be very helpful by obtaining a better evaluation of the statistical parameters of arteries and veins. The results are expressed as numbers of diameters, which corresponds to number of renal vessels, mean diameter, standard deviation of diameters, min and max diameters, segment length, the ratio diameter/length and tortuosity.

The measurement of the RA diameter was performed in the proximal segment of the renal artery at a distance of 1 to 1.5 cm from the ostium, where the renal artery reaches a uniform width [9]. The widest caliber was measured perpendicular to its long axis. If the kidney has single renal artery, this artery is accepted as a main renal artery (mRA) [5]. If the kidney has multiple arteries, the largest one is accepted as an mRA. On the base of the measured diameters the statistic parameters are calculated.

### III. EXPERIMENTAL RESULTS

The experiments are made in MATLAB 7.14 environment by using IMAGE PROCESSING TOOLBOX and WAVELET TOOLBOX. For the investigations are used 20 color Doppler US images with size 640x480 pixels of the kidney. Some of the obtained results are shown in the next figures below. In Fig. 3 is presented the original image and in Fig.4 is shown its enhancement modification after homomorphic filtration based on wavelet packet decomposition by using of adaptive threshold.

The obtained modifications after color segmentation based on k-mean clustering are shown in Fig.5 and Fig.6. The cluster images present the separated renal vessels in arteries and veins.



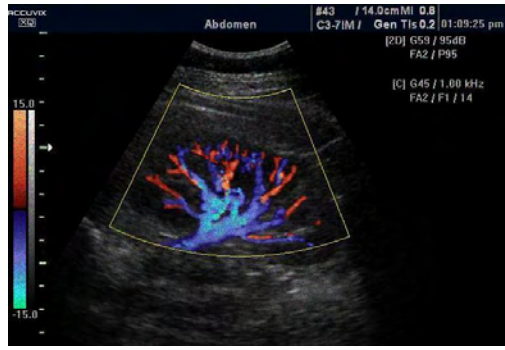


Fig. 3. Original US image

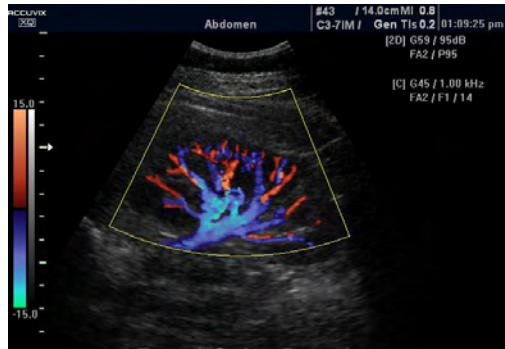


Fig. 4. US image after homomorphic filtration



Fig. 5. US cluster image with renal artery



Fig. 6. US cluster image with renal veins

The results from automatic calculated numbers of diameters, which corresponds to number of renal vessels, mean diameter, standard deviation of diameters, min and max diameters, segment length, the ratio diameter/length and

tortuosity, which are obtained by analysis in every cluster image are presented in Table 1.

The obtained third cluster image by the segmentation gives some information about the anomaly in blood vector flow in the renal vein. It is shown in Fig.7.

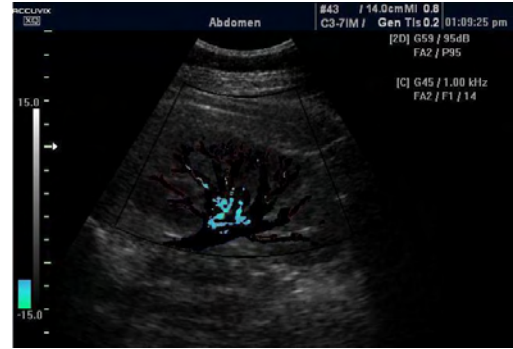


Fig. 7. US cluster image with renal vein anomaly

TABLE I  
EXPERIMENTAL RESULTS FOR STATISTIC PARAMETERS

Cluster image	N <sub>0</sub> diam.	Mean diam.	Stand. deviation	Min diam.	Max diam.	Segm. lenght	Diam. /lenght
Renal artery	30	9.31	1.19	7.28	10.1	28.2	0.32
Renal vein	6	3.41	0.38	3.13	4.04	11.8	0.28
Anomaly renal vein	5	3.46	0.32	3.13	3.94	4.22	0.82

The presented measured values of the parameters are in pixels. It can be calibrated in *cm* or *mm*. The tortuosity for the segments in all clusters is 1px.

The results are evaluated by calculating the measures accuracy, sensitivity and specificity. The values were calculated for each image and, then, averaged over all images. These results are compared to results obtained by evaluation of renal vessels in CT images [5]. The obtained results are given in Table 2.

TABLE II  
EXPERIMENTAL RESULTS FOR MEASURES ACCURACY, SENSITIVITY AND SPECIFICITY IN %

Images	Accuracy	Sensitivity	Specificity
CT	97	54	87
US	95.45	53.35	84.24

#### IV. CONCLUSION

In this paper is proposed an effective approach for automatic renal vessels evaluation in color Doppler US images. It consists noise reduction by homomorphic filter on the base of wavelet packet decomposition. Then an automatic color segmentation on the base of k-mean clustering is applied. It is very helpful by obtaining a better evaluation of the statistical parameters of arteries and veins. Experimental results showed that the proposed approach can evaluate the renal vessels such as arteria and vein accurately with average

accuracy, sensitivity and specificity of 95.45%, 53.35% and 84.24%, respectively, which is satisfactory. It shows that Doppler US can be a very effective method in screening for renal artery stenosis (RAS) or other diseases of the renal vessels. From other side color Doppler US has the advantages of being noninvasive and inexpensive.

Our future works will be focused in applying of other methods of segmentation for more effectiveness of automatic detection of different diseases in the renal vessels.

#### ACKNOWLEDGEMENT

The paper was supported by the PhD Students Research Project “Algorithms for US Image Segmentation based on Wavelet Transformation”, NIS № 152 PD0019-07/2015.

#### REFERENCES

- [1] Encyclopedia Britannica: [www.britannica.com](http://www.britannica.com)
- [2] D.Evans, J. Jensen, M.Nielsen, „Review, Ultrasonic colour Doppler imaging”, Journal Interface Focus, vol.1, pp.490–502, 2011.
- [3] G.Faustino, M. Gattass, P. Carvalho, S. Rehen and C. de Lucena, *Automatic Embryonic Stem Cells Detection and Counting in Fluorescence Microscopy Images*, Monografias em Ciência da Computação, ISSN 0103-9741, No. 04, 2009.
- [4] U.M.Hamper, M.R.Delong, L.M.Scott, „Ultrasound evaluation of the lower extremity veins”, Ultrasound Clin. Vol.4, pp.193–216, 2009.
- [5] Selma Uysal Ramadan, Hasan Yiğit, Dilek Gökharman, Işıl Tunçbilek, N. Anıl Dolgun, Pınar Koşar, Uğur Koşar, „Can renal dimensions and the main renal artery diameter indicate the presence of an accessory renal artery” , Cardiovascular Imaging, vol.17, issue 3, pp. 266-271, 2011.
- [6] V.Georgieva, S.Vassilev, “Despeckling of Medical Ultrasound Images based on Wavelet Decomposition”, Proceeding of CEMA’15, pp. Sofia, 2015.
- [7] B.Deka, P.Bora, Wavelet-based Despeckling of Medical Ultrasound Images, IETE Journal of research, Vol.59, pp.97–108, 2013.
- [8] Rakesh M, Ravi T. “Image segmentation and detection of tumor objects in MR brain images using FUZZY C-MEANS (FCM) algorithm”, International Journal of Engineering Research and Applications (IJERA) vol.2(3), pp. 2088-2094, 2012.
- [9] Aytac SK, Yigit H, Sancak T, Ozcan H., “Correlation between the diameter of the main renal artery and the presence of an accessory renal artery. J Ultrasound Med, vol. 22, pp.433–439q 2003.

# The Insertion of the Encrypted Low-Resolution Watermark in the Uncompressed Video

Zoran Veličković<sup>1</sup>, Zoran Milivojević<sup>1</sup> and Marko Veličković<sup>1</sup>

**Abstract** – The protection of the original video content is becoming an important activity of the author before publication. Inserting the encrypted watermark in the video is often an applied technique for this purpose. The watermark is first encrypted with multistage Arnold transformation and then inserted into the video with the reliable SVD algorithm. In this paper, it is shown that the lower resolution watermark gives better PSNR than higher resolution watermark. To determine the location into the frame where the low-resolution watermark is inserted, the original algorithm is developed. The extracted low-resolution watermark has a 25% higher SSIM index as compared to a higher resolution watermark.

**Keywords** – Digital video, Multistage Arnold transformation, Reliable watermarking SVD algorithm, H.264/AVC.

## I. INTRODUCTION

The use of the multimedia content has become the standard on the Internet thanks to the exceptional communications features of modern network technologies. In terms of accessibility, sharing of digital multimedia contents, especially video, is becoming the dominant form of network traffic [1]. This fact is especially contributed by the Web portal Youtube, which has provided opportunities for every Internet user to upload his own video, and thus, he becomes available globally.

These circumstances, as well as the characteristics of digital multimedia contents that, unlike analog, while copying does not lose its quality, favored the occurrence of piracy, i.e. illegal copying and distribution of multimedia content on the Internet. Financial losses as a result of piracy in the modern world are measured in billions of dollars [2]. The protection of the video content from illegal copying and distribution becomes bound activity when posting the original video content.

The standard cryptographic techniques are applicable for copy protection of digital multimedia contents, but for practical application, methods based on inserting a watermark into a multimedia content are more convenient. The inserted watermark remains permanently in the multimedia content, and it is not removed even in playback. The feature of the

digital watermark that never leaves multimedia, further increases the security level of the protected video. Also, this feature allows the survival of the watermark in the video content in the attempt of many malicious and destructive attacks on the video.

This paper deals with the insertion of the encrypted watermark in the uncoded video content in order to protect it from copying and proving ownership of the multimedia content [3], [4]. This concept of copyright protection is enabled by anti-piracy legislation, which includes the efficient and reliable extraction of the watermark from the multimedia content. The extracted watermark should match the inserted watermark and unambiguously identify the author by the content and appearance, that is, the owner of the multimedia content.

The techniques of the watermark include invisible digital images in the video content to the eye of the beholder [5], [6]. The strength of the watermark embedding in the video can be regulated by the inserting factor whose value should cause only imperceptible artifacts in the video. The larger inserting factor provides the extraction of better quality of the watermark while the video is lower quality. In this paper, the low-resolution watermark embedding technique that causes significantly less artifacts in the video content for the same value insertion factor is proposed.

For the proposed technique it is characterized that the dimensions and watermark resolution are reduced before insertion, and then the watermark is encrypted by an algorithm based on the invertible chaotic maps [7]. In this paper, the use of two-dimensional invertible chaotic maps [8] is proposed, unlike [7] where the hybrid techniques to encrypt the watermark are used. Instead of directly inserting the original content of the watermark, the watermark is inserted in each video frame transformed by multistage Arnold transformation. Applying multistage Arnold transformation, a cryptologic space is significantly increased and thus, the protection of the video content is set to a higher level. To obtain the original watermark from the encrypted one, the inverse multistage Arnold transformation [6] has to be applied. In addition to knowing the transformation parameters, it is necessary to know the initial conditions for the application of the inverse Arnold transformation, because the chaotic maps are especially sensitive on them.

The standard SVD algorithm [3], [5], [7] of the watermark insertion requires the same dimensions of the video and watermark, which, in this case, is not permitted. Before the insertion of the low-resolution watermark, it is necessary to determine a location in a frame where the watermark will be inserted. The selection of the optimal location of the watermark insertion is based on a choice part of a video frame with the greatest contrast of texture. The reason for this lies in human HVS [9] who is less sensitive to changes of parts of an

<sup>1</sup> Zoran Veličković is with College of Applied Technical Sciences, Nis, Serbia, A. Medvedeva 20, 18000 Niš, Serbia, E-mail: zoran.velickovic@vtsnis.edu.rs

<sup>1</sup> Zoran Milivojević is with College of Applied Technical Sciences, Nis, Serbia, A. Medvedeva 20, 18000 Niš, Serbia, E-mail: zoran.milivojevic@vtsnis.edu.rs

<sup>1</sup> Marko Veličković is with College of Applied Technical Sciences, Nis, Serbia, A. Medvedeva 20, 18000 Niš, Serbia, E-mail: marko.velickovic93@yahoo.com

image with greater texture contrast. Thus, the optimal location of the watermark should enter the least noticeable artifacts in the video frame. This idea has the consequence that the location of insertion depends on the content of the video frame. The locations where the watermark is inserted into individual frames can be used as a key in the watermark extraction process, and they are usually remembered in the form of a two-dimensional vector.

The insertion of the reduced and encrypted watermark in the video part of the frame is realized by the reliable SVD algorithm [3], [5]. Before presenting the protected video to customers it can be coded. In this paper we used the H.264/AVC encoder. The extracted watermarks are variable quality, and it is necessary to apply the advanced algorithm for watermark quality corrections [7], [8].

The structure of the work is as follows. The second section discusses the application of the multistage Arnold chaotic 2D map to encrypt the contents of the watermark. The third chapter shows the original algorithm to select the optimal location of the watermark in the frame based on the co-occurrence matrix. The fourth chapter presents the application of the proposed algorithm for the protection video encoded with an H.264/AVC encoder and the obtained results are analyzed. In the fifth chapter the corresponding conclusions based on conducted tests are derived.

## II. MULTISTAGE ARNOLD TRANSFORMATION

The cryptographic techniques based on two-dimensional (2D) maps are chaotic deterministic systems whose behavior is highly dependent on the initial conditions. The basic idea of the 2D chaotic map is to make relocation pixels of the original image for the spatial decorrelation of adjacent pixels. Spacious pixels decorrelation actually represents watermark encryption which makes it unrecognizable to the observer.

### A. Arnold transformation

For the image encryption invertible Arnold transformation [7] whose generalized form is defined with equations (1) and (2) is most commonly used.

$$\begin{bmatrix} x_{n+1} \\ y_{n+1} \end{bmatrix} = \begin{bmatrix} 1 & b \\ a & ab+1 \end{bmatrix} \begin{bmatrix} x_n \\ y_n \end{bmatrix} \bmod N \quad (1)$$

$$x_n, y_n \in [0, 1, 2, \dots, N-1] \times [0, 1, 2, \dots, N-1] \quad (2)$$

The arranged pairs  $(x_n, y_n)$  and  $(x_{n+1}, y_{n+1})$  are the coordinates of the image pixel (in our case the watermark) before and after transformation, respectively. The symbols  $a$  and  $b$  are positive integers that represent known parameters of Arnold transformation. It should be noted that, according to the definition, the transformation matrix determinant is equal 1, and thus, a transformational space is preserved. The parameter  $N$  is the dimension of the square image (watermark) in pixels. The main benefit of Arnold transformation is practically the watermark encryption, so that its content is not recognizable to the eye of the beholder. Another important feature of Arnold transformation is that it has a cyclical nature, i.e. after  $T$  iterations of Arnold transformation the initial image is received (in our case the original watermark).

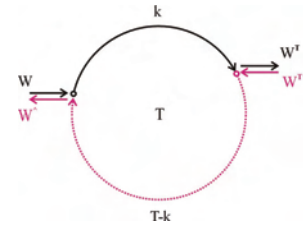


Fig. 1. The basic parameters of the generalized Arnold transformation.

The period  $T$  depends on transformation parameters  $N$ ,  $a$ , and  $b$ , but there are not generalized analytical method for its calculation. Determination of period  $T$  is usually done in practice, successive applications of Arnold transformation and analysis of the obtained results. If Arnold transformation is applying  $k$  times, the original image is obtained by the inverse Arnold transformation (further application of Arnold transformation  $T-k$  times). Fig. 1. shows the basic parameters of the generalized Arnold transformation. The original watermark is indicated by  $W$ , while the transformed watermark is labeled as  $W^T$ . A period of Arnold transformation is marked as  $T$ , and a number of the consecutive Arnold transformation for obtaining a transformed watermark is labeled as  $k$  (solid lines). In order to obtain the original watermark from the transformed one, it is necessary to perform the inverse Arnold transformation (more  $T-k$  consecutive Arnold transformation), which is shown in Fig. 1 (dashed lines).

### B. Multistage Arnold transformation

The basic idea of multistage Arnold transformation is based on the sequential application of multiple Arnold transformation (stages) with different parameters. The transformation parameters of  $i$ -th stage  $a_i$ ,  $b_i$ , the number of consecutive stages  $k_i$  and period of Arnold transformation stage  $T_i$  actually represent keys for watermark encryption/decription. It is clear that with the increasing number of stages in the multistage Arnold transformation the number of transformation parameters is growing which exponentially increases the transformation and search space.

In Fig. 2 multistage three-stage Arnold transformations has been shown, which has been used in the evaluation section of this paper. In Fig. 2 three stages have been applied - different Arnold transformation in a row.

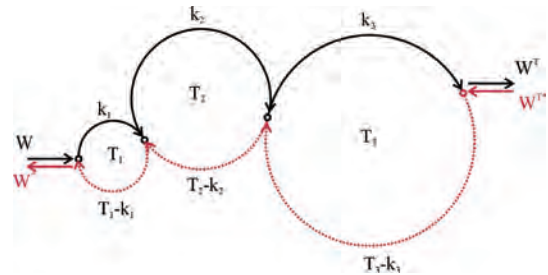


Fig. 2. Example of multistage Arnold transformation of three stages applied in the work with the following parameters:  $k_1=8$ ,  $T_1=24$ ,  $k_2=31$ ,  $T_2=48$ ,  $k_3=12$ ,  $T_3=96$ .

The first stage has a period  $T_1 = 24$ , the second  $T_2 = 48$  and the third  $T_3 = 96$ . The number of successive transformations of



the first stage is  $k_1=8$ , the second stage  $k_2=31$  and third stage  $k_3=12$ . In the case of Fig. 2, at the outlet of the third stage the transformed watermark is received that will later be inserted into each frame of the uncoded video.

### III. THE LOCATION OF THE LOW RESOLUTION WATERMARK

HVS (Human Visual System) is a dynamic visual human tool by which he learns, recognizes and understands the world around him. In the literature the characteristics of HVS related to sensitivity and brightness are often considered. Some of the HVS characteristics which are interesting for the process of the watermark insertion in the video are: image brightness, image edge and image texture.

#### A. Image Brightness

The Human visual system is able to recognize a large number of gray levels in the image, while it is significantly less sensitive to the presence of color. Although HVS is able to detect small changes in image brightness, there are limits below which it is not possible to detect changes.

#### B. Image Edge

The edges in the image represent the locations where the significant variations in brightness of pixels to a given direction are detected. By detecting the edges in the image, actually homogeneous discontinuity parts of the picture are detected. The watermark insertion in areas with pronounced edges in the image will be less noticeable.

#### C. Image Texture

The texture is an important feature of the image area, and in this paper is used for selection of the most suitable locations in the frame for the watermark insertion. By means of HVS the ability of human to get ahead of the changes in the brightness of pixels that appears in the homogeneous regions of the image compared to those with high contrast is explained. How can this fact affect the selection of a location for the watermark insertion in the video content? If you want to accepted this fact in the process of the watermark insertion, the insertion should be avoided in those parts of the frame that holds the homogeneous regions. In this paper, a texture as a part of the frame with the greatest contrast will be submitted for the watermark insertion. This conclusion can be formulated as follows: Insert the watermark in the parts of the frame with the highest texture contrast. The texture contrast  $Con$  can be determined on the basis of co-occurrence matrices as follows:

$$Con = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} (x-y)^2 c(x, y), \quad (3)$$

where  $c(x, y)$  is a co-occurrence matrix of size  $N \times N$  pixels. The co-occurrence matrix contains information how often a pixel with gray level  $i$  is located in horizontal direction of pixels with gray level  $j$ . If there is a significant variation in the picture, a co-occurrence matrix  $c(x, y)$  will generally have the elements outside the main diagonal, and the contrast of the

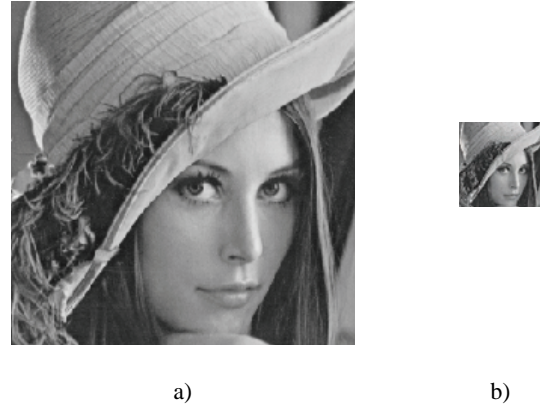


Fig. 3. Two watermarks used in the work a) high resolution watermark HiRLena 288×288 pixels and b) low resolution watermark LowRLena 72×72 pixels.

texture will be higher. The co-occurrence matrix with elements grouped in and around the main diagonal corresponds to a homogeneous region of the image. For the purpose of this study each frame is divided into 16 non-overlapped region of the same size 72×72 pixels whose contrast of texture is determined.

### IV. EXPERIMENTS AND RESULTS

For the comparison of the obtained results, a central part of the famous image "Lena" at the resolution of 288×288 pixels ("HiRLena") was used as the watermark in the first experiment, while in the second experiment the lower resolution watermark of 72×72 pixels ("LowRLena") was used. In Fig. 3 the watermarks which are used in a) the first and b) the second experiment are presented. Three-stage Arnold transformation is applied in order to encrypt watermarks, prior to their insertion in the video. The parameters of the applied multistage Arnold transformation are  $a_1=1$ ,  $b_1=1$ ,  $a_2=4$ ,  $b_2=2$ ,  $a_3=5$ ,  $b_3=8$ , and the parameter  $N$  for the first watermark was 288, and for the second 72. The encrypted watermarks are inserted into the famous uncoded video "Foreman.cif" in resolution of 288 × 288 pixels. The watermark "HiRLena" is inserted in each frame, while the watermark "LowRLena" is inserted only in the selected part of the frame size of 72×72 pixels. The watermark insertion is realized by the reliable SVD algorithm in all video frames with the constant insertion factor  $\alpha=0.05$ . After the video protection, video sequence encoding by means JM reference software of ITU in version 18.4 FRExt [10] was done. In Fig. 4 the 50th decoded video frames protected with watermarks a) HiRL and b) LowRLena are presented. By comparing the quality of frames displayed in Figure 4, there is evidence of greater presence of artifacts in the frame in Figure 4a). This result was expected thanks to the size of the inserted watermark. The watermark "LowRLena" is inserted in the top right corner of the frame, the presence of certain artifacts in this section can be identified with careful observation. The potential locations for inserting the low-resolution watermarks are frame segments of 72 × 72 pixels. For the shown experiment, there are a total of  $4 \times 4 = 16$  potential locations in the frame that have been marked by a number of columns and rows of non-overlapped segments. For

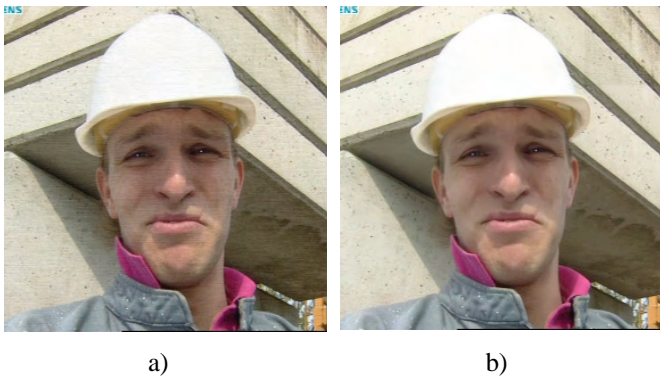


Fig. 4. The appearance of 50th decoded video frames protected with watermarks a) HiRLena and b) LowRLena.

the shown 50th frame, the algorithm for choice the location is selected location labeled (1,4) because in this frame it possesses the biggest textures contrast. The location of the low-resolution watermark insertion can be used as a key in the watermark extraction process which further increases the search space. PSNR (Peak Signal-to-Noise Ratio) is an objective measure of the quality of each frame of the video, and the video itself. In Fig. 5 the PSNR for the first 50 decoded video frames protected with the watermark inserted by HiRLena, or LowRLena is presented. In Fig. 5 it can clearly be seen that the quality of the videos inserted by the watermark LowRLena is almost 10dB better.

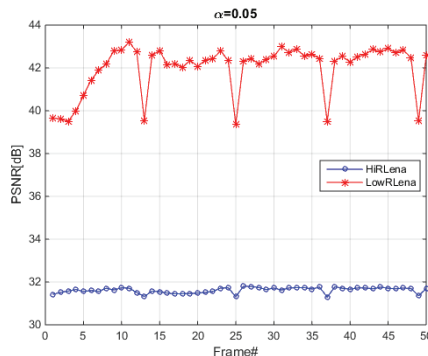


Fig. 5. PSNR values for the first 50 frames of decoded H.264/AVC video when the high-resolution watermark (HiRLena) and low-resolution watermark (LowRLena) are inserted.

Using the advanced algorithm for correcting the extracted watermark, the SSIM index of 0.84488 for the watermark LowRLena and 0.6765 for watermark HiRLena is realized. The presented data clearly show that the SSIM index was higher in extraction of the low-resolution watermark LowRLena of about 25%. Fig. 6 shows a) the progressive correction of the SIM index of the extracted watermark achieved in 26 iterations and b) the appearance of the repaired watermark.

## V. CONCLUSION

The protection of the original video content becomes very important activity before their publishing. The watermark insertion in the video itself has significant advantages over the use of standard cryptographic techniques because it permanently inserts into the video, and it is not removed during playback. In this paper, a watermark content is encrypted before the insertion by applying

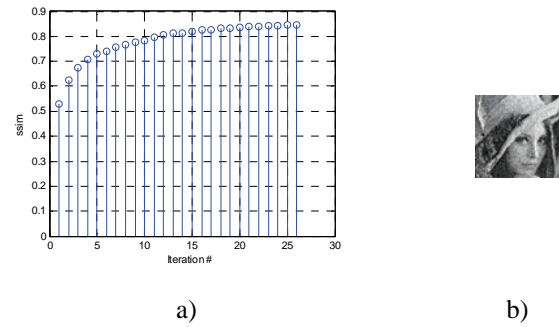


Fig. 6. a) The progressive correction of the SIM index of the extracted low-resolution watermark „LowRLena“ achieved in 26 iterations and b) the appearance of the repaired watermark with SSIM=0.84488.

multistage Arnold transformation. In this way, the encrypted space is significantly increased. The paper shows that by inserting the encrypted lower resolution watermark a higher quality of the protected video is obtained. The lower resolution watermark insertion requires determination of the location where it will be inserted into the frame. For these purposes, the original algorithm based on texture contrast of selected parts of the frame is developed. The watermark insertion is performed by the reliable SVD algorithm, thereby the occurrence of false watermarks is eliminated. After the extraction of the low-resolution watermark LowRLena, a greater SSIM index has been achieved by about 25% compared to the higher resolution watermark HiRLena. When this is not in contradiction with the implementation of the video, it is good to insert the lower resolution watermark because it will result in fewer artifacts in the protected video, and at the same time the quality of the extracted watermark will be significantly higher.

## REFERENCES

- [1] Cisco White Paper, “Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2015–2020”, Feb. 2016.
- [2] <http://www.havocscope.com/tag/movie-piracy/>
- [3] Z. Veličković, Z. Milivojević, M. Jevtović, “Reliable SVD-based Watermarking Algorithm Applied on Uncompressed Video”, ICEST 2015, pp. 254-257, Sofia, 2015.
- [4] D. Pröfrock, M. Schlauweg, E. Müllner, “A New Uncompressed-Domain Video Watermarking Approach Robust to H.264/AVC Compression”, Proc. of Signal Processing, Pattern Recognition and Applications – SPPRA, 2006.
- [5] Z. Veličković, Z. Milivojević, M. Jevtović, “The Improved Iterative Algorithm for the Enhancement of the Extracted Watermark from Video Streaming in a Wireless Environment”, *ETF Jour. of Electrical Engineering*, Vol. 21, pp. 93-107, 2015.
- [6] Z. Shahid, M. Chaumont, W. Puech, “Considering the reconstruction loop for data hiding of intra and inter frames of H.264/AVC”, *Sig. Image Video Proc.* Vol. 7, pp. 75-93, 2013.
- [7] Z. Veličković, Z. Milivojević, D. Blagojević, M. Jevtović, “Popravka kvaliteta ekstrahovanog vodenog žiga zasnovanog na haosu iz video strima”, INFOFEST, pp. 237-246, Budva-Milič, 2015.
- [8] Z. Veličković, Z. Milivojević, M. Jevtović, “Zaštita video sadržaja vodenim žigom kriptovanim multistage Arnoldovom transformacijom”, INFOTEH, pp. 316-321, Jahorina 2016.
- [9] C. C. Lai, “An improved SVD-based watermarking scheme using human visual characteristics”, *Optics Communications* Vol. 284, No. 4, pp. 938 – 944, 2011.

# Face Recognition using LBP and PCA under Different Lighting Condition

Dipak P. Patil<sup>1</sup>, Pallavi. S. Sonar<sup>1</sup> and Svetlin I. Antonov<sup>2</sup>

**Abstract** – Face is our primary and first focus of attention in social life which plays an important role in identifying an individual. Though there are variations in face due to aging and distractions like beard, glasses or change of hairstyles we can recognize a number of faces learned throughout our lifespan and identify them at a glance even after years. Face recognition system is a computer application for automatically identifying a person from digital image of a face. Making recognition more reliable under uncontrolled lighting condition is most important challenge in face recognition system. In this project we address this problem by combining the strength of robust illumination normalization, local binary pattern texture descriptor and principal component analysis. Illumination normalization is a simple and efficient pre-processing chain that eliminates most of the effects of changing illumination while still preserving the essential appearance details that are needed for recognition. Local binary pattern texture descriptor which labels the pixels of an image and gives output as a histogram of image is used for feature extraction. And principal component analysis feature dimensionality reduction algorithm is used to improve robustness. The crux of the work lies in obtaining a recognized image of a face. Such image is a recognized face image when Euclidean distance between a test image and training images is minimum.

**Keywords** – Illumination invariance, image preprocessing, local binary patterns, principle component analysis.

## I. INTRODUCTION

One of the most important aims of face recognition is to find out efficient and discriminative facial appearance descriptors which can counteract large variations in various illuminations, poses, facial expressions, ageing, partial occlusions and other related changes. Traditional approaches to deal with these issues can broadly be classified into three categories: appearance-based, normalization-based, and feature-based.

In appearance-based approach, training examples are collected under different lighting conditions which are directly used to learn a global model of the possible illumination variations, for example a linear subspace or manifold model [1]. The disadvantage of this method is that it requires large number of training images and an expressive feature set else it is crucial to include a good preprocessor to reduce illumination variations.

Normalization based approach seeks to reduce the image to a more “canonical” form in which illumination variations are suppressed e.g. histogram equalization. This has the merit of easy application to real images and lack of need for comprehensive training data.

The third approach extracts illumination-insensitive feature sets [1] [2], directly from a given face image. The feature sets may range from geometrical features [3] to image derivative features such as edge maps [2] local binary pattern (LBP) [2], [4], Gabor wavelets, and local autocorrelation filters [5].

In this paper, an integrative framework combines the strengths of all three of the above approaches. Therefore the complete process can be viewed as a pipeline involving image preprocessing, feature extraction, and, feature dimensionality reduction using PCA, classifier, as shown in Fig. 1. Each of the given stage shows increase in resistance to illumination variations and makes the information needed for recognition more evident [2].

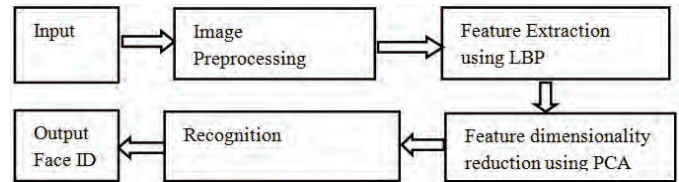


Fig.1. Stages of full face recognition method

The method focuses on robust visual features that are selected to capture as much as possible of the available information. A well-designed image preprocessing pipeline is designed to further enhance robustness.

In this paper, a simple image preprocessing pipeline is used that appears to work well for a wide range of visual feature sets, eliminating many of the effects of changing illumination while still preserving most of the appearance details needed for recognition. Finally the described LBP face descriptor, principle component analysis, measure Euclidean distance for face recognition.

## II. ILLUMINATION NORMALIZATION

Illumination Normalization is an important task in the field of face recognition. It is a preprocessing chain which is applied before feature extraction having a number of stages designed to resist the effects of illumination variation, local shadowing and different highlights. While doing so, it maintains all necessary characteristics of visual appearance. This preprocessing chain consists of following 3stages:

- A) Gamma correction;
- B) DoG filtering;
- C) Contrast equalization.

### A. Gamma Correction:

It is a nonlinear operation used to encode or decode luminance value in image. Gamma correction is, in easiest form defined as a power law expression:

<sup>1</sup>Dipak P. Patil and Pallavi. S. Sonar are with the Sandip Institute of Engineering & Management, Nashik, India, E-mail: dipak.patil@siem.org.in.

<sup>2</sup>Svetlin I. Antonov is with the Faculty of Automatics at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria. E-mail: svantonov@yahoo.com



$$I_{out} = A I_{in}^{\gamma} \quad (1)$$

where  $A$  is a constant and the input and output values are non-negative real values; in the common case of  $A = 1$ , inputs and outputs are typically in the range of 0–1. A gamma value  $\gamma < 1$  which is often called as an encoding gamma, and the process of encoding with this summarized power-law nonlinearity is called gamma compression; conversely a gamma value  $\gamma > 1$  is called a decoding gamma and the application of the expansive power-law nonlinearity is called gamma expansion.

Here, this is a nonlinear gray-level transformation that replaces gray level  $I$  with  $I^{\gamma}$  (for  $\gamma > 0$ ) or  $\log(I)$  (for  $\gamma = 0$ ), where  $\gamma \in [0,1]$  is a user defined parameter. This enhances the local dynamic range of the image in dark or shadowed regions while compressing it in bright regions and at peaks. An emphasized principle is that an intensity of the light reflected from an object is the product of the incoming illumination  $L$  (which is piecewise smooth for the most part) and the local surface reflectance  $R$  (which carries detailed object-level appearance information). A power law with exponent  $\gamma$  in the range (0 to 0.6) is a good compromise. Here use ( $\gamma = 0.3$ ) as the default setting [2].

#### B. Difference of Gaussian(DoG) Filtering:

Gamma correction does not remove the influence of overall intensity gradients such as shading effects. In computer vision, Difference of Gaussians is a grayscale image enhancement algorithm that involves the subtraction of one blurred version of an original grayscale image from one more less obscured variant of actual. Subtracting one image from other image conserves spatial information that lies between the ranges of frequencies that are preserved in the two obscured images. Therefore a difference of Gaussians is a band-pass filter that discards all but a handful of spatial frequencies that are present in the original grayscale image. As an image enhancement algorithm, the Difference of Gaussian (DOG) can be utilized to increase the visibility of edges and other detail present in a digital image. The difference of Gaussians algorithm can remove the high frequency detail that often includes random noise and this approach could be found well suitable for processing images with a high degree of noise.

The DOG impulse response is defined as [9]:

$$DoG(x, y) = \frac{1}{2\pi\sigma_1^2} e^{-\frac{x^2+y^2}{2\sigma_1^2}} - \frac{1}{2\pi\sigma_2^2} e^{-\frac{x^2+y^2}{2\sigma_2^2}} \quad (2)$$

where the default values of  $\sigma_1$  and  $\sigma_2$  are chosen as 1.0 and 2.0 respectively.

#### C. Contrast Equalization:

Contrast Equalization is also known as histogram equalization. Contrast Equalization is an approach to adjust image intensities to enhance contrast, means rescales the intensities of given image. A robust estimator is important to

be used because the signal typically contains extreme values produced by highlights, minute dark regions like that of nostrils, junk on the image edges, chin, forehead etc. Here a simple and rapid approximation is preferred based on a two stage process as follows:

$$I(x, y) = \frac{I(x, y)}{(\text{mean}(\min(\tau |I(x', y')|^\alpha))^{\alpha})^{1/\alpha}} \quad (3)$$

Here,  $\alpha$  is a strongly compressive exponent that reduces the influence of great values,  $\tau$  is an outset used for shorten large values later than the first phase of normalization, and the mean is over the whole (unmasked part) image. By default  $\alpha = 0.1$  and  $\tau = 10$  is used.



Fig.2. Output of preprocessing original RGB image, Gray image, Gamma correction image, DoG filter image, HE image.

### III. LOCAL BINARY PATTERN

Local Binary Pattern (LBP) is a simple yet very efficient texture operator which labels the pixels of an image by making a threshold on a neighborhood of each pixel and considers result as a binary number or local binary pattern histogram of an image. Due to its discriminative power and calculated clarity, LBP texture operator has become a prevailing modus operandi in various applications. Perhaps the most important characteristic of LBP operator in real-world applications is its robustness to monotonic gray-scale changes rooted, e.g. illumination variations.

An original LBP operator [6] forms labels for image pixels by threshold of  $3 \times 3$  neighborhood for each pixel with the center value and considering the result as a binary number. The histogram of these  $2^8 = 256$  heterogeneous labels can then be used as a texture descriptor. Using a circular neighborhood and bi linearly interpolating values at non-integer pixel coordinates allow any radius and number of pixels in the neighborhood.

Formally, the LBP operator takes the form [2]:

$$LBP(x_c, y_c) = \sum_{n=0}^7 2^n s(i_n - i_c) \quad (4)$$

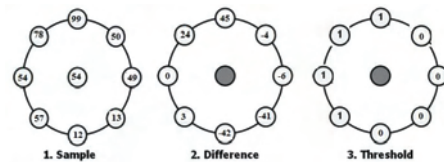


Fig. 3. An example of LBP computation

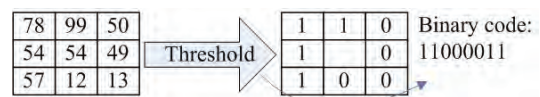




Fig.4. Illustration of the basic LBP operator for above example Where, in this case  $n$  runs over the 8 neighbors of the central pixel  $c$ ,  $i_c$  and  $i_n$  are the gray-level values at  $c$  and  $n$ , and  $s(u)$  is 1 if  $u \geq 0$  and 0 otherwise. The LBP encoding process is illustrated in Fig. 3.



Fig.5. I/P RGB image, Gray image, LBP image, Histogram of LBP

Two extensions of original operator were made in [7]. The first described LBP is for neighborhoods of various sizes, hence this makes it feasible to manage with textures of different scales. Other defined LBP uniform patterns are uniform if they contain at most one 0-1 and one 1-0 transition when viewed as a circular bit string. For example, the LBP code in Fig.4. is uniform. Uniformity is a vital concept in LBP methodology, which represent basic structural instruction such as edges, corners, etc. In these methods that histogram LBP is, a number of bins which can thus be considerably reduced by assigning all non-uniform patterns to a bin, constantly without dropping too much information. LBP image and their histogram image is shown in Fig.5.

#### IV. PRINCIPLE COMPONENT ANALYSIS

Principal Component Analysis (PCA) is one of the most successful techniques that have been used in face recognition. The functions which PCA does are prediction, redundancy removal, feature extraction, data compression, etc [8]. The main idea of using PCA for face recognition is to express the large 1-D vector of pixels constructed from 2-D facial image into the compact principal components of the feature space. It is known as eigenspace projection. Eigenspace is calculated by specifying the eigenvectors of the covariance matrix derived from a set of facial images (vectors).

Mathematics of PCA [8] - a 2-D facial image can be represented as 1-D vector by concatenating each row (or column) into a stretched narrow vector. Suppose there is  $M$  vectors of size  $N$  (=rows of image  $\times$  columns of image) representing a set of sampled images.  $p_j$ 's represent the pixel values.

$$x_i = [p_1 \dots p_N]^T, \quad (5)$$

$$i = 1, \dots, M$$

The images are mean centered by subtracting the mean image from each image vector. Mean image  $m$  is represented as:

$$m = \frac{1}{M} \sum_{i=1}^M x_i \quad (6)$$

And let  $w_i$  be defined as mean centered image

$$w_i = x_i - m \quad (7)$$

A goal is to find a set of  $e_i$ 's which have the largest possible projection onto each of the  $w_i$ 's. To find a set of  $M$  orthonormal vectors  $e_i$  for which the quantity

$$\lambda_i = \frac{1}{M} \sum_{n=1}^M (e_i^T w_n)^2 \quad (8)$$

is augmented with the orthonormality constraint

$$e_i^T e_k = \delta_{ik} \quad (9)$$

It has been shown that the  $e_i$ 's and  $\lambda_i$ 's are given by the eigenvectors and eigenvalues of the covariance matrix

$$C = WW^T \quad (10)$$

where  $W$  is a matrix composed of the column vectors  $w_i$  placed side by side. The size of  $C$  is  $N \times N$  which could be enormous. For example, images of size  $64 \times 64$  create the covariance of size  $4096 \times 4096$ . It is not practical to solve for the eigenvectors of  $C$  directly. A conventional theorem in linear algebra states that the vectors  $e_i$  and scalars  $\lambda_i$  can be obtained by solving for the eigenvectors and eigenvalues of the  $M \times M$  matrix  $W^T W$ . Let  $d_i$  and  $\mu_i$  be the eigenvectors and eigenvalues of  $W^T W$ , respectively.

$$W^T W d_i = \mu_i d_i \quad (11)$$

By multiplying left to both sides by  $W$ :

$$WW^T (W d_i) = \mu_i (W d_i) \quad (12)$$

which means that the first  $M-1$  eigenvectors  $e_i$  and eigenvalues  $\lambda_i$  of  $WW^T$  are given by  $W d_i$  and  $\mu_i$ , respectively.  $W d_i$  needs to be normalized in order to be equal to  $e_i$ . Since we only sum up a finite number of image vectors,  $M$ , the rank of the covariance matrix unable to exceed  $M-1$  (The -1 come from the subtraction of the mean vector  $m$ ).

The eigenvectors equivalent to nonzero eigenvalues of the covariance matrix produce an orthonormal basis for the subspace within which most image data can be represented with a small flaw. The eigenvectors are classified from high to low according to their correlating eigenvalues. The eigenvector associated with the highest eigenvalue is one that reflects the greatest variance in an image. It means lowest eigenvalue is related with the eigenvector that finds the least variance. A facial image can be projected onto  $M' (< M)$  dimensions by computing:

$$\Omega = [v_1 v_2 \dots v_{M'}]^T \quad (13)$$

where  $v_i = e_i^T w_i$ .  $v_i$  is the  $i^{th}$  coordinate of the facial image in a new space, came as a main component. The vectors  $e_i$  are also images, so called, *eigenimages*, or *eigenfaces* in this case. They can be viewed as images and indeed seem as faces. So,  $\Omega$  describe contribution of each eigenface in representing the facial image by treating eigenfaces as a core set for facial images. An easiest way to determine which face class provides the best description of an input facial image is to find the face class  $k$  that minimizes the Euclidean distance:

$$\delta_k = \Omega - \Omega_k \quad (14)$$

Where,  $\Omega_k$  is a vector describing the  $k^{th}$  face class.

## V. IMPLEMENTATION AND RESULT

The proposed method is tested on two different databases. One is available face database for different set of conditions and other is ORL face database. The use of only frontal face views other than lighting, expression and identity may vary.

Regarding the image size, cropping, and alignment of the datasets: Face database for different set of conditions image cropping to  $192 \times 128$  pixels.

For AT&T-ORL used the original images at  $112 \times 92$ . The default settings of different parameters are summarized in table I.

Table I  
DEFAULT PARAMETER SETTING

Procedure	Parameter	Value for available face database	Value for ORL face database
Gamma Correction	$\gamma$	0.2	0.4
DOG Filtering	$\sigma_0$	1	2
	$\sigma_1$	2	3
Contrast Equalization	$\alpha$	0.1	0.1
	$\tau$	10	10

A face image in 2-dimension with size  $M \times N$  can also be considered as one dimensional vector of dimension  $MN$ . For example, face image from trained database with size  $192 \times 128$  can be considered as a vector of dimension 24,576 or equivalently a point in a 24,576 dimensional space. Similarly face image from ORL (Olivetti Research Labs) database with size  $112 \times 92$  can be considered as a vector of dimension 10,304 or equivalently a point in a 10,304 dimensional space.

An ensemble of images maps to a collection of points in it takes huge space. Images of faces, being similar in overall configuration, will not be unspecific distribution in this huge image space and thus can be described by a relatively low dimensional subspace.

There are 10 different images of 10 different subjects for face database different set of conditions. And there are 10 different images of 30 different subjects for ORL face database. For some of the subjects, the images were taken at different times, varying lighting condition, facial expressions (open/closed eyes, smiling/non smiling) and facial details. Fig.6. shows results of preprocessing chain. Fig.7. shows histogram output of gray image and DoG filter image.

Computed recognition rate using face database for different set of conditions for 10 different subjects face images with 10 images of each person, is shown in table II. And for ORL database for 30 different subjects face images with 10 images of each person, is shown in table III.

Table II  
RECOGNITION RATE USING LBP AND PCA

Method	Recognition%
PP+LBP+PCA	94

TABLE III  
RECOGNITION RATE USING LBP AND PCA FOR ORL FACE DATABASE

Method	Recognition%
PP+LBP+PCA	92

## VI. CONCLUSION

A face recognition system is a computer application for automatically identifying a person from a digital image of face. This Project throws light on new methods for face recognition under different lighting conditions based on robust preprocessing, LBP local texture descriptor and PCA. The crux of the work lies in making three main contributions: (i) simple and efficient preprocessing chain, that eliminates most of the effects of changing illumination while still preserving the essential appearance details that are needed for recognition; (ii) a rich local texture descriptor called Local

Binary Pattern which labels the pixels of an image and gives output as a histogram of an image and (iii) Principle Component Analysis feature extraction algorithm is used to improve robustness. The crux of the work lies in obtaining a known face image. An image is a known face image when Euclidean distance between a test image and training images is minimum. And if this distance is not attained, it is an unknown face image.

## REFERENCES

- [1] Y. Adini, Y. Moses, and S. Ullman, "Face recognition: The problem of compensating for changes in illumination direction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 721–732, Jul. 1997.
- [2] X. Tan and Bill Triggs "Face Recognition under difficult lighting condition" *IEEE Transaction on Image processing* vol.19. no.7. June 2010.
- [3] R. Brunelli and T. Poggio, "Face recognition: Features versus Templates," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 10, pp. 1042–1052, Oct. 1993.
- [4] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006.
- [5] F. Guodail, E. Lange, and T. Iwamoto, "Face recognition system using local autocorrelations and multiscale integration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 10, pp. 1024–1028, Oct. 1996.
- [6] T. Ojala, M. Pietikainen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.
- [7] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.
- [8] K. Kim "Face Recognition using Principle Component Analysis" Department of computer science, University of Maryland. MD 20742, USA.
- [9] S. Anila and Dr. N. Devrajan "Processing Technique for Face Recognition under varying illumination condition". *Globe Journal of computer science and technology graphics and vision* vol.12. ver.1.0, 2001

# **Computer Systems and Internet Technologies**





# Formal Verification of Models for M2M Device Registration

Evelina Pencheva<sup>1</sup>, Ivaylo Atanasov<sup>1</sup> and Anastas Nikolov<sup>1</sup>

**Abstract** – Device management is a challenging task in Machine-to-Machine (M2M) communications. The increased number of connected devices and device diversity calls for device management that reduces the deployment time and operational costs. Open Mobile Alliance specified Lightweight M2M protocol for device management. In this paper, we propose device registration models based on Lightweight M2M procedures and a mathematical approach to behavior verification. Both server and client models are formally described, and a concept of weak bisimulation is used to prove that the models are synchronized.

**Keywords** – Device management, Finite state machines, Behavioral equivalence, Weak bisimulation

## I. INTRODUCTION

Machine-to-Machine (M2M) communications have various application areas such as automotive, home automation, smart cities, energy efficiency, industry, agriculture, safety and security, health, education and others. Despite the differences all these areas set common requirements for connected devices to communicate through different access networks, remote configuration and control. Device management is a challenging and critical issue due to rapidly growing number of connected devices and their diversity [1]. A plethora of devices and customized solutions are available on the market and a large amount of the employed technology is proprietary today [2], [3], [4]. This calls for abstraction of device management functions which has to hide the complexity and to be technology independent. Such an abstraction can be provided by OMA LWM2M [5], [6]. Lightweight M2M is a protocol from the Open Mobile Alliance for M2M device management. It defines device management procedures between a LWM2M server and a LWM2M client, which is located in a device. The protocol may be used to create device management solutions that apply the approach of software defined networks [7], [8]. The proposed solutions, based on LWM2M, consider high level architectural aspects and do not provide details on behavioral models that follow the M2M device management procedures. In this paper, we suggest an approach to formal verification of LWM2M server and client behaviour related to device management. Our models are compliant with ETSI M2M functional architecture [9] and Enabler Test Specification for Lightweight M2M [10]. First we start with formal description of device registration models

and using the well known concept of weak bisimulation [11] we prove formally that the models are synchronized. In addition to regular device registration functions, our models include functions related to server initiated device registration, updating the firmware version and server disabling, and prove that the models expose equivalent behaviour.

## II. DEVICE REGISTRATION MODELS

Device registration allows the server to maintain device reachability status. If the device is not registered it is not reachable. When the device sends a registration request ( $reg_{req}$ ) it moves to registering state, where it awaits the server answer. After receiving the server answer with registration acknowledgement, the device sets the registration timer ( $T_{reg}$ ) and moves to registered state. If the device is registered, it is with operational firmware and the server and device store registration-related information making it available, on request or based on subscription. When the registration timer expires the device refreshes its registration. When registered, the device may receive a soft offline request and then it sends a de-registration request to the server and becomes unregistered.

The device's view point on its registration state is as follows. In  $Unregistered_D$  state, the device is offline and it is not registered. In  $Registering$  state, the device is in a process of registration. In  $OperationalFw_D$ , the device is registered with operational firmware. In  $UpdateRegistration$  state a transport binding between the server and device is established and the device waits for registration update trigger message from the server. In  $WaitDeregAck$  state, the device waits for de-registration acknowledgement. In  $FirmwareDownloading_D$  state, the device downloads the new firmware version. The model representing the device's view on its registration state is shown in Fig.1.

The server's view point on device registration state is as follows. In  $Unregistered_S$  state, the device is not registered. In  $OperationalFw_S$  state, the device is registered with operational firmware. In  $NotificationStoring$  state, the device is registered and the server updates the notification storing object. In  $Disabling$  state, the server will be disabled. In  $Transport-Binding$  state, the device is registered and updates the registration. In  $WaitUpdateAck$  state, the server waits for acknowledgement of transport binding. In  $WaitFwVersion$  state, the device is registered and the server reads the current firmware version. In  $WaitDownloadAck$  state, the device is registered and the server initiates the download of a new firmware version. In  $FirmwareDownloading_S$  state, the device downloads the new firmware version. In  $WaitFwActionStatus$  state, the server asks for the firmware downloading status. In  $WaitFwUpdate$  state, the server waits for acknowledgement of

<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: enp@tu-sofia.bg; iia@tu-sofia.bg; nikolov.anastas@gmail.com

firmware update. In RemoveOldFirmware state, the server waits for acknowledgement that the old firmware version is removed. In Rebooting state, the server waits device rebooting. The model representing the server's view on device registration state is shown in Fig.2.

### III. FORMAL MODEL VERIFICATION

Both models are formally described as Labeled Transition Systems (LTS).

By  $R_D = (S_D, Act_D, \rightarrow_D, s_0^D)$  it is denoted an LTS representing the authorized device's view on its registration state, namely:

$$\begin{aligned}
 S_D &= \{ \text{Unregistered}_D, \text{Registering}, \text{OperationalFw}_D, \\
 &\quad \text{UpdateRegistration}, \text{FirmwareDownloading}_D, \\
 &\quad \text{WaitDeregAck} \}; \\
 Act_D &= \{ \text{online}, T_{\text{disable}}, \text{initialReg}_{\text{ack}}, \text{reReg}_{\text{ack}}, \text{fwVersion}_{\text{req}}, \\
 &\quad \text{removeFw}_{\text{req}}, \text{updateNS}_{\text{req}}, \text{Treg}_D, \text{bind}_{\text{req}}, \\
 &\quad \text{updateReg}_{\text{com}}, \text{disable}_{\text{req}}, \text{reboot}_{\text{req}}, \text{writeFw}_{\text{req}}, \\
 &\quad \text{readFw}_{\text{req}}, \text{updateFw}_{\text{req}}, \text{softOffline}, \text{deReg}_{\text{ack}} \}; \\
 \rightarrow_D &= \{ \tau_1^D, \tau_2^D, \tau_3^D, \tau_4^D, \tau_5^D, \tau_6^D, \tau_7^D, \tau_8^D, \\
 &\quad \tau_9^D, \tau_{10}^D, \tau_{11}^D, \tau_{12}^D, \tau_{13}^D, \tau_{14}^D, \tau_{15}^D, \tau_{16}^D, \tau_{17}^D \} \\
 s_0^D &= \{ \text{Unregistered}_D \}, \\
 \text{where} \\
 \tau_1^D &= (\text{Unregistered}_D \text{ online Registering}) \\
 \tau_2^D &= (\text{Unregistered}_D T_{\text{disable}} \text{ Registering}) \\
 \tau_3^D &= (\text{Registering initialReg}_{\text{ack}} \text{ OperationalFw}_D) \\
 \tau_4^D &= (\text{Registering reReg}_{\text{ack}} \text{ OperationalFw}_D) \\
 \tau_5^D &= (\text{OperationalFw}_D \text{ fwVersion}_{\text{req}} \text{ OperationalFw}_D) \\
 \tau_6^D &= (\text{OperationalFw}_D \text{ removeFw}_{\text{req}} \text{ OperationalFw}_D) \\
 \tau_7^D &= (\text{OperationalFw}_D \text{ updateNS}_{\text{req}} \text{ OperationalFw}_D) \\
 \tau_8^D &= (\text{OperationalFw}_D \text{ Treg}_D \text{ Registering}) \\
 \tau_9^D &= (\text{OperationalFw}_D \text{ bind}_{\text{req}} \text{ UpdateRegistration}) \\
 \tau_{10}^D &= (\text{UpdateRegistration updateReg}_{\text{com}} \text{ Registering}) \\
 \tau_{11}^D &= (\text{OperationalFw}_D \text{ disable}_{\text{req}} \text{ Unregistered}_D) \\
 \tau_{12}^D &= (\text{OperationalFw}_D \text{ reboot}_{\text{req}} \text{ Registering}) \\
 \tau_{13}^D &= (\text{OperationalFw}_D \text{ writeFw}_{\text{req}} \text{ FirmwareDownloading}_D) \\
 \tau_{14}^D &= (\text{FirmwareDownloading}_D \text{ readFw}_{\text{req}} \\
 &\quad \text{FirmwareDownloading}_D) \\
 \tau_{15}^D &= (\text{FirmwareDownloading}_D \text{ updateFw}_{\text{req}} \text{ OperationalFw}_D) \\
 \tau_{16}^D &= (\text{OperationalFw}_D \text{ softOffline} \text{ WaitDeregAck}) \\
 \tau_{17}^D &= (\text{WaitDeregAck deReg}_{\text{ack}} \text{ Unregistered}_D).
 \end{aligned}$$

By  $R_S = (S_S, Act_S, \rightarrow_S, s_0^S)$  it is denoted an LTS representing the server's view on authorized device registration state as follows:

$$S_S = \{ \text{Unregistered}_S, \text{OperationalFw}_S, \text{TransportBinding}, \\
 \text{NotificationStorings}, \text{Disabling}, \text{WaitUpdateAck},$$

$\text{WaitFwVersion}, \text{Rebooting}, \text{FirmwareDownloadings}_S, \\
 \text{WaitFwActionStatus}, \text{WaitFwUpdate}, \text{WaitDownloadAck}, \text{RemoveOldFirmware}, \};$

$$Act_S = \{ \text{initialReg}_{\text{req}}, \text{reReg}_{\text{req}}, \text{Tregs}, \text{deReg}_{\text{req}}, \text{notifStoring}, \\
 \text{updateNS}_{\text{ack}}, \text{disable}, \text{fwAvailable}, \text{updateReg}, \text{disable}_{\text{ack}}, \\
 \text{bind}_{\text{ack}}, \text{updateReg}_{\text{ack}}, \text{fwVersion}_{\text{res}}, \text{writeFw}_{\text{res}}, \text{fwT}_{\text{dl}}, \\
 \text{readFw}_{\text{res}}, \text{updateFw}_{\text{ack}}, \text{remove}_{\text{ack}}, \text{reboot}_{\text{ack}} \}$$

$$\begin{aligned}
 \rightarrow_S &= \{ \tau_1^S, \tau_2^S, \tau_3^S, \tau_4^S, \tau_5^S, \tau_6^S, \tau_7^S, \tau_8^S, \tau_9^S, \tau_{10}^S, \tau_{11}^S, \\
 &\quad \tau_{12}^S, \tau_{13}^S, \tau_{14}^S, \tau_{15}^S, \tau_{16}^S, \tau_{17}^S, \tau_{18}^S, \tau_{19}^S, \tau_{20}^S, \tau_{21}^S \}; \\
 -s_0^S &= \{ \text{Unregistered}_S \},
 \end{aligned}$$

where

$$\begin{aligned}
 \tau_1^S &= (\text{Unregistered}_S \text{ initialReg}_{\text{req}} \text{ OperationalFw}_S) \\
 \tau_2^S &= (\text{Unregistered}_S \text{ reReg}_{\text{req}} \text{ OperationalFw}_S) \\
 \tau_3^S &= (\text{OperationalFw}_S \text{ reReg}_{\text{req}} \text{ OperationalFw}_S) \\
 \tau_4^S &= (\text{OperationalFw}_S \text{ Tregs} \text{ Unregistered}_S) \\
 \tau_5^S &= (\text{OperationalFw}_S \text{ deReg}_{\text{req}} \text{ Unregistered}_S) \\
 \tau_6^S &= (\text{OperationalFw}_S \text{ notifStoring} \text{ NotificationStorings}_S) \\
 \tau_7^S &= (\text{NotificationStorings}_S \text{ updateNS}_{\text{ack}} \text{ OperationalFw}_S) \\
 \tau_8^S &= (\text{OperationalFw}_S \text{ disable} \text{ Disabling}) \\
 \tau_9^S &= (\text{OperationalFw}_S \text{ fwAvailable} \text{ WaitFwVersion}) \\
 \tau_{10}^S &= (\text{OperationalFw}_S \text{ updateReg} \text{ TransportBinding}) \\
 \tau_{11}^S &= (\text{Disabling} \text{ disable}_{\text{ack}} \text{ Unregistered}_S) \\
 \tau_{12}^S &= (\text{TransportBinding} \text{ bind}_{\text{ack}} \text{ WaitUpdateAck}) \\
 \tau_{13}^S &= (\text{WaitUpdateAck} \text{ updateReg}_{\text{ack}} \text{ Unregistered}_S) \\
 \tau_{14}^S &= (\text{WaitFwVersion} \text{ fwVersion}_{\text{res}} \text{ WaitDownloadAck}) \\
 \tau_{15}^S &= (\text{WaitDownloadAck} \text{ writeFw}_{\text{res}} \\
 &\quad \text{FirmwareDownloadings}_S) \\
 \tau_{16}^S &= (\text{FirmwareDownloadings}_S \text{ fwT}_{\text{dl}} \text{ WaitFwActionStatus}) \\
 \tau_{17}^S &= (\text{WaitFwActionStatus} \text{ readFw}_{\text{res}} \\
 &\quad \text{FirmwareDownloadings}_S) \\
 \tau_{18}^S &= (\text{WaitFwActionStatus} \text{ readFw}_{\text{res}} \text{ WaitFwUpdate}) \\
 \tau_{19}^S &= (\text{WaitFwUpdate} \text{ updateFw}_{\text{ack}} \text{ RemoveOldFirmware}) \\
 \tau_{20}^S &= (\text{RemoveOldFirmware} \text{ remove}_{\text{ack}} \text{ Rebooting}) \\
 \tau_{21}^S &= (\text{Rebooting} \text{ reboot}_{\text{ack}} \text{ Unregistered}_S).
 \end{aligned}$$

Intuitively, in terms of observed behavior, two state machines have bisimilar relation if one state machine displays a final result and the other state machine displays the same result [11]. Strong bisimilarity requires existence of homomorphism between transitions in both state machines. In practice, strong bisimilarity puts strong conditions for equivalence which are not always necessary. For example, internal transitions can present actions, which are internal to the system (i.e. not observable). In weak bisimilarity, internal transitions can be ignored. The concept of weak bisimilarity is used to study the modeling aspects of M2M device registration.

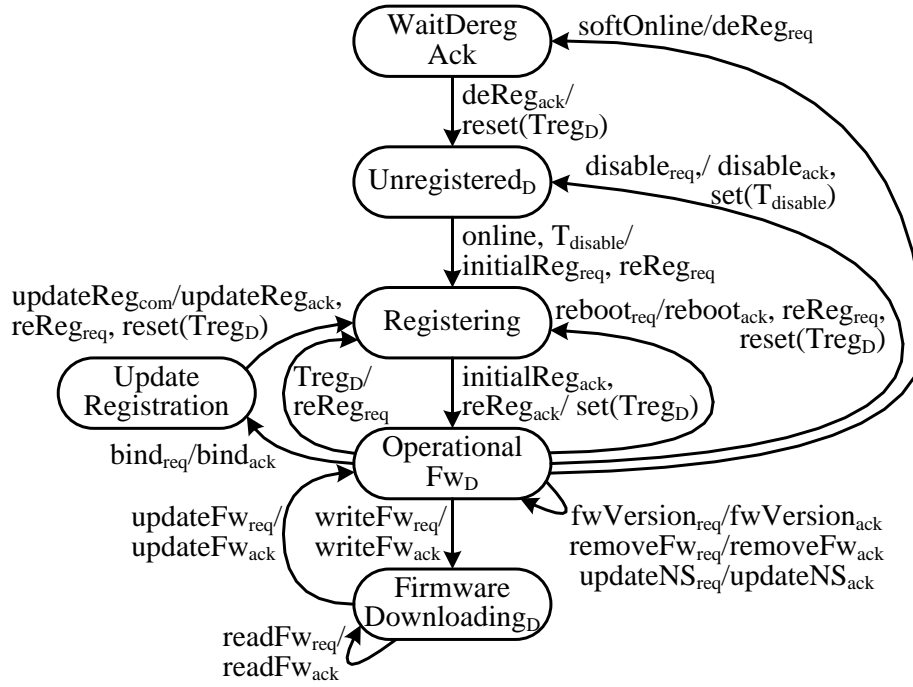


Fig. 1. Registration state of an authorized device as seen by the device

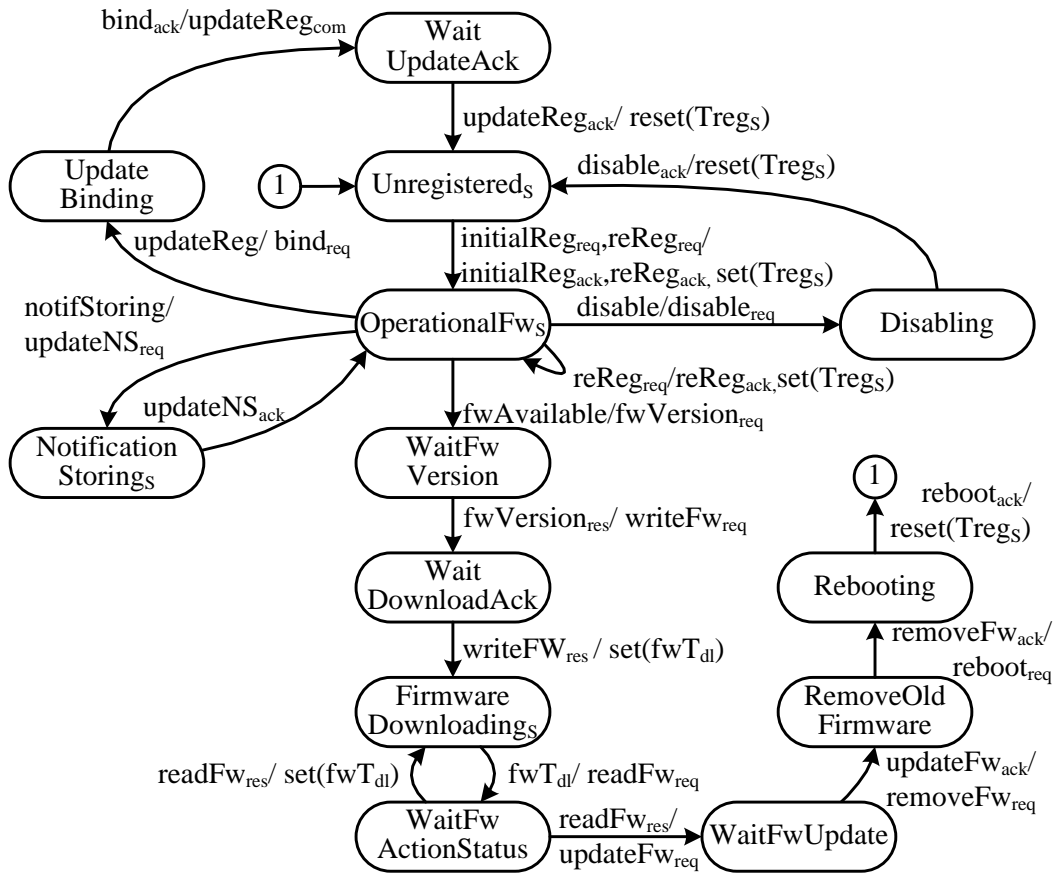


Fig. 2. Registration state of an authorized device as seen by the server

**Proposition:** The labeled transition systems  $R_S'$  and  $R_D'$  are weakly bisimilar.

**Proof:** Let  $U_{R'S'} = \{(\text{Unregistered}_D, \text{Unregistered}_S), (\text{OperationalFw}_D, \text{OperationalFw}_S)\}$ . Then:

1. For initial registration:  $\text{Unregistered}_D \{ \tau_1^D, \tau_3^D \}$   
 $\text{OperationalFw}_D \exists \text{Unregistered}_S \{ \tau_1^S \} \text{OperationalFw}_S;$
2. For re-registration after offline:  $\text{Unregistered}_D \{ \tau_1^D, \tau_4^D \}$   
 $\text{OperationalFw}_D \exists \text{Unregistered}_S \{ \tau_1^S \} \text{OperationalFw}_S;$
3. For de-registration:  $\text{OperationalFw}_D \{ \tau_{11}^D, \tau_{17}^D \}$   
 $\text{Unregistered}_D \exists \text{OperationalFw}_S \{ \tau_5^S \} \text{Unregistered}_S;$
4. For re-registration due to registration lifetime is over:  
 $\text{OperationalFw}_D \{ \tau_8^D, \tau_4^D \} \text{OperationalFw}_D$   
 $\exists \text{OperationalFw}_S \{ \tau_3^S, \tau_4^S, \tau_2^S \} \text{OperationalFw}_S;$
5. For update notification storing:  $\text{OperationalFw}_D$   
 $\{ \tau_7^D \} \text{OperationalFw}_D \exists \text{OperationalFw}_S \{ \tau_6^S, \tau_7^S \}$   
 $\text{OperationalFw}_S;$
6. For server disabling:  $\text{OperationalFw}_D \{ \tau_{11}^D \} \text{Unregistered}_D \exists$   
 $\text{OperationalFw}_S \{ \tau_8^S, \tau_{10}^S \} \text{Unregistered}_S;$
7. For re-registration when server enables:  $\text{Unregistered}_D$   
 $\{ \tau_2^D, \tau_4^D \} \text{OperationalFw}_D \exists \text{Unregistered}_S$   
 $\{ \tau_2^S \} \text{OperationalFw}_S;$
8. For update registration trigger:  $\text{OperationalFw}_D \{ \tau_9^D, \tau_{10}^D,$   
 $\tau_4^D \} \text{OperationalFw}_D \exists \text{OperationalFw}_S \{ \tau_{10}^S,$   
 $\tau_{12}^S, \tau_{13}^S, \tau_2^S \} \text{OperationalFw}_S;$
9. For update firmware version:  $\text{OperationalFw}_D \{ \tau_5^D,$   
 $\tau_{13}^D, \tau_{14}^D, \tau_{15}^D, \tau_{12}^D, \tau_4^D \} \text{OperationalFw}_D$   
 $\exists \text{OperationalFw}_S \{ \tau_9^S,$   
 $\tau_{14}^S, \tau_{15}^S, \tau_{16}^S, \tau_{17}^S, \tau_{18}^S, \tau_{19}^S, \tau_{20}^S, \tau_{21}^S, \tau_2^S \} \text{OperationalFw}_S.$

Therefore  $R_S$  and  $R_D$  are weakly bisimilar, i.e. they expose equivalent behavior.

#### IV. CONCLUSION

The paper presents models of M2M device registration status as viewed by the server and by the device. Starting with regular models representing just registered and unregistered device state, we expand the models with additional functionality including server triggered registration update, firmware version update and server disabling. We describe models formally and prove the model synchronization by using the concept of weak bisimilarity. The models are applicable to Device Reachability, Addressing and Repository Service Capability which allows re-use in different M2M applications.

#### REFERENCES

- [1] A. Sehgal, V. Perelman, S. Kuryla, J. Schönwälder. "Management of Resource Constrained Devices in the Internet of Things," IEEE Communications Magazine, December 2012, pp.144-149.
- [2] V. Tayur, R. Suchithra. "Software Defined Unified Device Management for Smart Environments," International Journal of Computer Applications, 2015, vol. 121, issue 9, pp.30-34.
- [3] D. Schulz, R. Gitzel, "Seamless maintenance - Integration of FDI Device Management & CMMS," IEEE Conference on Emerging Technologies & Factory Automation (ETFA), 2013, pp.402-407.
- [4] C.S. Shih, C. T. Chou, K. J. Lin, B. L. Tsai, C. H. Lee, D. Cheng, C. J. Chou "Out-of-Box Device Management for Large Scale Cyber-Physical Systems," IEEE International Conference on Internet of Things (iThings), and Green Computing and Communications (GreenCom), and Cyber, Physical and Social Computing (CPSCom), 2014, pp.402 – 407.
- [5] V. Cackovic, Z. Popovic. "Device Connection Platform for M2M communications," IEEE International Conference on Software, Telecommunications and Computer Networks (SoftCOM), 2012, pp.1-7. Software, Telecommunications and Computer Networks (SoftCOM), 2012, pp.1-7.
- [6] G. Klas, F. Rodermund, Z. Shelby, S. Akhouri, J. Höller, "Lightweight M2M: Enabling Device Management and Applications for the Internet of Things," 2014, Available at: <http://archive.ericsson.net/service/internet/picov/get?DocNo=1/28701-FGB101973>.
- [7] S. Datta, C. Bonnet, "Smart M2M Gateway Based Architecture for M2M Device and Endpoint Management," IEEE International Conference on Internet of Things (iThings), and Green Computing and Communications (GreenCom), and Cyber, Physical and Social Computing (CPSCom), 2014, pp.61-68.
- [8] A.A. Corici, R. Shrestha, G. Carella, A. Elmangoush, R. Steinke, T. Magedanz. "A solution for provisioning reliable M2M infrastructures using SDN and device management," International Conference on Information and Communication Technology (ICoICT), 2015, pp.81-86.
- [9] ETSI TS 102 690 "Machine-to-Machine communications (M2M); Functional architecture," v1.1.1, 2011.
- [10] Open Mobile Alliance, "Enabler Test Specification for Lightweight M2M Candidate Version 1.0 – 03 Feb 2015," OMA-ETS-LightweightM2M-V1\_0-20150203-C
- [11] L. Fuchun, Z. Qiansheng, C. Xuesong, "Bisimilarity control of decentralized nondeterministic discrete-event systems," International Control Conference CCC, 2014, pp.3898-3903



# Autonomous Agent Model for Connectivity Management in M2M Communications

Ivaylo Atanasov<sup>1</sup>, Anastas Nikolov<sup>1</sup> and Evelina Pencheva<sup>1</sup>

**Abstract** – The large amount of diverse connected devices which interact with the environment without human intervention and the variety of Machine-to-Machine (M2M) applications is a real challenge for network operators. Embedding autonomic features in M2M system may reduce the operational cost. In this paper, we present a model of autonomous agent for connectivity management of M2M devices. The agent observes connectivity parameters of the device and based on operator policies determines the best bearer that has to be used by the device. Temporal logic is used to formalize the agent behaviour.

**Keywords** – Remote device management, Network bearer control, Autonomic behaviour, Temporal logic

## I. INTRODUCTION

The ubiquitous penetration of Machine-to-Machine (M2M) communications in different application area, the increased amount of equipment and the explosion of M2M services become business and technical challenges for network operators [1]. The reduction of network management complexity can be achieved by embedding autonomic features in operation support systems [2]. The autonomic system exposes reactive or proactive behavior based on external stimuli, following goals that are required to fulfill, policies, capabilities, principles of operation, experience and knowledge.

Currently, there is a lot of work conducted on autonomics by the research community. In [3], the authors discuss challenges and enablers that allow connected machines to evolve and act in a more autonomous way and propose architectural approach based on situational knowledge acquisition and analysis techniques in order to make machines aware of conditions and events affecting systems behavior. In [4], the authors propose a middleware architecture that connects the appropriate devices and applications, and is based on software agents representing devices and applications negotiating between each other on the terms by which the data can be used. In [5], the authors propose network architecture for remote monitoring and surveillance M2M networks with broadband satellite connection. In [6], it is proposed a flexible multi-agent approach, leveraging semantic-based resource discovery and orchestration for home and building automation applications. In [7], a generic architecture for multi-goal, adaptable and open autonomic systems, exemplified via the development of a concrete autonomic application for the smart micro-grid is proposed.

<sup>1</sup>The authors are the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mails: iia@tu-sofia.bg, nikolov.anastas@gmail.com; enp@tu-sofia.bg

While presenting high level architectural aspects of autonomous systems, these works discuss proprietary solutions and do not consider autonomics in generic M2M communications.

In this paper, we present a model of autonomous agent for connectivity management in M2M communications. The goal of the agent is to optimize the device connectivity parameters by choosing the best bearer that device should use, which is based on device location and device operator policy. First we describe the logic behind the autonomous connectivity management. Next, an overview on the knowledge base in the domain is presented. The conclusion summarizes the contribution.

## II. M2M CONNECTIVITY MANAGEMENT AUTONOMOUS AGENT

An agent is a thing that perceives from and acts on an M2M device in such way that the device goes through a sequence of states maximizing the performance measures. The problem in M2M self-optimization includes a goal and set of means to achieve the goal. One of the goals of M2M self-optimization is the usage of the best network bearer. An agent dedicated to M2M self-optimization (Connectivity Management Agent) reasons about and follows actions in order to achieve the goal. The process of reasoning what means it can do is called search. The Connectivity Management Agent is goal-based and solves the problem deciding what to do by finding sequences of actions that lead to the desirable operational state of M2M devices with cellular or wireless connectivity. Further we consider M2M devices with cellular or wireless connectivity. The agent actions can be viewed as transitions between M2M device states.

The problem solving of an M2M Connectivity Management Agent includes four stages: goal formulation, problem formulation, searching solution and execution. The Agent explores the current situation and draws the goal which helps to organize behavior by rejecting actions that result in a failure to achieve the operational state of the M2M device. The Agent draws the problem by deciding what transitions and states to consider following the operational state of M2M device. In general, an M2M Connectivity Management Agent faces with several options of possible sequences of actions because it does not know enough about the current device state. For example, there may be different reasons for device not answering (a connectivity problem, low battery level, a firmware failure, etc.). The Agent searches the solution space by examining different sequences of action. Once the solution is found, the agent carries out the identified actions in the execution stage.

The autonomous behavior of the proposed Connectivity Management Agent is based on the OMA Lightweight M2M

(LWM2M) [8]. LWM2M is a standard for a communication mechanism between M2M devices and M2M server platforms. The main motivation behind it is to reduce the degree of fragmentation in the area of M2M remote management.

The LWM2M defines a protocol between M2M devices in a role of clients and the server. The server is responsible for device registration and deregistration, firmware updates, device rebooting, monitoring connectivity parameters on the device, controlling bearers used by the device, configuration of APN and retrieval of device location data [9].

The Connectivity Management Agent in a role of LWM2M server is responsible for device registration, observation of device connectivity parameters and selection of best bearer for the device. The agent functionality for device registration is simplified due to space restrictions and to the facts that the focus is on bearer control.

We assume that the device operator has determined preferred bearers for both specific and normal areas. Each device supports traps, which means that the device is capable of monitoring the event and send notifications whenever it detects the event. The Connectivity Management Agent has to register for the capability in order to use it. OMA DiagMon Trap Events specification defines a number of standardized traps [10]. Geographic trap goes to active when a device enters into a specific geographic area. Whenever the device leaves that specific geographic area, the trap goes to inactive. Collection of measurement data (e.g. signal strength and quality) is important for the connectivity management when the device communicates over the wireless network. Received power trap can help the connectivity optimization process by triggering the Connectivity Management Agent when the received power of the device drops below the server-specific value. Whenever a device's received power drops below an agent-specified value (TrapActivePower), it causes this trap to go active. Alternatively, when device senses power rises above another agent-specified value (TrapInactivePower), it causes this trap to go inactive. In cases that the trap goes active or inactive, the device notifies the registered agent. The device can have several instances of this kind of trap to monitor various network types (e.g. WiFi, WCDMA, LTE etc). At any time the agent can query the device about its location and about the connectivity parameters, i.e. parameters include the used network bearer, available network bearers, signal strength as well as network identities.

Usually, the goal of Connectivity Management Agent is that the managed device to use the best networks bearer. One of the policies of choosing the best network bearer is as follows. If the device is in the specified area and the signal strength of the preferred bearer in area is higher than the specified value of TrapActivePower, then the best bearer is the preferred one for this area. When the device is out of the specified area and the signal strength of the preferred bearer out of area is higher than the specified value of TrapActivePower, then the best bearer is the preferred out of area one. If the signal strength of the preferred bearer is lower than the specified value of TrapActivePower, then the best bearer is the available bearer with highest signal strength.

The logic behind the Connectivity Management Agent behavior might be described as a temporal sequence. On successful device registration, the agent configures geographic traps and received power traps. The agent queries the device about its location and about connectivity parameters. Based on the location, the signal strength of the used network bearer and available bearers, and the best bearer policy the agent performs a bearer selection procedure for the device. After selecting the best bearer the device is in operational state. During this state, the device may send notifications about traps in case the respective event occurs and the agent performs the bearer selection procedure.

### III. KNOWLEDGE BASE MODEL FOR DEVICE CONNECTIVITY MANAGEMENT

We use predicates to express the facts, to show the exchange of messages between the client and server, and to describe the device states as seen by the agent.

Excellent( $b, x$ ) becomes true when the received signal strength of bearer  $b$  by the device  $x$  is higher than the specified value of TrapInactivePower. Good( $b, x$ ) becomes true when the received signal strength of bearer  $b$  by the device  $x$  is between the specified values of TrapActivePower and TrapInactivePower. If the device senses signal strength of  $b$  below the specified value of TrapActivePower then the Bad( $b, x$ ) gets true. In case the device  $x$  uses bearer  $b$  then Used( $b, x$ ) is true. InArea( $a, x$ ) is true when the device  $x$  is in the area  $a$ . Predicates PreferredIn( $b, a$ ) and PreferredOut( $b, a$ ) are true when bearer  $b$  is preferred bearer in area  $a$  and out of area  $a$  respectively. The express the fact that bearer  $b$  is available for device  $x$  the Available( $b, x$ ) is used. When there are no available bearers for  $x$  except the used one then AvailableEmpty( $x$ ) is true. BadPreferred( $x$ ) is true the received signal strength of preferred bearer by the device  $x$  is bad. Best( $b, x$ ) is true if the received signal strength of  $b$  is the maximal one for device  $x$ . PowerTrapActive( $x, b$ ) gets true when the power trap goes active and the signal strength of used bearer  $b$  by device  $x$  becomes bad. PowerTrapInactive( $x, b$ ) gets true when the power trap goes inactive and the signal strength of used bearer  $b$  by device  $x$  becomes excellent.

The behavior of the Connectivity Management Agent is described by temporal logic. We use a minimal set of standard notations  $G$  for always,  $U$  for until, and  $N$  for next.

The agent considers the following statement when explores the current device state, formulates the problem, searches the solution and performs actions.

The device  $x$  is unregistered until a registration request is received:

$$G(\text{Unregistered}(x) \rightarrow \top U \text{reg}_{\text{req}}(x)) \quad (1)$$

After successful device registration, the server configures geo trap and power trap.

If the device  $x$  is unregistered and a registration request is received then a registration response is sent, and a request for geo trap configuration is sent, and the state becomes WaitGeoAck:

$$G(\text{Unregistered}(x) \wedge \text{reg}_{\text{req}}(x) \rightarrow \text{reg}_{\text{res}}(x) \wedge \neg \text{BadPreferred}(x) \wedge \text{configGeoTrap}_{\text{res}}(x) \wedge N \text{WaitGeoAck}(x)) \quad (2)$$

The device  $x$  is in WaitGeoAck state until the server receives a response of geo trap configuration:

$$\mathcal{G}(\text{WaitGeoAck}(x) \rightarrow \top \mathcal{U} \text{configGeoTrap}_{\text{res}}(x)) \quad (3)$$

If the state is WaitGeoAck and a response of geo trap configuration is received then a request for configuration of power trap is sent and the state becomes WaitPowerAck:

$$\mathcal{G}(\text{WaitGeoAck}(x) \wedge \text{configGeoTrap}_{\text{res}}(x) \rightarrow \text{configPowerTrap}_{\text{req}}(x) \wedge \mathcal{N}\text{WaitPowerAck}(x)) \quad (4)$$

The device  $x$  is in WaitPowerAck state until the server receives a response of power trap configuration:

$$\mathcal{G}(\text{WaitPowerAck}(x) \rightarrow \top \mathcal{U} \text{configPowerTrap}_{\text{res}}(x)) \quad (5)$$

After successful configuration of geo and power traps the server requests the device location and the device connectivity parameters.

$$\mathcal{G}(\text{WaitPowerAck}(x) \wedge \text{configPowerTrap}_{\text{res}}(x) \rightarrow \text{getLocation}_{\text{req}}(x) \wedge \mathcal{N}\text{WaitLocation}(x)) \quad (6)$$

The device  $x$  is in WaitLocation state until a location response is received.

$$\mathcal{G}(\text{WaitLocation}(x) \rightarrow \top \mathcal{U} \text{getLocation}_{\text{res}}(x)) \quad (7)$$

The location response will allow the server to determine whether the device  $x$  is in area  $a$ :

$$\mathcal{G}(\text{WaitLocation}(x) \wedge \text{getLocation}_{\text{res}}(x) \rightarrow \text{connParameters}_{\text{req}}(x) \wedge \mathcal{N}\text{WaitConnectivity}(x)) \quad (8)$$

The device  $x$  is in WaitConnectivity state until a connectivity parameters response is received. The connectivity parameters response will contain the signal strength of used bearer  $b$  by device  $x$  and available bearers for device  $x$ .

$$\mathcal{G}(\text{WaitConnectivity}(x) \rightarrow \top \mathcal{U} \text{connParameters}_{\text{res}}(x)) \quad (9)$$

Equations from (10) to (13) refer to bearer selection procedure when the device  $x$  is in area  $a$ .

When the used bearer of  $x$  is  $b$ , and  $b$  is the preferred bearer in area  $a$ , and the signal strength of  $b$  is excellent or good, then the state becomes Operational:

$$\mathcal{G}(\text{WaitConnectivity}(x) \wedge \text{connParameters}_{\text{res}}(x) \wedge \text{InArea}(a, x) \wedge \text{PreferredIn}(b, a) \wedge \text{Used}(b, x) \wedge (\text{Excellent}(b, x) \vee \text{Good}(b, x)) \rightarrow \mathcal{N}\text{Operational}(x)) \quad (10)$$

When the used bearer  $b$  of  $x$  is the preferred one, and the signal strength of  $b$  is bad, and  $c$  is available bearer for device  $x$  and  $c$  is the best bearer then a request to select bearer  $c$  is sent, and the state becomes WaitBearerAck:

$$\mathcal{G}(\text{WaitConnectivity}(x) \wedge \text{connParameters}_{\text{res}}(x) \wedge \text{InArea}(a, x) \wedge \text{PreferredIn}(b, a) \wedge \text{Used}(b, x) \wedge \text{Bad}(b, x) \wedge \text{Available}(c, x) \wedge \text{Best}(c, a) \rightarrow \text{select}_{\text{req}}(x, c) \wedge \text{BadPreferred}(x) \wedge \mathcal{N}\text{WaitBearerAck}(x)) \quad (11)$$

The bearer  $c$  selection procedure takes place when  $c$  is available and preferred bearer, and the received signal strength of  $c$  by device  $x$  is not bad:

$$\mathcal{G}(\text{WaitConnectivity}(x) \wedge \text{connParameters}_{\text{res}}(x) \wedge \text{InArea}(a, x) \wedge \text{Used}(b, x) \wedge \neg \text{PreferredIn}(b, a) \wedge \text{Available}(c, x) \wedge \text{PreferredIn}(c, a) \wedge \neg \text{BadPreferred}(x) \rightarrow \text{select}_{\text{req}}(x, c) \wedge \mathcal{N}\text{WaitBearerAck}(x)) \quad (12)$$

When the used bearer of  $x$  is not the preferred one, and  $c$  is available and preferred bearer in area  $a$ , and the received signal strength of  $c$  by device  $x$  is bad, and  $d$  is the best

available bearer then the server initiates bearer  $d$  selection procedure:

$$\mathcal{G}(\text{WaitConnectivity}(x) \wedge \text{connParameters}_{\text{res}}(x) \wedge \text{InArea}(a, x) \wedge \text{Used}(b, x) \wedge \neg \text{PreferredIn}(b, a) \wedge \text{Available}(c, x) \wedge \text{PreferredIn}(c, a) \wedge \text{BadPreferred}(x) \wedge \text{Best}(d, a) \wedge \text{Available}(d, x) \rightarrow \text{select}_{\text{req}}(x, d) \wedge \mathcal{N}\text{WaitBearerAck}(x)) \quad (13)$$

Equations from (14) to (17) refer to bearer selection procedure when the device  $x$  is out of area  $a$ .

When the used bearer  $b$  of  $x$  is the preferred one, and the signal strength of  $b$  is excellent or good, then the state becomes Operational:

$$\mathcal{G}(\text{WaitConnectivity}(x) \wedge \text{connParameters}_{\text{res}}(x) \wedge \neg \text{InArea}(a, x) \wedge \text{PreferredOut}(b, a) \wedge \text{Used}(b, x) \wedge (\text{Excellent}(b, x) \vee \text{Good}(b, x)) \rightarrow \mathcal{N}\text{Operational}(x)) \quad (14)$$

When the used bearer  $b$  of  $x$  is the preferred one, and the signal strength of  $b$  is bad, and  $c$  is the best available bearer then a bearer  $c$  selection procedure take place:

$$\mathcal{G}(\text{WaitConnectivity}(x) \wedge \text{connParameters}_{\text{res}}(x) \wedge \neg \text{InArea}(a, x) \wedge \text{PreferredOut}(b, a) \wedge \text{Used}(b, x) \wedge \text{Bad}(b, x) \wedge \text{Available}(c, x) \wedge \text{Best}(c, a) \rightarrow \text{select}_{\text{req}}(x, c) \wedge \text{BadPreferred}(x) \wedge \mathcal{N}\text{WaitBearerAck}(x)) \quad (15)$$

In case the used bearer  $b$  of  $x$  is not the preferred one, and  $c$  is available preferred bearer, and the received signal strength of  $c$  by device  $x$  is not bad, then the server request selection of bearer  $c$ :

$$\mathcal{G}(\text{WaitConnectivity}(x) \wedge \text{connParameters}_{\text{res}}(x) \wedge \neg \text{InArea}(a, x) \wedge \text{Used}(b, x) \wedge \neg \text{PreferredOut}(b, a) \wedge \text{Available}(c, x) \wedge \text{PreferredIn}(c, a) \wedge \neg \text{BadPreferred}(x) \rightarrow \text{select}_{\text{req}}(x, c) \wedge \mathcal{N}\text{WaitBearerAck}(x)) \quad (16)$$

When the signal strength for the preferred bearer  $c$  is bad and  $d$  is the best available bearer then the server request selection of bearer  $d$ :

$$\mathcal{G}(\text{WaitConnectivity}(x) \wedge \text{connParameters}_{\text{res}}(x) \wedge \neg \text{InArea}(a, x) \wedge \text{Used}(b, x) \wedge \neg \text{PreferredOut}(b, a) \wedge \text{Available}(c, x) \wedge \text{PreferredIn}(c, a) \wedge \text{BadPreferred}(x) \wedge \text{Best}(d, a) \wedge \text{Available}(d, x) \rightarrow \text{select}_{\text{req}}(x, d) \wedge \mathcal{N}\text{WaitBearerAck}(x)) \quad (17)$$

When there are no available bearers for device  $x$ , and the received signal strength of  $b$  by device  $x$  is good the state becomes Operational:

$$\mathcal{G}(\text{WaitConnectivity}(x) \wedge \text{connParameters}_{\text{res}}(x) \wedge \text{Used}(b, x) \wedge (\text{Excellent}(b, x) \vee \text{Good}(b, x)) \wedge \text{AvailableEmpty}(x) \rightarrow \mathcal{N}\text{Operational}(x)) \quad (18)$$

The server considers the device  $x$  unregistered when the received signal strength of the used bearer  $b$  is bad and there are no available bearers:

$$\mathcal{G}(\text{WaitConnectivity}(x) \wedge \text{connParameters}_{\text{res}}(x) \wedge \text{Used}(b, x) \wedge \text{Bad}(b, x) \wedge \text{AvailableEmpty}(x) \rightarrow \mathcal{N}\text{Unregistered}(x)) \quad (19)$$

The device  $x$  is in WaitBearerAck state until the server receives a response of bearer selection procedure:

$$\mathcal{G}(\text{WaitBearerAck}(x) \rightarrow \top \mathcal{U} \text{selectBearer}_{\text{req}}(x, b)) \quad (20)$$

When a bearer selection response is received, the server waits for device re-registration:



$$G(\text{WaitBearerAck}(x) \wedge \text{selectBearers}_{\text{res}}(x) \rightarrow \neg \text{WaitReregistration}(x)) \quad (21)$$

The device  $x$  is in WaitReregistration state until the server receives a registration request:

$$G(\text{WaitReregistration}(x) \rightarrow \neg \text{Ureg}_{\text{req}}(x)) \quad (22)$$

Upon successful device re-registration the server request device location:

$$G(\text{WaitReregistration}(x) \wedge \text{reg}_{\text{req}}(x) \rightarrow \text{reg}_{\text{res}}(x) \wedge \neg \text{BadPreferred}(x) \wedge \text{getLocation}_{\text{req}}(x) \wedge \neg \text{WaitLocation}(x)) \quad (23)$$

The device  $x$  is in Operational state until the server receives a notification about signal strength change of used bearer by  $x$  or a notification about change of  $x$  location:

$$G(\text{Operational}(x) \rightarrow \neg \text{U}(\text{notifyPower}_{\text{req}}(x) \vee \text{notifyGeo}_{\text{req}}(x))) \quad (24)$$

In Operational state, when a notification about location change of device  $x$  is received, the server sends a response of geo trap notification and requests device connectivity parameters:

$$G(\text{Operational}(x) \wedge \text{notifyGeo}_{\text{req}}(x) \rightarrow \text{notifyGeo}_{\text{res}}(x) \wedge \text{connParameters}_{\text{req}}(x) \wedge \neg \text{WaitConnectivity}(x)) \quad (25)$$

When the power trap becomes inactive in Operational state, the server sends a response of power trap notification and the state remains Operational:

$$G(\text{Operational}(x) \wedge \text{notifyPower}_{\text{req}}(x, b) \wedge \text{PowerTrapInactive}(x) \rightarrow \text{notifyPower}_{\text{res}}(x, b) \wedge \neg \text{Operational}(x)) \quad (26)$$

Activation of power trap in Operational state means that the signal strength becomes bad and the server sets the margin timer:

$$G(\text{Operational}(x) \wedge \text{notifyPower}_{\text{req}}(x, b) \wedge \text{PowerTrapActive}(x) \rightarrow \text{notifyPower}_{\text{res}}(x, b) \wedge \text{setTmargin}(x) \wedge \neg \text{WaitMargin}(x)) \quad (27)$$

The device  $x$  is in WaitMargin state until the server receives a notification about signal strength change of used bearer by  $x$  or a notification about change of  $x$  location:

$$G(\text{WaitMargin}(x) \rightarrow \neg \text{U}(\text{notifyPower}_{\text{req}}(x, b) \vee \text{notifyGeo}_{\text{req}}(x) \vee \text{Tmargin}(x))) \quad (28)$$

Notification that the power trap is inactive in WaitMargin state means that the signal strength becomes excellent and the server and resets the margin timer:

$$G(\text{WaitMargin}(x) \wedge \text{notifyPower}_{\text{req}}(x, b) \wedge \text{PowerTrapInactive}(x) \rightarrow \text{notifyPower}_{\text{res}}(x, b) \wedge \text{resetTmargin}(x) \wedge \neg \text{Operational}(x)) \quad (29)$$

In WaitMargin, when the margin timer expires, the server sends requests device connectivity parameters:

$$G(\text{WaitMargin}(x) \wedge \text{Tmargin}(x) \rightarrow \text{connParameters}_{\text{req}}(x) \wedge \neg \text{WaitConnectivity}(x)) \quad (30)$$

The device may change its location while it is in WaitMargin state:

$$G(\text{WaitMargin}(x) \wedge \text{notifyGeo}_{\text{req}}(x) \rightarrow \text{notifyGeo}_{\text{res}}(x) \wedge \neg \text{WaitMargin}(x)) \quad (31)$$

When the device is unregistered due to connectivity problems, it may wait for some time and try to register again.

For brevity sake the behavioral model of the Connectivity Management Agent is given without parts like re-registration

which is caused by registration timer expiry, APN activation due to change of network bearer etc.

#### IV. CONCLUSION

Embedding autonomic features in M2M systems reduces operational costs. Following the goal of ensuring a reliable and efficient M2M network, autonomous agents that perceive from the M2M ecosystem and act upon it can contribute to self-optimization.

We present an algorithm for choosing the best bearer and model the behavior of Connectivity Management Agent which selects radio technologies in wireless environment according to current radio conditions. The agent monitors the connectivity parameters on M2M devices and based on the device location and policies defined by the device operator, selects the best network bearer to be used by the device. The model is formally described by temporal logic. Our future work will include more details related to APN configuration, device rebooting and firmware update.

#### REFERENCES

- [1] P. Watson, K. C. Gupta, "EM-ANN Models for Microstrip Vias and Interconnects", IEEE Trans., Microwave Theory Tech., vol. 44, no. 12, pp. 2395-2503, 1996.
- [2] C. Pereira, A. Aguiar, "Towards Efficient Mobile M2M Communications: Survey and Open Challenges," Sensors 2014, no. 14, 19582-19608; doi:10.3390/s141019582, pp.19582-19608.
- [3] J. Sachs, N. Beijar, P. Elmdahl, J. Melen, F. Militano, P. Salmela, "Capillary networks – a smart way to get things connected," Ericsson Review, 2014, no. 8, pp. 2-8.
- [4] D. Kyriazisa, T. Varvarigoua, "Smart, autonomous and reliable Internet of Things," International Workshop on Communications and Sensor Networks (ComSense'2013), International Workshop on Communications and Sensor Networks, ComSense'2013, Procedia Computer Science, 2013, pp. 442 – 448.
- [5] S. Vassaki, G. Pitsiladis, C. Kourogorgas, M. Poulakis, A. Panagopoulos, G. Gardikis, S. Costicoglou, "Satellite-based sensor networks: M2M sensor communications and connectivity analysis", International Conference on Telecommunications and Multimedia (TEMU), Greece, 2014, pp.132–137.
- [6] K. Misura, M. Zagar, "Internet of things cloud mediator platform," International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2014, pp.1052-1056
- [7] M. Ruta, F. Scioscia, G. Loseto, E. Di Sciascio, "Semantic-Based Resource Discovery and Orchestration in Home and Building Automation: A Multi-Agent Approach," IEEE Transactions on Industrial Informatics, vol. 10, no. 1, 2014, pp.730-741.
- [8] S. Frey, A. Diaconescu, D. Menga, I. Demeure, "Towards a generic architecture and methodology for multi-goal, highly-distributed and dynamic autonomic systems," International Conference on Autonomic Computing (ICAC), 2013, pp.201-212.
- [9] G. Klas, F. Rodermund, Z. Shelby, A. Akhouri, J. Holler, "Lightweight M2M": Enabling Device Management and Applications for the Internet of Things," white paper, 2014.
- [10] Open Mobile Alliance, Enabler Test Specification for Lightweight M2M, 2015, OMA-ETS-LightweightM2M-V1\_0-20150203-C.
- [11] Open Mobile Alliance, Diagnostics and Monitoring Trap Events Specifications, 2013, OMA-TS-DiagMonTrapEvents-V1\_2-20131008-A



# Integrated HPC Data-log Monitoring System

Kostadin Mishev<sup>1</sup>, Sonja Filiposka<sup>2</sup>, and Anastas Mishev<sup>3</sup>

**Abstract** – Monitoring HPC systems is crucial for the increase of their efficient usage. Integrated monitoring, based on data from various monitoring sources gives a broader picture and enables in depth understanding of the systems, reducing their power consumption and developing of new, highly effective HPC software.

**Keywords** – HPC, monitoring, efficient usage.

## I. INTRODUCTION

High Performance Computers (HPC) represent one of the the main building blocks of the todays e-Infrastructures [1]. Together with cloud-based solutions and efficient distributed large scale storage on top of ultrafast agile network, they are a crucial element that provides the processing power needed to tackle with big data problems.

However, achieving their optimal usage comes with a whole plethora of new problems and challenges. With the most important resource being processing power achieved with a huge number of cores (CPUs and accelerator cores like GPGPUs and Xeon Phi's), the goal of maximum performance needs to be reached under a complex set of constraints like efficient core usage, minimum inter process communication, but also minimized power consumption. Thus, the effective usage of these computing giants needs thorough monitoring that will reveal the usage of the available resources and the way they respond to different loads and tasks [2].

Most of the current monitoring solutions are vendor dependant proprietary solutions like IBM Platform HPC [3], HP Insight Cluster Management Utility [4], Fujitsu HPC Cluster Suite [5]. They provide various types of monitoring of HPC systems, mostly offering chassis and operating system (OS) level analysis and visualization tools.

Notable open source monitoring solutions include Ganglia [6], Nagios [7], and OVIS [8]. These tools are usually based on probes running on the hosts themselves and different ways of collecting and representing the collected data.

Important questions that arise regarding the HPC monitoring [9] include the understanding of the HPC monitoring process, its evolution, technologies and vendor

lock-ins vs neutrality. Vendor specific proprietary tools usually offer higher levels of details, while the open source solutions give more flexibility.

There are several perspectives of the HPC monitoring that should be taken into account when implementing a monitoring solution: OS level monitoring, chassis level monitoring, power consumption monitoring etc. The integration of different monitoring data sources gives a more complete and wider perspective [10]. Using integrated data, such as host or chassis level power consumption combined with the CPU and memory usage can help in developing power aware algorithms, schedulers etc. One of the main challenges when integrating monitored data from different sources is the time based association of the data that is gathered using different sampling rates, i.e. monitoring data with different resolution.

HPC monitoring systems can be implemented as real-time monitors or log analysis tools [11]. Although real time monitoring provides valuable data for the current state of the HPC cluster, monitoring based on log analysis offers more in-depth view of the overall system parameters and can lead to important conclusions regarding the correlation of the values of measured parameters obtained from different sources and levels.

In this paper we present a HPC data-log monitoring system acting as a collection server for monitoring data that stems from different sources. Upon collection, the system adapts the data into one common format, aligning the different sampling resolutions of the available monitored data. The summarized output of the system can be used for various purposes such as performance analysis, power consumption examination etc.

The rest of the paper is organized as follows. Part II describes the sources of data aggregated by the proposed system. Part III explains the system architecture, while Part IV discusses the possible applications of the system. Part V concludes the paper and points toward the future work.

## II. MONITORING SOURCES

HPC monitoring data can be integrated from various sources that will provided information concerning different aspects of the system like node level monitoring, enclosures parameters, and out-of-band system management interfaces.

In order to obtain monitoring data from the OS perspective, a cluster management software (CMS) [12] can be used. Upon deployment, each computing node runs an agent of the cluster management utility that sends data to the appointed management node. Usually, the utility solution provides a tool for real time monitoring and visualization of the collected data. This data includes: CPU load and frequency, memory usage, OS version, disk space usage, etc. In order to obtain logged data for a specific time period, a command line tool can be used. This tool enables the user to select a subset of nodes and the parameters that will be exported to a file with a flat structure (timestamp, nodeID, parameter, value). This

<sup>1</sup>Kostadin Mishev is with the Faculty of Computer Science and Engineering at Ss. Cyril and Methodius University in Skopje, Rudjer Boshkovikj 16, Skopje 1000, Macedonia, E-mail: [kostadin.mishev@finki.ukim.mk](mailto:kostadin.mishev@finki.ukim.mk).

<sup>2</sup>Sonja Filiposka is with the Faculty of Computer Science and Engineering at Ss. Cyril and Methodius University in Skopje, Rudjer Boshkovikj 16, Skopje 1000, Macedonia, E-mail: [sonja.filiposka@finki.ukim.mk](mailto:sonja.filiposka@finki.ukim.mk).

<sup>3</sup>Anastas Mishev is with the Faculty of Computer Science and Engineering at Ss. Cyril and Methodius University in Skopje, Rudjer Boshkovikj 16, Skopje 1000, Macedonia, E-mail: [anastas.mishev@finki.ukim.mk](mailto:anastas.mishev@finki.ukim.mk).

dump can then be combined and aligned with the other collected datasets.

Another common monitoring data source is the so-called Light-out management system (LOM) [13]. LOM uses a dedicated management channel to monitor and remotely manage nodes regardless of whether they are powered on, or if an operating system is installed or functional, i.e. compared to the cluster management utility, it provides an external monitoring view of the nodes and their enclosures. The data acquired from this system includes chassis information, power usage and distribution by chassis elements, aggregated thermal monitoring data, etc.

Although the LOM solutions are specialized implementations of the Intelligent Platform Management Interface (IPMI) [14], the direct use of IPMI enables further enrichment of the gathered data. IPMI is in fact a specification for interfaces used for out-of-band management of computer systems and monitoring of their operation. Using IPMI one can gather more detailed data including current power consumption, various temperature sensor readings and fan speed on node level, compared to the LOM's chassis level. Unlike CMUs that logs the monitoring data autonomously, using IPMI the data must be manually pulled in real-time with a given resolution. This can be achieved using the *ipmitool* command.

### III. SYSTEM OVERVIEW

In order to obtain a holistic, yet, comprehensive, overview of the monitored system, the data gathered from the three available monitoring sources described in the previous section must be composed in a single collection. For this reasons, we propose an integrated data-logging monitoring system that will act as a collection server for the monitoring sources and thus provide a single point of reference when monitoring the usage of the HPC resources and related components.

In this section we describe the proposed system architecture and its implementation for a given HPC.

#### A. System architecture

Data gathered from the sources described above, are merged and saved into a database as a raw data. The captured data series are organized in reservations enabling the start and the end of data capturing. The data captured by the LOM tool is the basis for establishing the time flow and data resolution. On top of this data, the captured measurements from the other two sources are added. All data imported in the monitoring software is automatically merged and aligned using the provided timestamps.

The complete raw data is organized in different reservations characterized with reservation start and end time. Inside each reservation, the raw data that belongs to the reservation is aligned according to the LOM's data resolution. Each reservation can be visualized using a chart panel, or it can be downloaded as a set of flat files for external evaluation.

The deployed example system for integrated HPC monitoring, given on Fig. 1, is implemented using Java Spring 4.0 framework using MySQL database for data logging.

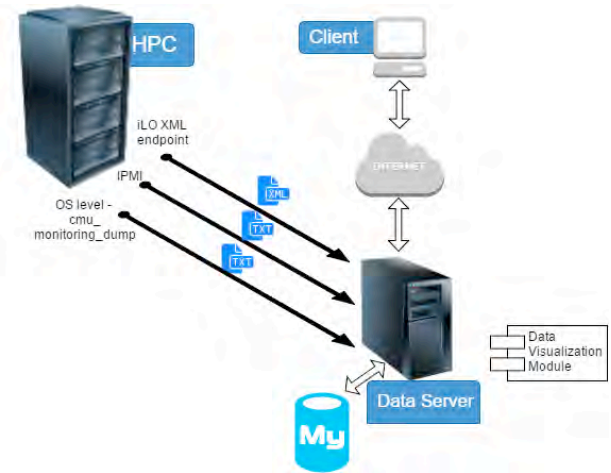


Fig. 1: System architecture

#### B. Use-case scenario

The described system is used to monitor a Hewlett-Packard HPC system located at the Faculty of Computer Science and Engineering in Skopje. The system is based on HP Cluster Platform 7000 enclosures with 42 blades BL2x220c G7 with dual motherboard and dual Intel Xeon L5640 @ 2.26Ghz (total 1008 cores). The blades are grouped in three enclosures.

Using HP's Insight CMU [4] each blade can have its own monitoring station which measures several parameters such as temperature, CPU frequency, CPU load, memory used and process memory, which determine the current state of the blade depending of the current load handled by the appropriate blade. The data from the CMU is gathered using the *cmu\_monitoring\_dump* CLI tool.

The Light-out management system used for monitoring data is based on the HP's implementation of the LOM, the HP integrated Lights Out (iLO) [15]. iLO uses a custom build scripting language, based on XML, to provide detailed information about the HPC enclosure and its elements. Each iLO-based enclosure serves as an endpoint that returns XML files presenting the values of the sensors. The enclosure which wraps a set of blades, has its own control unit. The control unit monitors the overall system health:

- the rotations per minute of cooling fans
- the enclosure manager provides information about hardware state, firmware versions and temperature of monitoring unit
- temperature of switches
- power consumption and power output
- ambient temperature

Additionally, using IPMI fine-grain monitoring data is gathered on the node level, specifically node power consumption and node level CPU temperature.

More detailed information on the various available measurement parameters that can be obtained using the merged data log monitoring approach is given in Table I that can be found at the end of this paper.

Within the developed solution (available at [16]), each monitoring session is presented as a reservation. In order to

conserve storage, the system is not running all the time, but is acting upon scheduled user created reservations. To enable the log during specified time period, the user must create a new reservation and specify the time interval of the active session together with the enclosures that are to be monitored. Upon creating the new reservation, using the main page, the user can monitor the current status of the reservation presented via a green slider (see Fig. 2).

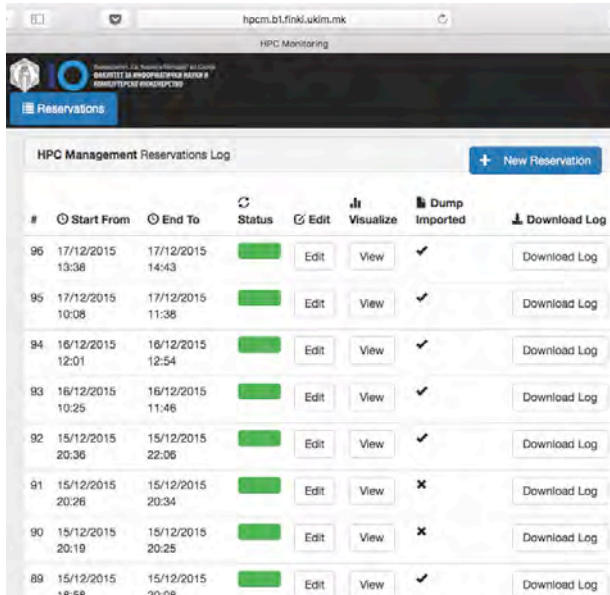


Fig. 2: Monitoring reservations

Afterwards, the user can import the dump files obtained from the other sources by navigating to the Edit section of the reservation. This page assists the user to execute the appropriate commands on the native interfaces to obtain the right dump file (see Fig. 3). The file should be uploaded by using the file upload option on this page. The system calls background data merge services which matches the appropriate data parameters from different sources. The product of such merge, can be observed by navigating to the Visualize section by clicking the View button.

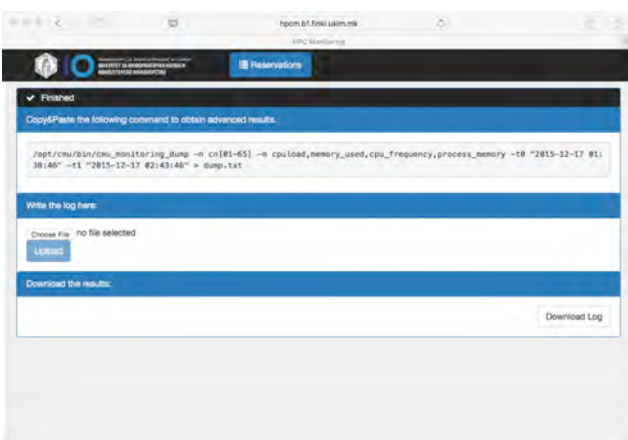


Fig. 3: Import of the data from other sources

On the navigation menu, the user chooses the parameters that he will observe. The result is a chart series which visualize the parameters changes over time (see Fig. 4). Also, there exists an option for downloading a zip file of a set of flat files for all parameters by reservation for additional processing.

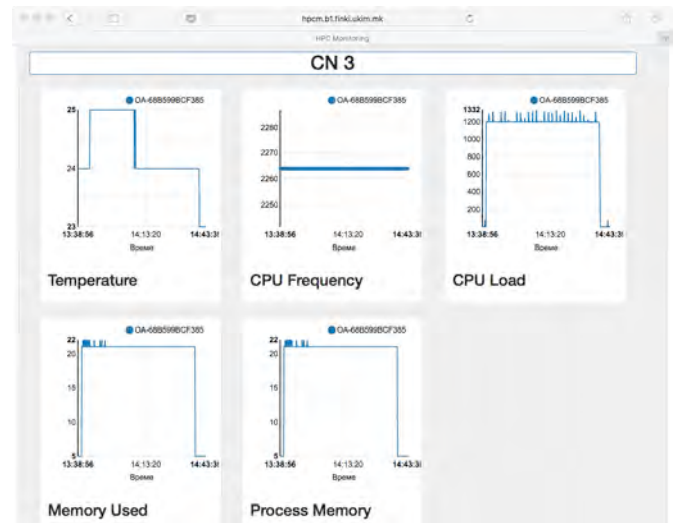


Fig. 4 Visualization of the data

## IV. APPLICATIONS

The proposed system offers a wide spectre of applications. Analysis and visualization of the execution of specific applications can help the developers optimize their solutions toward better resource utilization.

If we take power efficiency as an example, the collected data can be used for investigating power aware algorithms or schedulers, providing optimal power consumption.

Thermal related data (temperature sensors measurements, fan rotation speeds) can help avoid hot spots in the chassis by varying the scheduling and the distribution of the processes by nodes.

Analysis of the data from the networking elements can help develop intelligent parallel algorithms that will optimize interposes communication.

## V. CONCLUSION

In this paper we presented an integrated monitoring solution for HPC systems. It is based on data collections from various sources including server management interfaces and OS level probes. The integrated data can be visualized or exported for further in-depth analysis. The acquired and integrated datasets can be used for better system understanding, development of novel HPC schedulers, increased resource usage under power restrictions, etc.

Future versions of the system will be aimed towards enabling finer resolution of the data sampling, real time import and alignment and wider integration and scalability.

## ACKNOWLEDGEMENT

This work is supported by the project Design and analysis of green ICT systems, financed by the Faculty of computer science and engineering, UKIM.

## REFERENCES

- [1] N. Mustafee, and S. JE Taylor. High-performance simulation and simulation methodologies. 2013.
- [2] J.C Browne, R.L. DeLeon, A.K. Patra, W.L. Barth, J. Hammond, M.D. Jones, T.R. Furlani, B.I. Schneider, S.M. Gallo, A. Ghadersohi, and R.J. Gentner. Comprehensive, open-source resource usage measurement and analysis for HPC systems. *Concurrency and Computation: Practice and Experience*, 26(13), pp.2191-2209, 2014.
- [3] IBM Platform HPC, <http://www-03.ibm.com/systems/platformcomputing/products/hpc/>, visited March 2016.
- [4] HP Insight Cluster Management Utility, <http://www8.hp.com/us/en/products/server-software/product-detail.html?oid=3296361>, visited March 2016.
- [5] Fujitsu Cluster Suite, <http://www.fujitsu.com/fts/microsites/hpc/products-services/index.html>, visited March 2016.
- [6] Ganglia Monitoring System, <http://ganglia.info>, visited March 2016
- [7] Nagios, <https://www.nagios.org>, visited March 2016.
- [8] J. M. Brandt, A. C. Gentile, D. J. Hale and P. P. Pebay, "OVIS: a tool for intelligent, real-time monitoring of computational clusters," *Parallel and Distributed Processing Symposium, 2006. IPDPS 2006. 20th International*, Rhodes Island, 2006.
- [9] W. Allcock, E. Felix, M. Lowe, R. Rheinheimer, and J. Fullop. Challenges of HPC monitoring. In *State of the Practice Reports* (SC '11). ACM, New York, NY, USA, 2011.
- [10] S. Bohm, C. Engelmann and S. L. Scott, "Aggregation of Real-Time System Monitoring Data for Analyzing Large-Scale Parallel and Distributed Computing Environments," *High Performance Computing and Communications (HPCC), 2010 12th IEEE International Conference on*, Melbourne, VIC, pp. 72-78, 2010.
- [11] C. L. Moore, et al. Monitoring High Performance Computing Systems for the End User. *Cluster Computing (CLUSTER), 2015 IEEE International Conference on*. IEEE, 2015.
- [12] N. Bonvin, S. Vaudenay, and P. Junod. Cluster Management Software. No. LSIR-STUDENT-2009-001. 2003.
- [13] H. Bojinov, E. Bursztein, E. Lovett, D. Boneh. *Embedded Management Interfaces: Emerging Massive Insecurity*. Black Hat USA 2009.
- [14] IPMI – Intelligent Platform Management Interface, <http://www.intel.com/content/www/us/en/servers/ipmi/ipmi-home.html>, visited March 2016.
- [15] HPE Integrated Lights Out (iLo), <http://www8.hp.com/us/en/products/servers/ilo/index.html>, visited March 2016.
- [16] Integrated HPC monitoring, <http://hpcm.b1.finki.ukim.mk/>.

TABLE I  
SELECTED MEASURED PARAMETERS BY SOURCE

Parameter	Applies to	CMU	iLO	IPMI
Temperature	Node		X	
CPU frequency	Node	X		
CPU load	Node	X		
Memory used	Node	X		
Process memory	Node	X		
Current fan RPM	Chassis		X	
Output power	Chassis		X	
Input power	Chassis		X	
Power supplies	Chassis		X	
Ambient temp	Chassis		X	
Power meter	Node			X
CPU temp	Node			X
Inlet ambient temp	Node			X
DIMM temp	Node			X
Memory zone temp	Node			X
Chipset temp	Node			X
NIC temp	Node			X
System exhaust temp	Node			X
Virtual fan	Node			X
Enclosure status	Node			X
Memory status	Node			X
Health LED	Node			X
Per switch temp	Chassis		X	
Manager temp	Chassis		X	
Power on switch	Chassis		X	
Power off switch	Chassis		X	



# E-content Development and Task Solution Using the Content Management System of Online-CADCOM

Galia Marinova<sup>1</sup> and Ognyan Chikov<sup>2</sup>

**Abstract** – The paper presents different types of e-content in Online-CADCOM and the methodology steps for its development, using Content Management System. The methodology is illustrated by adding e-content for a new online tool WEBENCH Clock Architect (TI). Later a design and verification task is solved using the new tool from the CMS and the simulator ORCAD Design Suit for verification.

**Keywords** – E-content, Content Management System, E-content development, Design and verification task solution, PLL loop filter design.

## I. INTRODUCTION

Online-CADCOM [1,2] content is dynamically generated from the Content Management System (CMS) [4], once the content is successfully developed and uploaded through the user interface. The platform has unique and major capabilities which turn Online-CADCOM in powerful web platform that has the following features through the CMS platform:

- Automated templates – new content is automatically applied to the existing templates and views, as also all content can be changed from a central place.
- Access control – user groups allow Online-CADCOM to be controlled on different hierarchical levels and they determine how registered users interact with the CMS system
- Scalable expansion of content – the CMS platform offers ability to expand content in future work as adding more files, new categories and etc.
- Easily editable content – the content is separated from the visual presentation of the web platform and it is very easy to edit and manipulate by non-technical users, which is the main focus and idea of using CMS approach.
- Workflow – technique used for uploading a content in the platform which gives abilities for workflow management and multiuser parallel tasks.
- Collaboration – the content of the platform is retrieved and uploaded by one or many authorized users. Changes can be made from all authorized users, and this can be considered as advanced form of collaboration that allows multiple users to make changes of a content material at the

same time, from different geographical locations.

- Delegation – user levels of access provide limited access for different user groups over specific sections of platform, which divide into small part the responsibility for the management.
- Content Virtualization – CMS platform allows each user to work with a virtual copy of the whole web site, documents and resources included. This feature enables changes to multiple resources to be viewed prior to submission.
- Multilingual – with the developing of the CMS, the multilingual functionality is embedded. This gives the ability to display the content in different languages, not only the native.

The paper is organized as follows: Types of e-content and methodology steps for e-content development, using the CMS menus are described firstly. The process is illustrated for the integration of a new tool WEBENCH Clock Architect (WBCA) [7]. Then task solution using the platform is illustrated through an example of design and verification of PLL (Phase-locked loop) loop filter using WBCA tool and ORCAD/PSpice simulator.

## II. E-CONTENT AND METHODOLOGY STEPS FOR ITS DEVELOPMENT IN ONLINE-CADCOM

### A. E-content in Online-CADCOM

E-content development using CMS platform is an intuitive process, thanks to the proper scalability of the platform as described in [3]. The content of Online-CADCOM would not be limited only to online tools. The platform is built by different types of e-content with different specifications and functions. In CMS, the approach used to develop and create e-content is using methods for CRUD (Create, Read, Update, Delete). These main methods are implemented in buttons that are visible in the bottom of each section of platform – “Add”, “Show”, “Edit” and “Delete”. Moreover, the content is highly protected using a login screen for the platform and also different user-type levels as described in [3] with different levels of access and platform rights. Development of e-content can be conditionally divided into:

- Categories,
- Tools,
- Portals,
- Standards,
- Knowledge Base,
- Estimation tools,
- Prototype tools,
- Optimization tools.

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

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

Categories in the platform are related with Panel 1 and 2, which cover the Daniel-Gajsky Y model used in Online-CADCOM and represent the content related with the categories of this model. These categories can contain both subcategories and online tools with individual tool passports. Adding a category does not affect tools and vice versa. This means that the two sections are separate and independent and the content is not connected, but separated in the categories.

In CMS, the tools section gives the opportunity to develop content related with different online tools. A tool passport is used as described in [3] for identification of the tool before integration in the platform. Here, different tools are added, edited and deleted. During the process of development the content provides an option to choose the related category or subcategory for the tool which is added.

The section with portals gives the possibility of adding different type of engineering portals that can contain useful and important information about all users that will use Online-CADCOM. The information is not limited to text, but also to images and video files, which make this section an extremely useful for quick and reliable information and future links to useful products that are out of Online-CADCOM scope.

"Standards" is another section in the platform that provide the users with information about different standards in telecommunications and system design. Information here is not limited to name, date and title of the standard but a standard can also be downloaded from the platform using the button "Download". The platform is developed in a way that content can be up to one document per title and the document formats are limited to \*.doc, \*.docx and \*.pdf. Limitation is needed for storage optimization and also virus, worms and adware protection of the E-content.

The Knowledge Base contains a large number of materials in different categories. Visitors of Online-CADCOM can view and download knowledge materials such as how-to documents, videos and another knowledge material.

"Estimation tool" section of CMS allows the administrators to upload estimation tools in the platform as the logic is the same as the tools in the Y model of Panel 1 and Panel 2, but without subcategories. This limitation is done because of the used method of showing items from this menu – using dynamic number of pages according to the number of Estimation tools. Here the same approach for tools classification, using tool passports is covered.

The section "Outcome to prototype" is developed for e-content related with tools that can be used for prototype engineering work.

Optimization menu is developed for tools that are used for optimization task and solutions.

The CMS is developed to facilitate the work of Online-CADCOM administrators and to make content development and upload easy and fast task, without need of programming skills. The scope of using the platform is tasks solving related with the e-content in Online-CADCOM. Some of the options of CMS are:

- New Category/Subcategory creation through CMS;
- Adding a new tool with passport in a specific Category;
- Developing/Adding e-content in the Knowledge Base.

### *B. Methodology steps for e-content development in Online-CADCOM*

The main methodology steps for e-content development in Online-CADCOM, using the CMS are: Adding a new category/subcategory; Adding a new tool in a category; Passport filling for a new tool; Development and adding e-content in the Knowledge Base to help task solution with a set of tools.

These steps are considered in details and they are illustrated through the example of adding the new Online tool WBCA, which is a module in the Online Platform WEBENCH Design Center of TI [8], in a new subcategory "PLLs". The new tool permits design, simulation and optimization of clock tree solutions for given input and output frequencies. The main circuit used is the PLL. Details about PLL design and important building blocks as Voltage-controlled oscillator (VCO) and PLL loop filter can be found in [6]. More about datasheet specification reading of a PLL device is presented in [5].

#### *Adding new subcategory*

As described in [3], categories in Online-CADCOM are related with Panel 1 and Panel 2 of the front page. More categories or subcategories can be added through the Content Management System and they automatically stand behind the main categories that build up the Panel 1 and Panel 2. This approach is used because of Daniel-Gajsky Y model and tools classification according to it. Category creation is easy and can be made from the platform menu – "Categories". After opening this menu, the main screen shows all the current categories that are created in the system. These categories correspond to the names of the buttons of Panel 1 and Panel 2. As described above, a category can contain a subcategory. For creation of a new category, the button "Add" must be pressed. Categories can be deleted using button "Delete".

For the needs of the new online tool, a new subcategory "PLLs" has to be created in the existing Category "RF and Microwave".

#### *Adding a new online tool*

Tools can be added to the platform using the section – Tools. First when this button is clicked, a window with all tools is shown on the screen. A new tool can be added using the "Add" button. The process is the same as for adding a new category/subcategory, but here the user must fulfill the form that represents the tool passport with all information related to the specific tool. The development of the e-content for the online tool content must be consistent with the categories. If there is no category, no tools can be added. Also when a tool is added, a selection form with the category must be chosen, otherwise the platform generates an error. Tools can be easily removed using the "Delete" button.

Figure 1 illustrates the creation of the e-content for the new online tool WBCA. The creation of the content for the new tool begins with the name of the tool and an image symbolizing the tool. The image chosen to symbolize the WBCA is a clock tree block diagram. The web address of the new online tools is added in order to perform "Launch"

function for accessing and running the new tool. Figure 2 illustrates the vision of the new tool on Panel 1, subcategory “PLLs” of the Category “RF and Microwave”.

#### Passport filling for a new tool

Figure 3 illustrates the parameter filling in the passport of the tool WBCA and Figure 4 shows the e-content of the passport filled, visualized by Online-CADCOM. The new tool is a module in the online platform WEBENCH Design Center of TI. It's acting at the transistor level from the Daniel Gajsky Y-model. The application area is defined as PLL design and simulation, the functions are selected from a predefined set, then input/output connections with other existing tools are enumerated and ORCAD/PSPICE is pointed as a verification tool. Some quantitative and qualitative features are added in the last 2 fields of the passport. The input/output connections show possibilities for combining the tool WBCA with other tools for solving complex design tasks.

#### Developing and adding e-content in the Knowledge Base

The Knowledge Base contains different type of materials and files, which help engineers to solve complicated tasks. Knowledge Base materials can be of any kind of files – documents, pictures and etc. Uploading a Knowledge Base material is an automated task that requires only to fulfil the forms of Knowledge Base section and to point a file to be uploaded. A new knowledge material can be created with the button “Add”. Then a window with forms for creation of the material is shown. After completion of filling forms, with the button “Save”, the new knowledge material is added to the database and is ready to be visualized in Online-CADCOM. Instructions for how to combine tools in Online-CADCOM, in order to solve complex design and verification tasks, are part of the e-content in the Knowledge base. A new instruction for verification of the PLL loop filter designed with WBCA, using the simulator ORCAD/PSPICE is uploaded to the Knowledge Base.

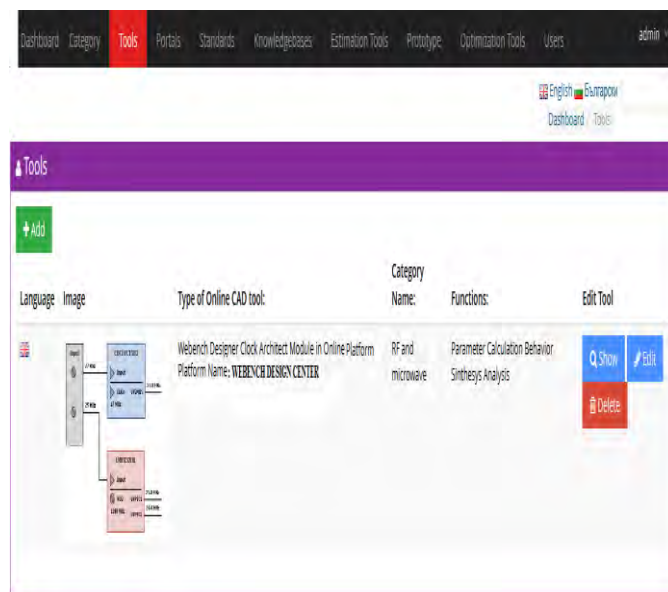


Fig.1. Adding New tool WEBENCH Clock Architect in the category “RF and Microwave”, subcategory “PLLs”

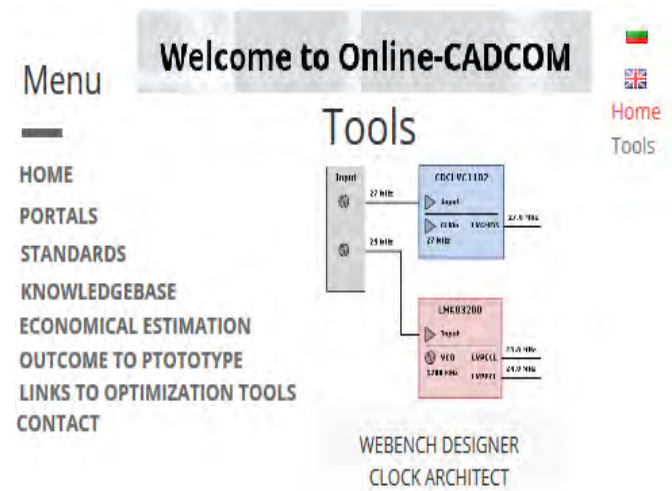


Fig.2. Visualization of the tool WEBENCH Clock Architect and its image on Panel 1

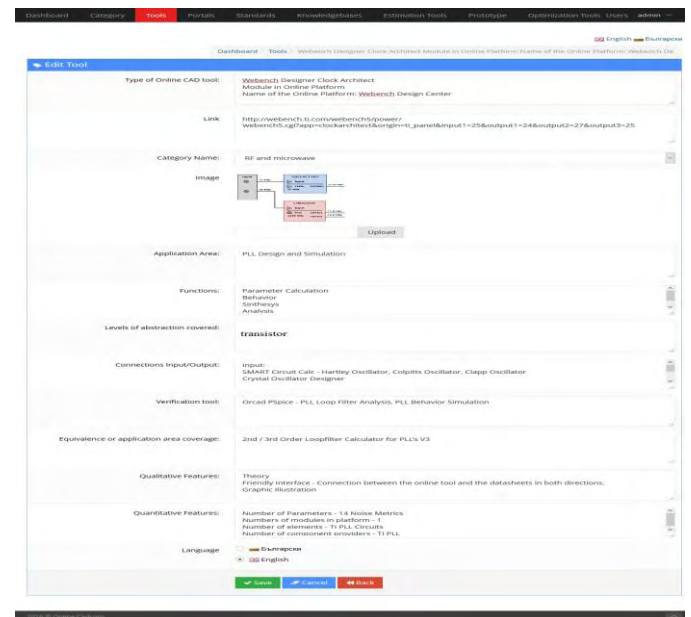


Fig.3. Filling e-content for the passport of WEBENCH Clock Architect



Fig.4. Visualization of the passport filled in Online-CADCOM



### III. DESIGN TASK SOLUTION IN ONLINE-CADCOM

Once the WBCA tool is integrated in Online-CADCOM, it is used for design verification combining the online tool with Orcad Design Suit 16.6 program [9] and PSpice simulator, following the instructions given in the Knowledge Base.

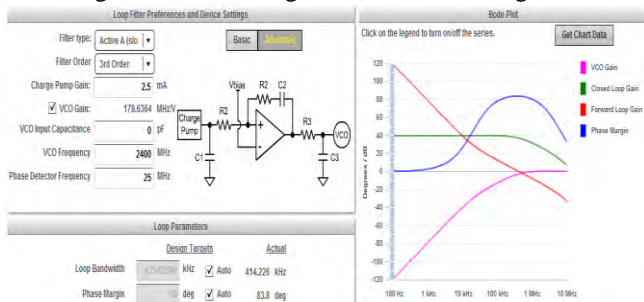


Fig.5. Loop filter design circuit, parameters and simulation results in WEBENCH Clock Architect

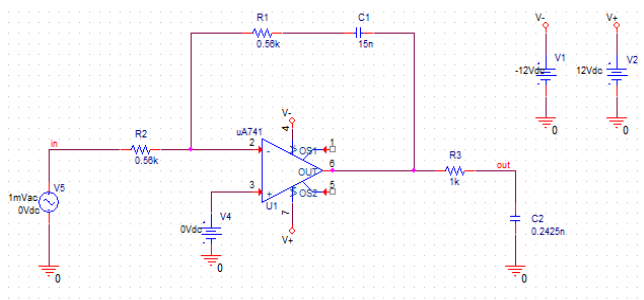


Fig.6. Circuit of the loop filter in ORCAD/Capture 16.6

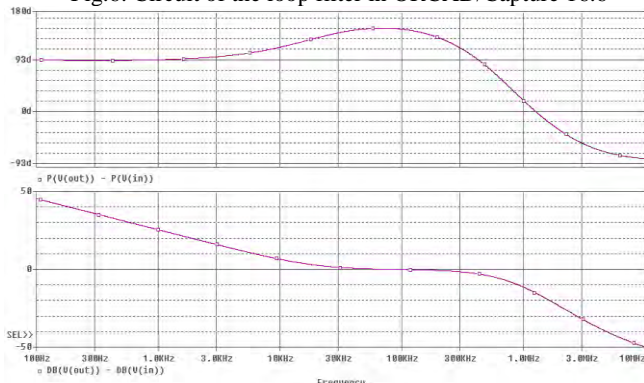


Fig.7. Gain and Phase simulation curves of the loop filter from Fig.6 simulated in ORCAD/PSpice 16.6

First a clock tree with specification for an input frequency of 25 MHz and 3 output frequencies of 24 MHz, 27 MHz, 24 MHz is designed. One of the 10 proposed solutions is considered. Then the **LoopFilter>Advanced** options are selected. The circuit selected is Active A (low slew rate) 3rd Order. The filter circuit and the curves simulated in WBCA are presented on Figure 5. The parameters of the loop filter obtained by WBCA are: Bandwidth – 414 kHz and Phase Margin – 83.8°.

Then the circuit of the loop filter is drawn in ORCAD Capture as shown on Figure 6. The circuit is with negative feedback which follows the theory in [6] and gives meaningful simulation results. The filter from Figure 6 is simulated in frequency area (AC) and the curves for the Gain and the Phase are presented on Figure 7. The loop filter parameters

determined from the simulated curves are Bandwidth – 428 kHz and Phase Margin – 92.5°. The curves from Fig.7 are similar in form as the curves of Forward loop gain and Phase Margin from Fig.5, but there are some differences in the values of the parameters obtained. It would be good to know the motivation for using the positive feedback circuit in WBCA on the other hand, it's worth to verify the designs with ORCAD/PSpice and to search the most realistic solution, as basis for further optimization of the clock tree design.

### IV. CONCLUSION

Using the CMS for e-content development and for task solution is a job that does not require special programming skills. E-content can be easily created with help of few sections and buttons. The platform is intuitive and easy to operate with. Content can be changed dynamically and anytime. The ease-of use is illustrated by adding e-content for the online tool WBCA to Online CADCOM. A process of adding e-content for hundreds of online tools is foreseen in near future. The results from the design and verification task show that such a study is needed in order to compare results and to fix better parameters obtained and to optimize designs.

### ACKNOWLEDGEMENT

This research is partly supported by Project №152PD0039-07 in Technical University – Sofia.

### REFERENCES

- [1] G. Marinova, V. Guliashki, O. Chikov, "Concept of Online Assisted Platform for Technologies and Management in Communications – OPTIMEK", ICCSIST 2014, 7-9 November 2014, Durres, Albania, 2014, pp.55-62
- [2] O. Chikov, G. Marinova, MVC Framework Approach in the Online assisted platform form Computer-aided Design in Communications – Online-CADCOM, Computer and Communications Engineering, TUS, 2016.
- [3] G. Marinova, O. Chikov, Methodology for tools integration in the Online Assisted Platform for Computer-aided design in communications, Proc. of papers of ICEST'2015, 24-26 June 2015, Sofia, pp.31-36, ISBN:978-619-167-182-3
- [4] O. Chikov, "Software development of Content Management System in Online-CADCOM platform", submitted to ICEST'2016, Ohrid, Macedonia
- [5] S. Damphousse, How to estimate the phase noise of a PLL with basic datasheet specifications, Texas Instruments, March 2, 2016, [https://e2e.ti.com/blogs\\_/b/analogwire/archive/2016/03/02/understanding-datasheet-phase-noise-specifications-of-a-phase-locked-loop?DCMP=lmx25x2&HQS=sva-sds-fcp-lmx25x2-exah-null-null-eu](https://e2e.ti.com/blogs_/b/analogwire/archive/2016/03/02/understanding-datasheet-phase-noise-specifications-of-a-phase-locked-loop?DCMP=lmx25x2&HQS=sva-sds-fcp-lmx25x2-exah-null-null-eu)
- [6] D. Banerjee, PLL Performance Simulation and Design Handbook - 4th Edition, 13 October 2006
- [7] WEBENCH Clock Architect, TI <http://www.ti.com/lscs/ti/analog/webench/clock-architect.page>
- [8] WEBENCH Design Center, TI, [http://www.ti.com/lscs/ti/analog/webench/overview.page?DCMP=sva\\_webench&HQS=webench-btechday](http://www.ti.com/lscs/ti/analog/webench/overview.page?DCMP=sva_webench&HQS=webench-btechday)
- [9] Orcad Design Suit 16.6, Users' Guide, Cadence, 2014



# SCADA system for monitoring and control of small industrial producing unit

Hristo Nenov<sup>1</sup>

**Abstract** – The topic of this paper is design and development of SCADA system for remote autonomous production units (Remote Automation Unit - RAU) from a "small farm" type. For the purpose of developing the material base of the farm is reduced to a limited number of controlled functional units which could vary depending on the particular implementation on a real production site.

**Keywords** – SCADA, MODBUS, Java, JPA, PLC.

## I. INTRODUCTION

In modern economies it is almost impossible to imagine industries where computers and machines have not have taken control of a significant part of the production processes. The main reason behind this is the need to increase the competitiveness of production in the sector. The formula for achieving this goal is to increase the production quantity and reduce the production costs. The implementation of different tools for mechanization of the production creates an opportunity for that. Agriculture is no exception and modern technologies are becoming more widely used in this industry. The combination of the latest information technologies, global information network Internet, more available control devices and the use of renewable energy sources make the appearance of modern agriculture - autonomous, more efficient and predictable.

## II. ARCHITECTURE OF THE SYSTEM

### A. General scheme

Remote Automation Unit (RAU) from "small farm" type consist the following modules:

1. Greenhouses with the following adjacent manageable systems and sensors:
  - Drip irrigation system with controllable valve;
  - Sprinkling system with controllable valve;
  - Ventilation;
  - Measuring devices for air temperature;
  - Measurement devices for humidity.
2. Wells - local inexhaustible sources with the following adjacent manageable systems and sensors:
  - Pumping systems;

- Sensors for water level.

3. Water storage reservoir - capacitive buffer source, which is fed with water from local inexhaustible sources and supplies water for the entire farm, the following adjacent manageable systems and sensors:

- Pumping systems for water;
- Water level sensor.

For the purpose of real testing, simplified configuration is selected and it is deployed as follows:

- Greenhouse with an individual drip irrigation system, sprinkling system, ventilation system, temperature sensor and an air sensor for humidity;
- Water storage supplying the greenhouse, with a pumping system for water supply and one sensor for water level;
- Two wells, each with a pumping system connected to the water supply and water retaining with one sensor for water level

### B. Physical layer

Physical layer - covers the management and transfer of data to and from various sensors and controllable devices (sensors, valves, relays). This layer is designed to the specific realities of production sites and goals.

### B. Industrial Interface layer

The layer is implemented on one or more programmable logic controllers (PLC) and includes logic for gathering and initial processing of data to and from physical sensors and controllable devices and providing them to the layers of higher-level system. PLC interface has its own logic, which is connected to the other layers of the system by appropriate industrial protocols.

### C. Application layer

The layer implements the application management logic. It consist two aspects:

- Management of autonomous remote objects (small farms);
- Business logic implemented on the Central System Proxy Server.

### D. Presentation layer

The layer implements Human Machine Interface (HMI) for developing SCADA systems. This is **layer of the highest level** and provides a graphical interface for system operators

<sup>1</sup>Hristo Nenov is with the Faculty of Computer Science and Automation at Technical University - Varna, 1 "Studentska" str., Varna 9000, Bulgaria, E-mail: h.nenov@tu-varna.bg.

through which they can monitor the working condition, to read different measurements, to oversee occurred events, alarms and archival data

#### E. Data layer

The layer implements mechanisms for temporary and permanent storage of system data. It covers two aspects:

- Temporary storage of data within the RAU (individual small farms)
- Permanent storage of data source RAU on a main server.

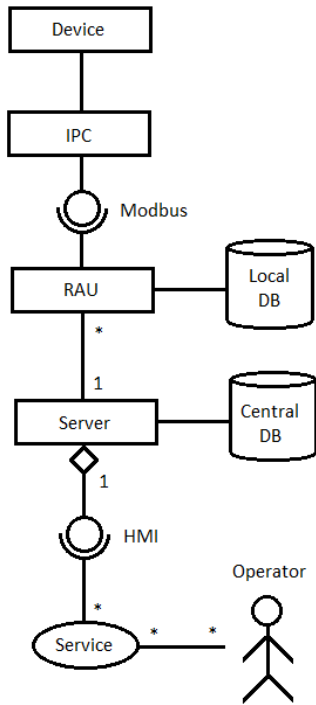


Fig.1. General architecture of the system.

### III. REALIZATION OF THE SYSTEM

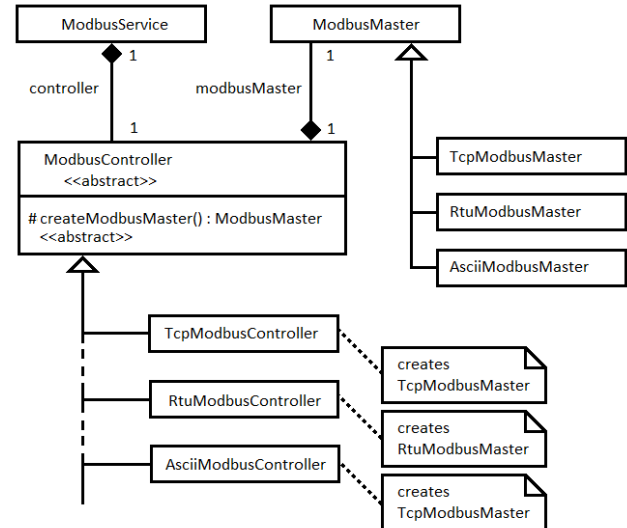
Each Remote Automation Unit (RAU) is designed to operate in unreliable network connectivity to the central server, but it should provide reliable and trouble-free autonomous management of the production process based on the current configuration. In implementation ScheduledExecutor Service (Java 1.7+) is used to start services at regular configurable intervals:

- Activities of gathering information from the sensory part of the system, primary processing and storage of information in a local database
- Logic of autonomous management of physical devices (pumps, valves, relays, etc.) based on a current configuration to ensure the needs of production
- Generate and store events and alarms;
- Send the collected data, events and alarms to the central server.

Data collection (Snapshots) and their replication to a central server are performed by the SnapshotProducerService and SnapshotConsumerService, working in parallel on RAU and implementing the principle Producer / Consumer.

The exchange service for MODBUS contains abstract controller class Modbus Controller, which implements operations via MODBUS composite instance of ModbusMaster - base library class in modbus4j. The heirs of ModbusController implement different MODBUS modes - TCP, RTU, ASCII, as appropriate instantiation type - successor to ModbusMaster.

Fig.2. Service and Controller for data exchange over MODBUS.



Updating the status of RAU (Remote Automation Unit). The internal state of RAU is managed by the class DeviceManager, which maintains a collection of devices – “Device”, which represent the inputs and outputs of the PLC. Synchronization of values with those of the registers of the PLC, is carried out by ModbusService, which uses class ModbusController, encapsulating work with MODBUS. ModbusService and DeviceManager sharing data for the state through class Snapshot as the exchange is initiated only from ModbusService.

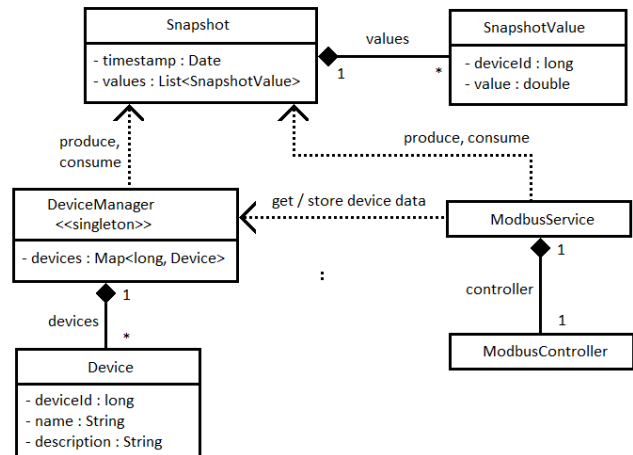


Fig.3. Updating of state of RAU.

Model of automatic control of RAU.

Automatic control is realized by class AutomationController, which applies logical chains of rules. Each type of rule implements specific logic. This mechanism realizes design pattern “Chain of responsibility” and relies on polymorphism technique.

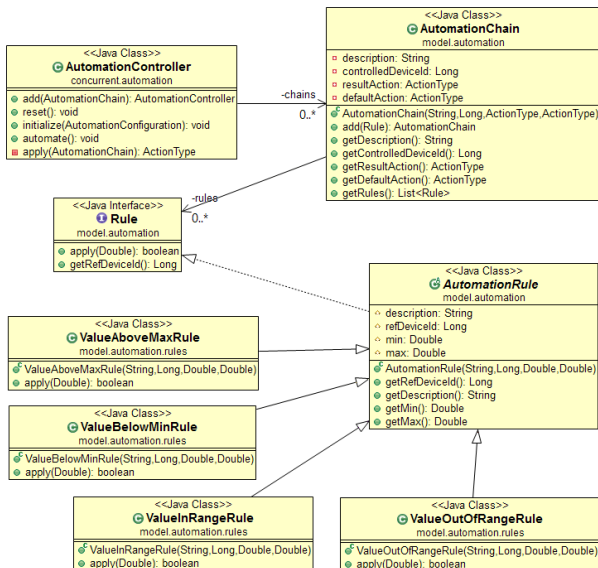


Fig.4. Model of automatic control of RAU

#### IV. TESTING OF THE SYSTEM

For testing purposes of the developed SCADA systems, experimental arrangement that reflects the structure of a small autonomous farm RAU and a central server are realized. The experimental setting is made up of the following components:

- Central server on which the application runs RauServer. Implemented on a separate computer that is running Apache Tomcat with tuned Jersey Servlet, REST service requests from RAU. On the same machine instance is running for MySQL database on which stored data are received from RAU;
- RAU (Remote Automation Unit), which was launched as a standalone application (service) on a separate computer. On the same machine is running instance of MySQL database on which the data is temporarily stored until it is sent to them to the central RauServer;
- MODBUS Controller.

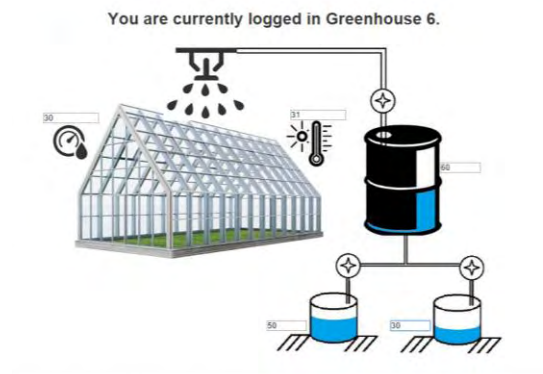


Fig.5. Monitoring of sensors and devices in real time.



You are currently logged in Greenhouse 6.

Date From:

2015-05-01

Date To:

2015-05-02

Load

Id	Greenhouse Id	Description	Timestamp
2	6	Tank Pump on	May 1, 2015 9:55:56 AM
8	6	Well 1 Pump on	May 1, 2015 9:44:00 AM
9	6	Well 1 Pump off	May 1, 2015 11:10:23 AM
16	6	Pump on	May 1, 2015 11:00:00 AM
21	6	Tank Pump off	May 1, 2015 9:55:56 AM

Fig.6. Monitoring of events in real time.

#### V. CONCLUSION

The system is fully operational. All modules are working properly, generating and exchanging correct data. The main objective of the study is achieved. There are a few aspects, which are provided for future development.

#### REFERENCES

- [1] David Geary, Cay S. Horstmann “Core JavaServer Faces” 3rd.Edition,May.2010.
- [2] Erich Gamma, Richard Helm, Ralph Johnson, John Vlissides “Gang Of Four - Design Patterns, Elements Of Reusable Object Oriented Software - Addison Wesley
- [3] Kevin Mukhar, Chris Zelenak, James L. Weaver, Jim Crume JavaEE 5, From Novice to Professional
- [4] Leonard Richardson, Sam Ruby –RESTful Web Services 2007.
- [5] www.modbus.org

This Page Intentionally Left Blank



# Project of Crochet Software Development

Elena Zaharieva-Stoyanova<sup>1</sup>

**Abstract** – The paper treats problems related to crochet software development. Such applications are intended for Computer-aided design of handmade textile. Generally, they help people dealing with traditional home crafts. Moreover, these software applications can be used for storing samples of the old knitting and crochet which are ethnographic exhibits. These software systems can be developed with tools or wizards oriented to ethnographic researches. Detailed overview of graphic representation of crochet charts is done. Examples of such charts are created. According to the technology, the paper suggests an approach for crochet software development.

**Keywords** – Crochet software, Knitting software, Computer-aided design, CAD systems.

## I. INTRODUCTION

Computer science and technologies enter in almost all branches in the modern life. They are applied even in branches where until recently the modern technologies were not in use. Such example is a handmaid textile.

Traditional home crafts as weaving, knitting, embroidery, crochet, etc. have a long history. Because they are related to the manufacture of cloths and household articles, they have practiced for centuries all over the world. Nowadays few people continue the tradition of these home crafts, although they are practiced rather as a hobby, than a necessity [3], [14]. However, it has to be mention that the interest to homemade textile provokes another tendency – putting on the market home made knitwear, embroideries, crochet as souvenirs or as customized boutique products, turning home crafts into a business [1], [2], [8].

The design process in which pre-established model of a product, is required whether for manual or machine-made product. There are many software applications for Computer-aided design of home made textiles. Most of existing hand-craft software is oriented towards embroidery and tapestry [13], [14], [15]. There are also applications for Computer-Aided Design of knitwear, so called **knitting software** [6], [7], [11], [12], [17]. Embroidery design software can be used as knitting software, but imposes restrictions on the type of knitted models. Also, knitting software may be used for crochet designing. Desiring to present their product as a multipurpose, companies state theirs application more than one activity [5], [10]. At all, the idea for multipurpose software for Computer-aided design of home made textile is not good. The main reason is that nevertheless the design process in embroidery, knitting and crochet are resemble each

other, each home craft has its own peculiarities and different way of describing the models.

The software applications for Computer-aided design of crochet charts are much less common than those for designing embroidery and knitting. Usually, the companies dealing with knitting software state that products are usable for crochet design [5], [10]. It's valid just only of models, whose patterns can be described in rectangular grid. The most crochet patterns are worked in the round. For this kind of models such software is inapplicable.

The examples of crochet software are Crochet Charts [4] and Filet Crochet Software [9]. First one is oriented to models crocheted in the round; second one is for fillet patterns worked flat. The advantages of Crochet Charts are: the product is tailored to the manufacture of crochet; presents the crochet diagram with crochet instructions, a well-done user interface. The product has some flaws. As the user designs the structure stitch by stitch, so the editing process is complicated. Another disadvantage is that software does not respect the rules of next crochet row coming. This makes the diagram unrealistic. Accordingly, it is possible to create crochet charts, which in practice can not be realized.

This paper represents an approach for crochet software development. The project takes into account both feature of crochet patterns and the way they have been made. The application can be used for designing new models, as well as for preserving and storing samples of old, existing crochet. In this sense, the software can be used for storage of ethnographic exhibits.

## II. FEATURES OF CROCHET PATTERN DESCRIPTIONS

### A. Descriptions of crochet patterns

Crochet patterns are described by crochet charts and/or crochet (work) instructions. Crochet charts include graphic symbols, each of them correspond to a crochet stitch. At all, a crochet chart follows the shape of the real crochet pattern as it's represented in fig. 1. In most cases a legend describing stitches is added to the crochet chart. Figure 2 represents crochet chart with a description of the stitches.

Crochet symbols are type of semi-graphical characters. The most frequently used crochet symbols standardized by Craft Yarn Council are represented in fig. 3 [16]. Usually, a legend describing crochet symbols is added to crochet charts as it is shown in fig. 2.

<sup>1</sup>Elena Zaharieva-Stoyanova is with the Department of Computer Systems and Technologies at Technical University of Gabrovo, 4 H. Dimitar str., Gabrovo 5300, Bulgaria, E-mail: zaharieva@tugab.bg.



Fig. 1. Crochet pattern and its chart

To get accurate information of pattern workmanship, the crochet charts adds work instructions. They give details how to crochet each row. Practically, sometimes operating instructions are an alternative manner to describe a crochet pattern. It should be mentioned though that this does not always work; instructions completely replace the crochet pattern. People dealing with crochet often prefer to hold a crochet chart as a pattern description. A crochet chart follows the shape of the pattern, so it represents how the crochet pattern looks like. Looking at the chart the master can find errors leading to inability to make the model. Using charts skips also the language barrier. If the user does not speak the language of written operating instructions but he is aware with crochet symbols and charts, it is sufficient to understand how to realize the pattern. Work instructions in turn complement the chart and bring any necessary background information. For example, in precise work instructions no need for the user to count the chart how many crochet symbols are in a row. Overall, the full description of a crochet pattern should include both the charts and work instructions. These instructions also include materials used and the hook's gauge.

#### B. Graphic representation of crochet charts

Development of computer aided design software always takes into account the peculiarities both the description of the models and technology of their implementation. In this case, it is necessary to take into account both the characteristics of the

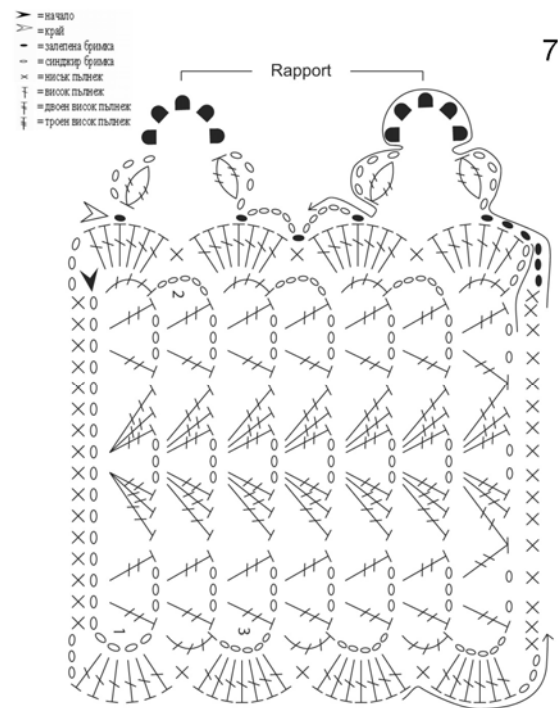


Fig. 2. Crochet chart and legend

graphical representation of the charts and the way that makes the product.

As it has been already mentioned, the graphic description of crochet patterns using diagrams or charts represents their handmade. Unlike knitting patterns where diagrams are described in a rectangular grid, crochet charts could be represented in various shapes. Just a few models can be represented in a rectangular grid as each item corresponds to a stitch. In general, crochet symbols are put on the screen in more open order. Moreover, following the pattern shape each symbol could be rotated from 0 to 360 degrees. These features greatly make the Computer-aided design of the crochet charts more difficult than Computer-aided design of knitting patterns.

Crochet can be worked either **flat** or **in the round**. The model in fig 1 is an example of crochet worked in the round and the model in fig. 2 is an example of flat worked crochet. The second type of crochet charts are resembled to the knitting charts – the stitches described by symbols are put on a rectangular grid. Unlike knitting charts, because there is no precise layout of crochet stitches, the rotation of crochet symbols is possible (see fig 2). For patterns worked in the round, stitches are placed in the circle and crochet symbols follow this layout (see fig. 1). It's possible to modify the circular shape of the pattern to an oval, triangle, hexagon or any other polygon. Crochet symbols are positioned to follow the shape of the pattern.

Both of the kind of patterns (worked in flat or in the round) begins with a **chain** - a number of stitches named **chain**. Practically, the first chain stitches form the base of the crochet. Also, each row begins with one, two or more chain stitches. For patterns worked in the round, the chain stitches are closed with a slip stitch, thus the base is formed as a circle. For patterns worked in flat, the chain base does not close and the crochet continues as knitting forming odd and even rows. There are patterns worked in the round in which the base does not close in a circle. This kind of patterns acquires an oval

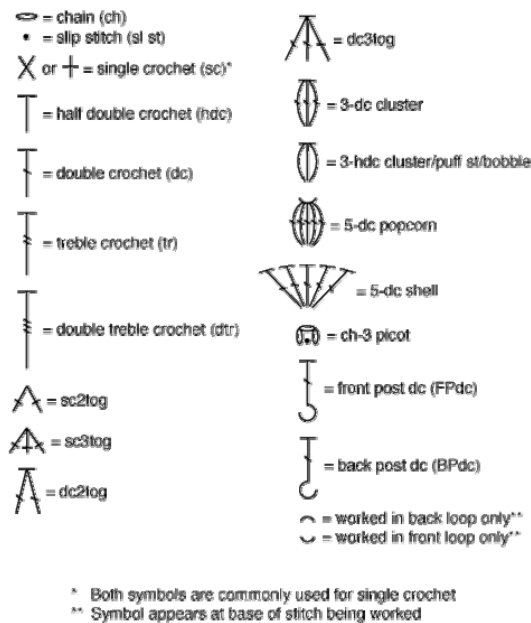


Fig. 3. Crochet symbols

shape.

They are common patterns that comprise a plurality of repeating segments. This technology is named “patchwork” or Irish crochet. Developing crochet software, it must be provided for editing this type charts - the location of the segments and the links between them.

### III. APPROACH TO CROCHET SOFTWARE DEVELOPMENT

In existing solutions [4], [9], crochet software development follows the approach to knitting software development – the charts are edited symbol by symbol as each symbol corresponds to a stitch. The paper suggests another approach – to represent a crochet pattern with their chart and work instructions together. The chart and its operating instructions must be interconnected, so change the chart to reflect to work instructions and vice versa – edit the working instructions to change the chart.

Initially, it’s necessary to determine the working method – crochet in flat or work in the round. The way of crochet significantly affects how will be represented graphically subsequently chart. In the first way may be used a rectangular raster grid - like knitting software development. An example of such crochet char is given in fig. 4. However, it should



Fig. 4. Graphic representation of crochet chat worked in flat

provide that, unlike knitting symbols in crochet charts very often violated the right rectangular order of crochet symbols. In terms of graphics representation, this means that should be allowed each of the symbols in order to be rotated around its own starting point.

In the second way, crochet symbols are arranged in the round in the chart – the shape is a circle or an ellipse. This type of the charts suggests using a polar coordinate system for their graphical representation. An example of such a chart is presented in fig. 5. Provision should be made that possible circular or oval shape can be transformed into a triangular, square, hexagonal or other polygonal shape. It is also possible circular shape be modified in more complex shape, a star, for example, as the chart in fig.2. It’s convenient, according to the type of figure, the user alone to determine the number of axes in the polar grid. Due to the specifics of crochet in this type of chart should be allowed rotation symbols around their starting points.

At all, the description of crochet chart includes its shape and the number its rows. The crochet chart comprises a plurality of rows, each of which contains a certain number of crochet symbols. Respectively can be defined the following drawing objects:

CHART → ROW → SYMBOL

These three really existing objects determine classes needed for a chart description.

There are two approaches to creating a crochet pattern:

- To create a pattern by editing symbol by symbol, to shape lines that form the entire chart.

- Initially create the overall shape of the chart, separate it as rows and set the characters in each line.

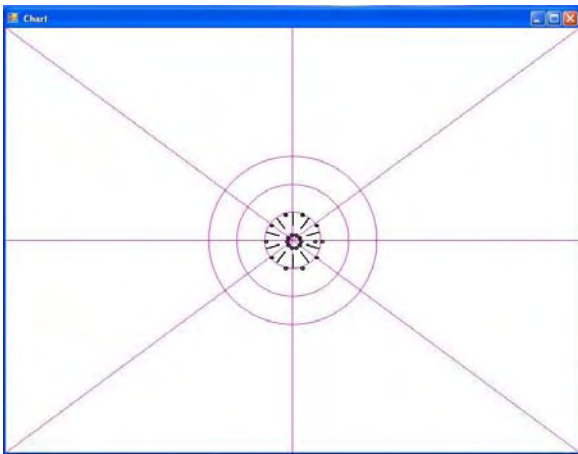


Fig. 5. Graphic representation of crochet chat worked in the round

Both approaches are recommended for well-developed software. It's necessary to declare classes corresponding to the objects chart, row and symbol. Classes must have methods for graphic transformations: translation, rotation, scaling. The chart is defined by its shape – circle, ellipse or polygon. The rows follow the chart shape but each row can be somewhat of changing its size. Each row consists of crochet symbols. Each symbol has type, size, and location.

To avoid possible technical errors in the graphic construction should be laid:

- An initial row includes a set of chain stitches
- Start point for each row marked with a number and required number of chain stitches and slims, if it necessary.

Regardless of the type of chart depending on the working mode, each of crochet symbols can be rotated. It is necessary to provide for a graphical transformation rotation of each symbol.

#### IV. CONCLUSION

Despite the fact that in the modern world, unlike in the past, people are not required to make handmade products, there are some people who deal with this either as a hobby or as a small business. Moreover, traditional crafts and home occupations are also part of the cultural heritage; in practice people who deal with them keep traditions alive. Modern information technology can be successfully used for presentation and preservation of cultural heritage. Development of knitting and crochet software helps the people dealing with these home crafts. Moreover, these software applications can be used for storing samples of the old knitting and crochet which are ethnographic exhibits. These software systems can be developed with tools or wizards oriented to ethnographic researches.

In comparison with knitting and embroidery software, there are fewer crochet software applications. The reason is the

larger difficulties in graphics system development due to the essence of the crochet technology. This paper treats the problems related to crochet software development. Detailed overview of graphic representation of crochet charts is done. Examples of such charts are created. According to the technology, the paper suggests an approach for crochet software development.

#### REFERENCES

- [1] Ani embroidery, <https://anibroderia.alle.bg/>, 2016
- [2] Axioma Hobby Shop, <http://www.axioma-hobby-shop.com/bg/index.html>, 2016.
- [3] Coats, <http://www.coatsbulgaria.bg/>, 2016.
- [4] Crochet Charts 1.2.0, Sept 2015, Stitch Work Software, <http://stitchworkssoftware.com/>, 2016.
- [5] Creative Design ChartMaker specs, PCWorld, <http://www.pcworld.com/product/970803/creative-design-chartmaker.html>, 2016.
- [6] DesignaKnit8, Soft Byte LTD, <http://www.softbyte.co.uk/designaknit.htm>, 2016.
- [7] EnvisioKnit User's Manual, Knitting Pattern Design Software, <http://www.envisioknit.com/>, 2016.
- [8] Faena Hand Knith Fashion, <http://www.faena-fashion.com/bg/>, 2016 r.
- [9] Filet Crochet Software, Sand Castle Designs, <http://www.crochetdesigns.com/software/default.htm>, 2016.
- [10] Intwined Pattern Studio, <http://intwinedstudio.com/>, 2016.
- [11] Knit Visualizer User Manual, The Knit Foundry, 2009.
- [12] Knitting Chart Editor User Manual, Stitchmastery 2014, <http://www.stitchmastery.com/knitting-chart-editor/>, 2016.
- [13] Mac Stitch and Win Stitch, Ursasoftware, <http://www.ursasoftware.com/>, 2016.
- [14] Pattern Maker for cross stitch, Hobby Ware, <http://www.hobbyware.com/>, 2016.
- [15] Stitch & Motif Maker, Software Informer, <http://stitch-motif-maker1.software.informer.com/>, 2016.
- [16] Yarn Standards, Craft Yarn Council, <http://www.craftyarncouncil.com>, 2016.
- [17] Zaharieva-Stoyanova E. , Stefan Bozov, CAD Module for Knitting Patterns Design, Intr. Conf. on Computer Systems and Technologies - CompSysTech'11, 15-16 June 2011, Vienna, Austria.



# Graph Construction Algorithm for finding the Shortest Path in a Maze

Milena Karova<sup>1</sup>, Ivaylo Penev<sup>2</sup>, Mariana Todorova<sup>3</sup>, Hristiyan Bobev<sup>4</sup>, Neli Kalcheva<sup>5</sup>

**Abstract** – The paper presents implementation of an algorithm for movement of a robot from a start position to a final destination in a maze. The algorithm solves two main problems: transformation of the maze into a graph and finding the shortest path from an initial to a final position. The algorithm is implemented as a part of an application, using already generated text image, obtained by a picture of the maze from top. Walls (obstacles), empty spaces, starting position of the robot and the final destination are marked on the image. The maze is presented as a bit set to form a graph which vertexes are valid positions of the robot (i.e. the robot can rotate without touching a wall). The algorithm finds the shortest path, marks movement commands and saves them into a text file.

**Keywords** – Bit Set, Breadth First Search, Graph, Image, Maze, Shortest Path

## I. INTRODUCTION

The movement of a robot in a maze is a problem with many applications into lots of areas. Therefore the problem is an object of a broad research interest.

The main problems, which have to be solved, are presentation of the maze in a proper form and implementation of an algorithm for movement of the robot from initial position to a final destination (exit) into the maze. In the existing solutions the robot sensors or the camera of the robot are usually used [1, 3, 4]. Using these devices only a part of the maze could be shot. As a result information for the obstacles (the walls) in the maze is obtained during the robot movement. This approach is suitable, if the robot moves in an unknown or dynamically changing environment, but has some disadvantages:

- Requires permanent recalculation of the movement trajectory;
- Troubles the application of effective algorithms for path finding.

Various shortest path finding algorithms for robot movement in a maze have been researched. Some of the commonly used algorithms are: Dijkstra algorithm, A\* algorithm, breadth-first search (BFS), depth-first search (DFS) and etc.

Dijkstra's algorithm is one of the simplest algorithms.

<sup>1</sup>Milena Karova, <sup>2</sup>Ivaylo Penev, <sup>3</sup>Mariana Todorova, <sup>5</sup>Neli Kalcheva are with the Faculty of Automation and Computing at Technical University of Varna, 1 Studentska str., Varna 9010, Bulgaria, E-mail: [mkarova@ieee.bg](mailto:mkarova@ieee.bg), [ivailo.penev@tu-varna.bg](mailto:ivailo.penev@tu-varna.bg), [mgtodorova@yahoo.com](mailto:mgtodorova@yahoo.com), [n\\_kalcheva@abv.bg](mailto:n_kalcheva@abv.bg)

<sup>4</sup>Hristiyan Bobev is a student with the Faculty of Automation and Computing at Technical University of Varna, 1 Studentska str., Varna 9010, Bulgaria, E-mail: [hristiyan\\_bobev@abv.bg](mailto:hristiyan_bobev@abv.bg).

Starting from the initial vertex where the path should start, the algorithm marks all direct neighbors of the initial vertex with the cost to get there. It then proceeds from the vertex with the lowest cost to all of its adjacent vertices and marks them with the cost to get to them via itself if this cost is lower. Once all neighbors of a vertex are checked, the algorithm proceeds to the vertex with the next lowest cost [2, 5].

A\* is like Dijkstra's algorithm in that it can be used to find a shortest path. A\* algorithm is the most popular choice for path finding, because it's fairly flexible and can be used in a wide range of contexts. It is one of a family of graph search algorithms that follow the same structure. These algorithms represent the map as a graph and then find a path in that graph. Depending on the environment, A\* algorithm might accomplish search much faster than Dijkstra's algorithm [5].

Graph traversing in depth (Breadth-First-Search or BFS) is algorithm for path searching in trees or graphs [6]. Searching starts from a given vertex in the graph (or from the root in case of a tree) and all the unvisited neighbors are visited first. The algorithm is proposed by E. F. Moore, who used it for finding the shortest path in a maze.

The algorithm, proposed in this paper, is called AlgoHris. It solves two main problems:

- Construction of a graph using the image of the maze, shot from above (for example by the camera of a drone);
- Finding the shortest path from an initial position to a final destination in the graph.

The AlgoHris algorithm is compared to three other algorithms: Backtracking, A\* and Genetic Algorithm GAPP.

## II. ALGORITHM ALGOHRIS FORMULATION INSTRUCTIONS FOR THE AUTHORS

The AlgoHris algorithm considers an image of the maze. The image is shot from above, for example by the camera of a drone. Thus AlgoHris is integrated into an application for shooting a maze from above, finding shortest path in the maze and moving a robot from an initial position to a final destination in the maze (Fig. 1).

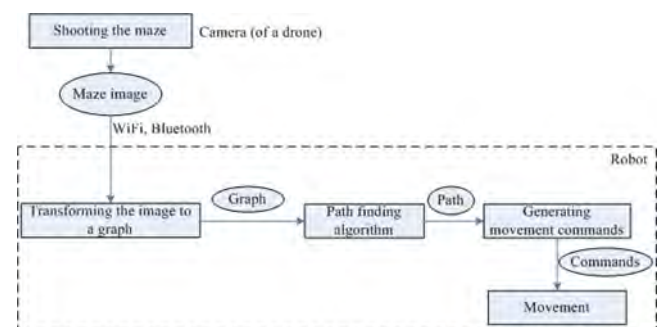


Fig. 1. Schema of the whole project

The proposed algorithm consists of two parts:

- Construction of a graph, presenting the maze;
- Finding the shortest path from an initial position to a final destination, using the graph.

#### A. Graph construction age

The construction of a graph is an essential problem for the robot movement. This problem is hard due to the limited resources of the robot platform (computing power, memory).

The proposed AlgoHris algorithm uses presentation of the maze as a set of characters 0 and 1, where 0 marks free (possible) position for the robot movement and 1 marks busy position (i.e. an obstacle or a wall). Fig. 2 presents a simplified maze and a graph with an initial position and a final destination.

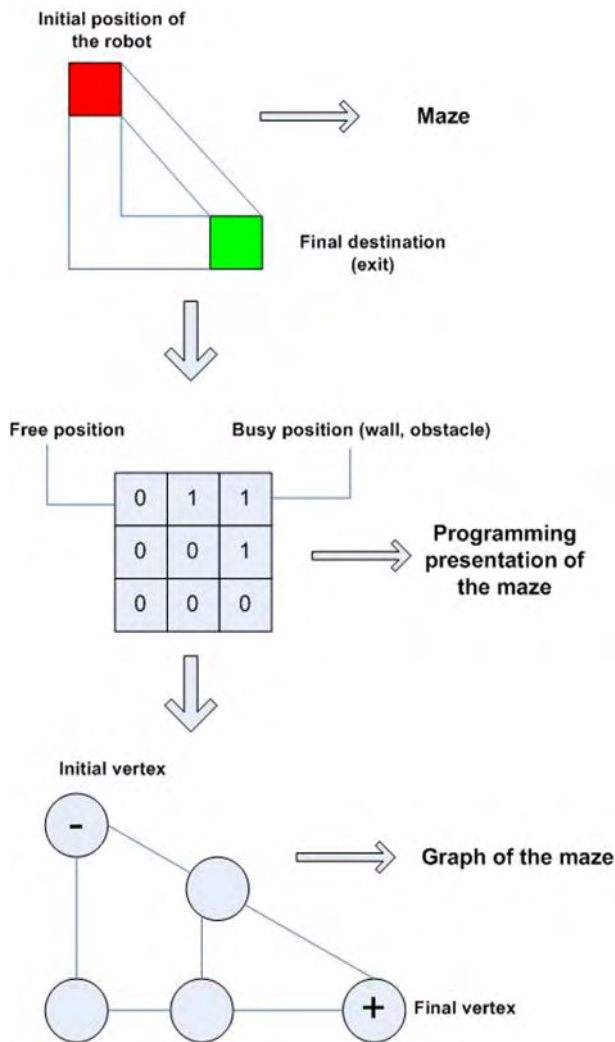


Fig. 2. Construction of a graph for an example maze

The algorithm constructs unweighted, indirect graph, formed by all pixels, which could be a valid position of the robot in the maze. A position is valid, if all pixels up, down, left and right in a distance equal to the robot size are free. Such construction of the graph guarantees, that only the

possible routes are presented in the graph. Over this graph various algorithms for path finding could be applied.

The graph construction algorithm uses radius of the circle, described around the robot (Fig. 3). This presentation is used for checking the valid positions of the robot in horizontal and vertical direction.

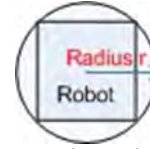


Fig. 3. Robot presentation, using described circle

#### B. Check for valid position in horizontal direction

The valid positions in horizontal direction are checked according to the radius of the described circle (i.e. according to the robot size) (Fig. 4).

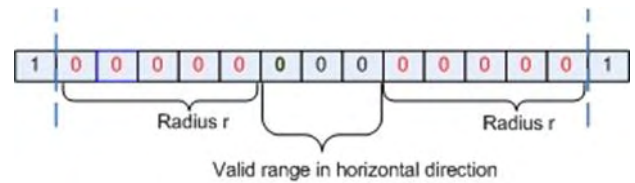


Fig. 4. Valid positions in horizontal direction

If there are no valid positions in horizontal direction, no vertex is added to the graph for these positions (Fig. 5).

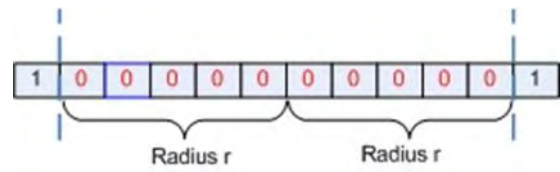


Fig. 5. No valid positions in horizontal direction

#### C. Check for valid position in vertical direction

The valid positions of the robot in vertical direction are formed by OR operation of each pair of rows from the set, presenting the maze (Fig. 6).

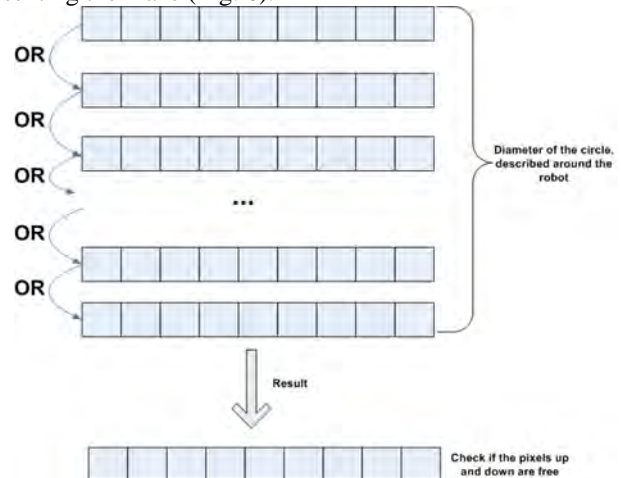


Fig. 6. Check for valid positions in vertical direction

If the OR operation of a pair of neighbor rows results in a character 1, then an obstacle is available on this position and the robot could not be in the position. No vertex is added to the graph for the position (Fig. 7).

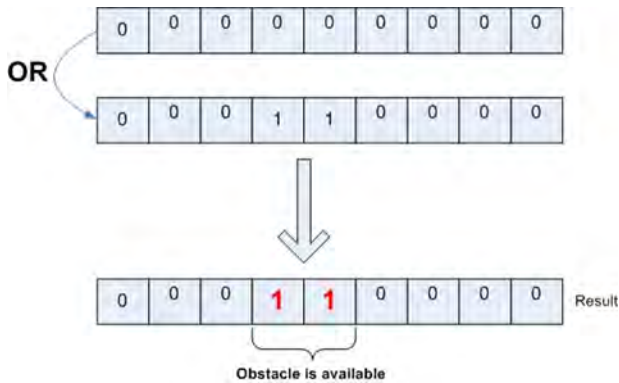


Fig. 7. An obstacle in vertical direction

If the OR operation of a pair of neighbor rows results in characters 0 only, then the positions in vertical direction are free. A vertex is added to the graph (Fig. 8).

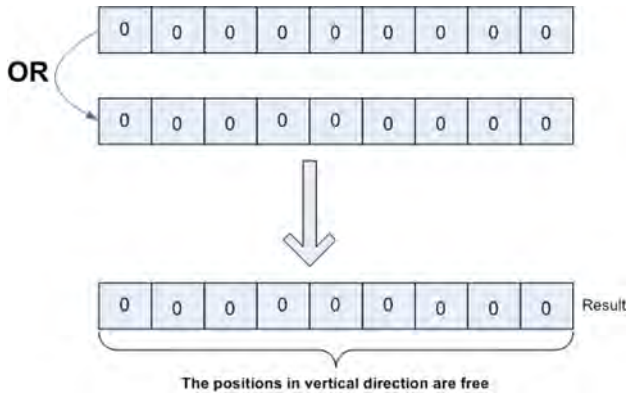


Fig. 8. Example for valid positions in vertical direction

#### D. Finding the shortest path

If the described graph, containing only the possible positions of movement, is available, an algorithm for finding the shortest path could be applied over the graph.

The algorithm, proposed in this paper, implements breadth first search. The algorithm finds the shortest path between an initial and a final vertex, passing through least number of visited vertices.

The algorithm uses the following data structures:

- Queue of the graph vertices;
- List of the visited vertices;
- List of the possible paths.

First the initial vertex is marked as visited and is added to the queue. Afterwards the neighbor vertices are traversed. If a neighbor vertex is not visited, in the field "previous" of the vertex the number of the previous neighbor is written. The vertex is marked as visited and is added to the queue. The loop continues until the queue is not empty and the final vertex is not visited.

Fig. 9 presents the algorithm execution for an example graph.

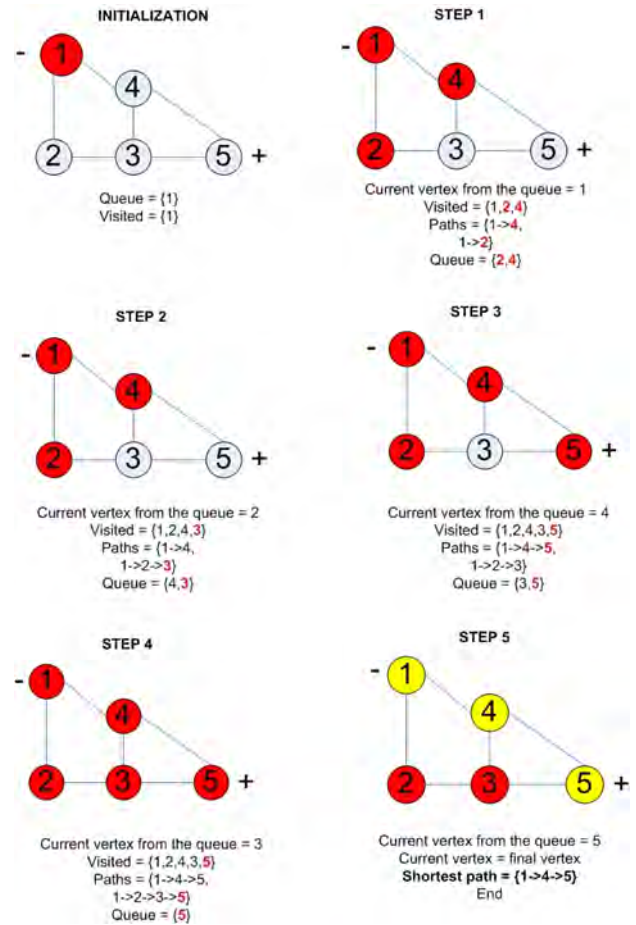


Fig. 9. Algorithm for finding the shortest path in the constructed graph

#### E. Programming implementation

The most challenging problem in the implementation of the described algorithm is proper presentation of mazes with large dimensions (e.g. 1600x1013 pixels). Such images have to be quickly transformed to a graph, considering the limited memory and computing power in the selected robot platform.

In the implementation of the AlgoHris algorithm the maze is presented as a set of bits. Each bit is a structure of elements with only 2 possible values: 0 (true) or 1 (false). A programming class is realized, emulating an array of Boolean elements. Each element reserves only one bit, which is eight times less than the simple character data type. Each bit could be accessed individually as in ordinary array.

### III. EXPERIMENTAL RESULTS

The experiments are carried out in two directions:

- Comparing the times for graph construction for mazes with various dimensions;
- Comparing the algorithm with three other algorithms, using the same dimensions (height and width in pixels) of the maze – Backtracking, A\* and genetic algorithm GAPP.



In the first experiment the times for completing the separate steps of the whole algorithm are presented in Table I.

TABLE I  
EXPERIMENTAL RESULTS WITH IMAGES OF EXAMPLE MAZES WITH  
DIFFERENT RESOLUTION

Width of the maze (pixels)	Height of the maze (pixels)	Time for graph construction (ms)	Time for shortest path finding (ms)
480	304	45	5
480	304	45	5
480	304	46	6
1000	820	72	23
1000	820	73	24
1600	1013	105	52

The results show, that the implemented algorithm processes relatively large image (1600x1013 pixels) and finds the shortest path for less than 160 ms.

The data, obtained by the second experiment, are presented on Fig. 10. In the case of small dimensions of the maze there is no significant difference between the times for shortest path finding. The genetic algorithm is comparatively slow and inapplicable to large dimensions of the maze. AlgoHris achieves shorter time than the Backtracking and even than the A\* algorithm when increasing the dimensions of the maze.

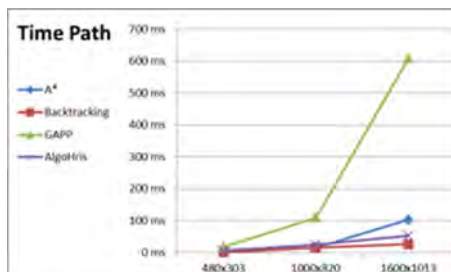


Fig. 10. The shortest path finding time for four different algorithms

On Fig. 11 the approximation function for AlgoHris is shown. The function is

$$y = 23x - 18.667, R^2 = 0.9845$$

The dependence between the time for path finding and increasing the maze dimensions is proportional and the algorithm is stable.

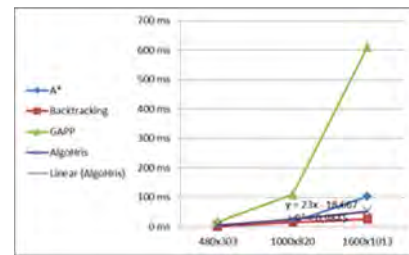


Fig. 11. Trendline correlation for AlgoHris

## IV. CONCLUSION

The proposed AlgoHris algorithm for movement of a robot in a maze has the following advantages in comparison to the existing implementations of other algorithms:

- It uses an image of the maze, which makes possible fast processing of the pixels in the image to find a path in the maze;

- The maze is presented as a bit set, which significantly increases the graph construction, containing only the possible paths for movement of the robot.

Although in the current implementation AlgoHris uses the breadth first search method to find the shortest path, other methods for shortest path finding as Dijkstra's algorithm and A\* algorithm could be effectively applied over the graph, produced by AlgoHris.

The future work will consider implementation of algorithms for finding paths in mazes, which are not shot in advance. The camera or the sensors of the robot will be used. This way the robot will be able to move in dynamically changing environment.

## REFERENCES

- [1] Chauhan S., Bajpai A., Collision Free Autonomous Robot Path Planning. *International Journal of Engineering Research & Technology*, Vol. 1, Iss. 8, e-ISSN: 2278-0181, 2012.
- [2] Hachour O., The proposed path finding strategy in static unknown environments. *International Journal of Systems Applications, Engineering & Development*, Iss. 4, Vol. 3, 2009.
- [3] Hachour O., Path planning of Autonomous Mobile robot. *International Journal of Systems Applications, Engineering & Development*, Iss. 4, Vol. 2, 2008.
- [4] Sedgewick R., Wayne K., *Algorithms, 4th Ed.*, ISBN-13: 978-0-321-57351-3, Pearson Education, 2011.
- [5] Naumov V., Karova M., D. Zhelyazkov, M. Todorova, I. Penev, V. Nikolov, V. Petkov, Robot Path Planning Algorithm, *International Journal of Computers and Communications*, ISSN: 2074-1294, NAUN, 2015.
- [6] Correll N., *Introduction to Autonomous Robots*, 1st edition, ISBN-13:978-1493773077, 2014.



# **Informatics and Computer Science**



# A Hybrid Metaheuristic Algorithm for Flexible Job Shop Scheduling "PSO&TS"

Asen Tochev<sup>1</sup> and Vassil Guliashki<sup>2</sup>

**Abstract** – In this paper is presented a new developed hybrid metaheuristic algorithm, designed to solve flexible job shop scheduling problems. The algorithm combines a Particle Swarm Optimization procedure with Tabu search technique, using neighborhoods with one operation movement, as well as a final local search using neighborhoods with one and two operations movement. An illustrative example is presented. The novel algorithm is called "PSO&TS" and is suitable for large sized flexible job shop scheduling problems.

**Keywords** – Flexible Job Shop Scheduling Problem, Particle Swarm Optimization, Tabu Search.

## I. INTRODUCTION

In the job shop scheduling problem (JSSP), there are  $n$  jobs that must be processed on a group of  $m$  machines. The set of machines is denoted by  $M$ . Each job  $i$  consists of a sequence of  $h$  operations ( $O_{i1}, O_{i2}, \dots, O_{ih_i}$ ), where  $O_{ik_i}$  (the  $k_i$ -th operation of job  $i$ ) must be processed without interruption on a predefined machine  $j$  for  $p_{ikj}$  time units. The operations  $O_{i1}, O_{i2}, \dots, O_{ih_i}$  must be processed one after another in the given order and each machine can process at most one operation at a time.

In this paper we study the flexible job shop scheduling problem (FJSSP). The FJSSPs are an extension of classical JSSP taking into account the production flexibility. Unlike the classical JSSP, where each operation is processed on a predefined machine, each operation  $O_{ikj}$  in the FJSSP can be processed on one machine  $M_l$  out of several machines  $M_l \in M_{ij}, M_{ij} \subseteq M$ . If  $M_{ij} \subset M$  for at least one operation, then it is a partial flexibility of JSSP (P-FJSSP); while if  $M_{ij} \equiv M$  for each operation, then it is a total flexibility of JSSP (T-FJSSP).

In a flexible job shop, each job  $i$  consists of a sequence of  $h_i$  operations ( $O_{i1}, O_{i2}, \dots, O_{ih_i}$ ). The processing time of operation  $O_{ik}$  on machine  $j, j = 1, \dots, m$ ; is  $p_{ikj} > 0$ , where  $k = 1, \dots, h_i$ . The FJSSP problem is to choose for each operation  $O_{ik_i}$  a machine  $M(O_{ik_i}) = M_{ik_i}$  and a starting time  $s_{ik_i}$  at which the operation must start its processing.

Detailed definition of FJSSP is given (see [1-4]) as follows:

- A set of  $J$  independent jobs.
- Each job  $J_i, i = 1, \dots, n$ ; can be processed on a given set of machines  $M_i$ .
- The  $O_{i,k_i}$  represents the  $k_i$ -th operation of  $J_i$ . The machines set waiting for processing the  $O_{i,k}$  noted by  $M_j \subseteq M$ .

<sup>1</sup>Asen Tochev is with the Institute of Information and Communication Technologies – BAS, "Acad. G. Bonchev" Str. Bl. 2, 1113 Sofia, Bulgaria, E-mail: tochevassen@yahoo.com.

<sup>2</sup>Vassil 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.

- We use  $p_{i,k_i,j}$  to represent the processing time of  $O_{i,k}$  operated on the  $j$ -th machine.
- There have two assumptions: a started operation cannot be interrupted; each machine can process only one operation at the same time.

The objective in our paper is to find the minimum time to finish all the operations.

In FJSSP one job can be processed on different machines and the distinct jobs can have different number of operations. In flexible job shop, job  $J_i$  has  $h_i$  operations to be performed in a strong sequence, corresponding to the index consequence  $1, \dots, h_i$ ; each job/operation is attributed to more than one machine which can perform operations on each of them. In this article we assume that each operation can be performed on at least two machines as alternatives.

This model represents the expansion of the upper JSSP - model. Unlike the previous model, in which each operation can be performed only on a single machine, in this model an operation  $O_{ik_i}$  can be performed on a machine  $M_l$  from a number of possible machines (machines subgroup). This subgroup is naturally different for each operation. In other words, it is not known in advance, any transaction of which one machine must be fulfilled.

Flexible job shop scheduling problem (FJSSP) can be decomposed naturally into two tasks: 1) object routing (appointment) at which each transaction is assigned to a corresponding machine chosen among the set of machines, on which its processing is possible; 2) subtask scheduling (sequence) where you have to calculate the start times of the appointed operations of all machines to obtain acceptable schedule (which does not violate the technological limitations) and to optimize predefined criteria (objective functions).

JSSP is NP hard, hence FJSSP is also NP hard (see [5]). Its optimal solution is difficult to be obtained even when the objective is to minimize the makespan (the time for processing all the jobs). For this reason, many studies are focused on the development of heuristic procedures for this problem. Well known among them are the particle swarm optimization (PSO) algorithm (see [9, 10]) and the metaheuristic Tabu search (TS) [14-17]. According [18] the job shop instances larger than  $15 \times 15$  have to be solved by heuristics and the Tabu search heuristic is the best known for JSP. In [9] is demonstrated that PSO obtains better results (mainly considering the convergence of the search process) in a faster and cheaper way in comparison to other evolutionary methods. Based on these statements the idea arises, to combine PSO and TS techniques using the advantages of these very successful optimization procedures and to create a hybrid metaheuristic algorithm, designed to solve flexible job shop scheduling problems.

The paper is organized as follows: in section II the

subprocedure PSO is presented; in s. III the subprocedure TS is presented. The phases of the new algorithm are considered in s. IV. The PSO&TS algorithm is presented in s. V. An illustrative example is given in s. VI. In s. VII some conclusions are drawn.

## II. Subprocedure PSO

One powerful heuristic is the Particle Swarm Optimization (PSO). It has been first developed by Eberhart and Kennedy [6-8]. The basic idea in PSO is to imitate the intelligent swarming behavior, observed in flocks of birds, schools of fish, swarms of bees, etc. Each particle in PSO represents one solution and it makes steps from its current position to a new position. This movement is determined as a sum of three vectors: *inertia*, *competition* and *cooperation*. The inertia vector is determined by the current velocity of the particle  $v(t)$  weighted by a constant  $w$ . In this way the tendency of the particle to maintain its current velocity is reflected. The competition vector links the current position of the particle  $x(t)$  to its personal best position found during the search process. This vector is weighted using a uniformly distributed random diagonal matrix  $Dr1$ . The cooperation vector links the current position of the particle  $x(t)$  to the global best position, found by the particles. This vector is weighted using a second uniformly distributed diagonal matrix  $Dr2$ . It is clear that the cooperation among particles is important for finding the global optimum solution. The inertia and the competition are necessary for the particle to avoid trapping in the local minima.

Suppose that the search space is  $d$ -dimensional, and then the  $i$ -th particle of the swarm can be represented by a  $d$ -dimensional vector,  $x_i = (x_{i1}, x_{i2}, \dots, x_{ij}, \dots, x_{id})$ ,  $d = m * n$ . The velocity of the particle can be represented by another vector,  $v_i = (v_{i1}, v_{i2}, \dots, v_{ij}, \dots, v_{id})$ . The best solution achieved by the particle is denoted as  $p_i = (p_{i1}, p_{i2}, \dots, p_{ij}, \dots, p_{id})$ , and the global best solution/particle is denoted by  $g = (g_1, g_2, \dots, g_j, \dots, g_d)$ . The particle updates its velocity and position by means of equation (1) and (2) as follows.

$$v_{i=L} = w * v_i + b * Dr1 * (p_i - x_i) + c * Dr2 * (g - x_i) \quad (1)$$

$$x_i^{(k+1)} = x_i^{(k)} + v_i \quad (2)$$

The appropriate choice of inertial weight  $w$  in (1) can provide a balance between global and local research's operating capabilities of the swarm, and resulted in an average less reps to find enough optimum solution. Accelerating constants  $b$  and  $c$  in equation (1) represent the weight of stochastic acceleration terms that draw every particle  $i$  in  $p_i$  and  $g_i$  directions. In this study  $b = 1.8$  and  $c = 2.2$ . Lower values lead to small convergence speed, while higher values result in jerky movements through the search region.

In the last years PSO algorithms have been successfully applied in many research areas for different applications. It is demonstrated – see [9], that PSO obtains better results (mainly considering the convergence of the search process) in a faster and cheaper way in comparison to other evolutionary methods. In [10] another hybrid algorithm, combining particle

swarm optimization and a tabu search technique to solve FJSSP is proposed.

## III. Subprocedure TS

This is a widely used robust metaheuristic created by Fred Glover (see [11-13]). The main idea of TS is to perform a strategic search procedure, starting from one solution and avoiding cycles that lead to visiting the same solution by means of a memory (more precisely a set of memories), storing some characteristics of solutions or movements (steps in given directions) which are forbidden (tabu) for a certain number of iterations. In this manner is avoided the trap of local optimality. The list of forbidden characteristics or movements stored in the memory is called tabu list. Typically the stopping criteria correspond to the iterations limit and to the number of consecutive iterations without improving the best obtained solution. Considering FJSSP the Tabu search technique requires a starting solution and a neighborhood structure [19]. An initial solution is used to start TS. It is stored as a current seed and the best solution. Exploring the neighborhood  $V(x_i)$  of  $x_i$  are produced new solutions, called candidate solutions. The best according its objective function value candidate solution, which is not a tabu or which satisfies preliminary defined aspiration criteria, is selected as a new seed solution. This selection is called move. The new seed solution is compared with the current seed solution. If it is better, it is stored as new best solution. To avoid cycling the last  $L$  solutions and/or found are stored in the so called tabu list. Iterations are repeated until the stop criteria are satisfied.

Single objective TS algorithm for FJSSP is presented in [14]. Recently a Pareto-based tabu search algorithm for multi-objective FJSSP is proposed in [16]. Another effective hybrid Tabu search algorithm for the same multi-objective problem is developed in [15]. In [17] the multi-objective FJSSP is solved using a novel path-relinking algorithm based on the state-of-the-art Tabu search algorithm.

## IV. Proposed phases of the new algorithm

- 1) **Phase** generation of one initial solution like the procedure described in [20].
- 2) **Phase** Initialization of initial swarm in the PSO.
- 3) **Phase** PSO
- 4) **Phase** TS - as an initial parameter of the algorithm is taken the `global_best` solution obtained by means of PSO. The path through the critical machine is assumed to be "tabu".

## V. Hybrid algorithm PSO & TS

The main idea of the proposed algorithm is the following: First a good solution to be found out by means of PSO heuristic. Then the "critical path" (the sequence of operations on the machine finishing last its processing) is assumed to become "tabu". After that a TS procedure is used, attempting to improve `global_best` generated by PSO. The phases 3)



"PSO" and 4) "TS" are repeated iteratively until the stop condition is met as shown on Fig. 1.

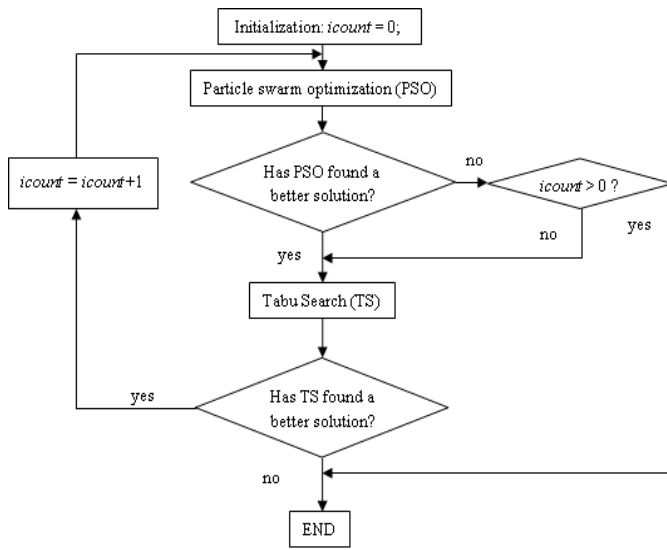


Fig. 1. Flow chart of PSO&TS algorithm

In PSO, each solution is a point in the search space and may be considered as a particle. Initially, a set of particles is randomly created and are moved through the search space. In their movement, the particles have memory, and each particle regulates its position on the base of its own experience as well as the experience of a next particle by using the best position faced by itself and its neighbors. The previous best value is known as the *pbest* (personally best) of the particle, and the best value of all the particles' *pbest* in the swarm is known as the *gbest* (globally best). In each iteration, the velocity and the position of each particle will be updated on the basis of its *pbest* and *gbest*. Let the search space is *d*-dimensional, and the position of the *i*-th particle at the *t*-th iteration is shown by  $X_i^t = (x_{i1}^t, x_{i2}^t, \dots, x_{iD}^t)$ . The velocity of the *i*-th particle at the *t*-th iteration is  $V_i^t = (v_{i1}^t, v_{i2}^t, \dots, v_{iD}^t)$ . The best position of the *i*-th particle found out till the *t*-th iteration is .

$P_i^t = (p_{i1}^t, p_{i2}^t, \dots, p_{iD}^t)$ . The best position detected for all swarm is denoted as  $G^t = (g_1^t, g_2^t, \dots, g_D^t)$  at the *t*-th iteration. The velocity and position of particles are updated by equations (3) and (4), corresponding to (1) and (2) in the standard PSO:

$$v_i^{t+1} = w \times v_i^t + c_1 \times rand() \times (p_i^t - x_i^t) + c_2 \times rand() \times (g_i^t - x_{id}^t) \quad (3)$$

$$x_i^{t+1} = x_i^t + v_i^{t+1} \quad (4)$$

where  $x_i^t$  and  $v_i^t$  indicate the position and velocity of the *i*-th component of one particle at the *t*-th iteration, respectively. *w* is the inertia weight represented to balance between the global and local search capabilities, and *rand()* is a random number creator with a uniform distribution over [0, 1]. *c*<sub>1</sub> and *c*<sub>2</sub> are the acceleration constants, which suggest the weighting of stochastic acceleration terms that force each particle toward *pbest* and *gbest*, respectively. The updating formula forces

particles moving toward compound vector of local and global optimal solutions. Therefore, opportunity for particles to get the optimal solution will increase.

The pseudocode of the algorithm is presented as follows:

**Step 1.** Initialization.

1) PSO:

- Initialize a swarm of particles;
- Give initial values of *w*, *b*, *c*, *iter\_lim*;
- Put the counter *i* = 0.

2) TS:

- Initialize to the algorithm to generate the initial solution *S*<sub>0</sub>; // global\_best
- *S*<sub>best</sub> = *S*<sub>0</sub>; // initialize global\_best
- TabuList = null;

**Step 2.** Iterative process:

L1: while (the stop criterion is not met)

- { - generate a next swarm according to equations for particles;
- find new particle\_best of each particle and global\_best of swarm;
- update particle\_best of each particle and global\_best of the swarms;
- If global\_best has not been improved
- Then counter *i* = *i* + 1;
- End If }

If counter *i* == *iter\_lim*

{// for particle global\_best to perform TS

until (it is not satisfied the stop condition)

CandidateList ← []

BestCandidate ← null

for (Scandidate in Senvironment )

if ((not TabuList contains Scandidate)

and (fitness(Scandidate) > fitness(BestCandidate)))

then BestCandidate ← *S*<sub>candidate</sub>

end if

end for

*S* ← BestCandidate

If (fitness(BestCandidate) > fitness(*S*<sub>best</sub>))

*S*<sub>best</sub> ← BestCandidate

end if

push into TabuList (BestCandidate);

if (TabuList.size > max TabuSize)

then TabuList.deleteFirst ()

end if }

end if

If global\_best improved

Then GOTO L1

return global\_best = *S*<sub>best</sub>

**Step 3.** End. Show the optimization solution.

Variables *w*, *b*, *c*, *iter\_lim* should be chosen by the decision-maker / the user. In the next example the following values are used: *w* = 1; *b* = 1.8; *c* = 2.2; *max\_gen* = 100.

## VI. Illustrative example

Below is solved an example of FJSSP of Fattahi et al. called M2J2O4, where M means machines, J – jobs and O – operations.

TABLE I  
PROCESSING TIMES

$p_{ijh}$		<i>M</i> <sub>1</sub>	<i>M</i> <sub>2</sub>
J1	<i>O</i> <sub>11</sub>	25	37
	<i>O</i> <sub>12</sub>	32	24
J2	<i>O</i> <sub>21</sub>	45	65
	<i>O</i> <sub>22</sub>	21	65

1)  $S_{km} = 0$  for every  $O_{km} \in A$

Then

$S_{11} = 0$  for  $O_{11} \in A$

$S_{21} = 0$  for  $O_{21} \in A$

$t(A) = \min(s_{km} + p_{km}) = \min(25, 65) = 25$

$M = \{0\}$

$(k, m) \in A$

$O_{11} \in A$ .  $SO_{11} = 25 = t(A)$ . It follows that  $SO_{11}$  is not smaller of  $t(A)$ .

$O_{21} \in A$ .  $SO_{21} = 25 = t(A)$ . It follows that  $SO_{21}$  is not smaller of  $t(A)$ .

Then  $G = \{0\}$ .

$m^* = M2$

$O_{21}$  runs on machine 2.

Then  $G = \{O_{21}\}$

$(k, m^*) = (k, 2) = O_{21}$

Earliest time of  $O_{21} = 65$ .

$k^* = J_2$

$A = \{O_{11}\}$

$A = \{O_{11}, O_{22}\}$

$R = \{O_{21}\}$

With this ends the first iteration.

On second iteration:

$R = \{O_{21}, O_{22}\}$

On third iteration:

$R = \{O_{21}, O_{22}, O_{11}\}$

On forth iteration:

$R = \{O_{21}, O_{22}, O_{11}, O_{12}\}$  is the initializing schedule for PSO.

2) Initial schedule  $R = \{O_{21}, O_{22}, O_{11}, O_{12}\}$ .

**3) PSO:**

The schedule  $R = \{O_{21}, O_{22}, O_{11}, O_{12}\}$  is not improved.

**4) TS:**

1)  $O_{11}, O_{12}$  are run on  $M_1$ .  $O_{21}, O_{22}$  – on  $M_2$ . Makespan = 130.

$O_{21}, O_{11}, O_{12}$  are run on  $M_1$ .  $O_{22}$  – on  $M_2$ . Makespan = 102. This is better than 130.

$O_{11}, O_{12}, O_{22}$  are run on  $M_1$ .  $O_{21}$  – on  $M_2$ . Makespan = 86.

$O_{12}$  is run on  $M_1$ .  $O_{11}, O_{21}, O_{22}$  – on  $M_2$ . Makespan = 167.

$O_{11}$  is run on  $M_1$ .  $O_{21}, O_{12}, O_{22}$  – on  $M_2$ . Makespan = 154.

2)  $O_{11}, O_{12}, O_{22}$  are run on  $M_1$ .  $O_{21}$  – on  $M_2$ . Makespan = 86.

Minimum.

$O_{11}, O_{12}, O_{21}, O_{22}$  are run on  $M_1$ .  $M_2$  - deadtime. Makespan = 187.

$O_{12}, O_{22}$  are run on  $M_1$ .  $O_{11}, O_{21}$  – on  $M_2$ . Makespan = 167.

3)  $O_{11}, O_{12}$  are run on  $M_1$ .  $O_{21}, O_{22}$  – on  $M_2$ . Makespan = 130.

Minimum.

$O_{22}$  is run on  $M_1$ .  $O_{11}, O_{21}, O_{12}$  – on  $M_2$ . Makespan = 126.

$O_{11}, O_{21}, O_{22}$  are run on  $M_1$ .  $O_{21}$  – on  $M_2$ . Makespan = 91.

$O_{11}, O_{12}, O_{22}$  are run on  $M_1$ .  $O_{21}$  – on  $M_2$ . Makespan = 130.

$O_{21}, O_{22}$  are run on  $M_1$ .  $O_{11}, O_{12}$  – on  $M_2$ . Makespan = 66.

Minimum. Solution. (The optimal solution is obtained.)

## VII. CONCLUSION

The received results are encouraging. The PSO&TS algorithm has good performance and could be applied to solve large size problems. As a future step of this study PSO&TS will be tested on a set of benchmark examples.

## REFERENCES

[1] Jun-qing Li, Quan-ke Pan, Sheng-xian Xie, Yu-ting Wang, "A Fast TS Algorithm for Solving the Flexible Job-shop Scheduling

Problem". In Proceeding of International Conference on Information Technology (ICIT'2009), 2009, pp:5-4.

[2] Bruker P, Schlie R, (1990), "Job-shop scheduling with multi-purpose machines", *Computing*, 1990, 45(4), pp:369-375.

[3] Brandimarte P, (1993), "Routing and scheduling in a flexible job shop by tabu search", *Annals of Operations Research*, 1993, 22, pp:158-183.

[4] Liouane N, Saad I, Hammadi S, Borne P, (2007), "Ant Systems & Local Search Optimization for Flexible Job-Shop Scheduling Production", *International Journal of Computers, Communications & Control*, 2007, pp: 2174-184.

[5] Lawler, E., J. Lenstra, A. Rinnooy Kan, and D. Shmoys. (1993) Sequencing and scheduling: algorithms and complexity. In A.H.G. Rinnooy Kan S.C Graves and P.H. Zipkin, editors, *Logistics of Production and Inventory*, volume 4, Chapter 9, pages 445–522. Elsevier, 1993.

[6] Eberhart R. C. and J. Kennedy, (1995), "A new optimizer using particle swarm theory", *Proceedings of the Sixth International Symposium on Micromachine and Human Science*, Nagoya, Japan, pp. 39-43, 1995.

[7] Kennedy J. and R. C. Eberhart, (1995), "Particle Swarm Optimization", *Proceedings of IEEE International Conference on Neural Networks*, Piscataway, N. J., 1995, pp. 1942-1948.

[8] Kennedy J., R. C. Eberhart and Y. Shi, (2001), „*Swarm intelligence*“, San Francisco: Morgan Kaufmann Publishers, 2001.

[9] Ge H. W., Y. C. Liang, Y. Zhou, Guo, X.C., (2005), A Particle Swarm Optimization-based Algorithm for Job-shop Scheduling Problems, *Int. J. Computational Methods*, Vol. 2, No. 3, 2005, 419-430

[10] Zhang, G. H., Shao, X.Y., Li, P. G., & Gao, L. (2009). An effective hybrid particle swarm optimization algorithm for multi-objective flexible job-shop scheduling problem. *Computers and Industrial Engineering*, 56 (4), pp. 1309-1318.

[11] Glover F (1990) Tabu search-part II. *ORSA J. Comput.* 2:4–32

[12] Pirlot M (1996) General local search methods. *European Journal of Operational Research*, 92:493–511.

[13] Glover F. and M. Laguna, *Tabu Search*, Kluwer: Boston, MA, 1997.

[14] Saidi-Mehrabad M., Fattahi P., (2007), "Flexible job shop scheduling with tabu search algorithms", *International Journal of Advanced Manufacturing Technology*, March 2007, 32(5):563-570.

[15] Li J., Q. Pan and Y. Liang, (2010), "An effective hybrid tabu search algorithm for multi-objective flexible job-shop scheduling problems", *Computers & Industrial Engineering*, Vol. 59, Issue 4, November 2010, pp. 647-662.

[16] Li J., Q. Pan, S. Xie and J. Liang, (2010), "A Hybrid Pareto-Based Tabu Search for Multi-objective Flexible Job Shop Scheduling Problem with E/T Penalty", *Advances in Swarm Intelligence, Lecture Notes in Computer Science*, Vol. 6145, pp. 620-627.

[17] Jia S. and Z.-H. Hu, (2014), "Path-relinking Tabu search for the multi-objective flexible job shop scheduling problem", *Computers & Operations Research*, Vol. 47, July 2014, pp. 11-26.

[18] Brucker, P., (2007) The Job-Shop Problem: Old and New Challenges, 3-rd Multidisciplinary International Scheduling Conference, Paris, 28.-31. August, 2007.  
<http://www.mistaconference.org/2007/papers/The%20Job%20Shop%20Problem%20Old%20and%20New%20Challenges.pdf>

[19] J. P. Watson and J. Christopher Beck, "Problem Difficulty for Tabu Search in Job-Shop Scheduling", Elsevier Science, 2002, vol. 21.

[20] Hongwei Ge, Wenli Du, Feng Qian, "A Hybrid Algorithm based on PSO and SA for Job Shop Scheduling",  
<http://www.paper.edu.cn>.

# NoTree

Iliya Tronkov<sup>1</sup>

**Abstract** – The present article introduces an external merge based index data structure called NoTree. What is distinctive for it is that it does not work with individual keys, but with intervals of keys. NoTree inherits the message model from buffer based trees, but goes one step forward as it allows selecting and specific multiway merging of buffers from different levels. This gives a possibility for choosing an optimal path of buffers which can be merged with minimum number of I/O operations. For comparison of the different families of indexing technologies, a unified approach is applied (an interval model) which highlights the relation between them. The studies include B+-tree, Buffer Trees, LSM-tree, Fractal Index Tree, WaterfallTree, NoTree and the accent is not on the number of comparisons that are made between the keys, but on the number of I/O operations. The NoTree index is suitable for applications with intensive flow of input data which is constantly been queried with range queries that include a significant subset of the available data in the index.

**Keywords** – Database, Benchmark, Index Data Structure, Indexing, Comparative Analysis.

## I. INTRODUCTION

The keys from each leaf of the B-tree [1-3] based trees define an *interval* which is [minimal key from leaf, maximal key from leaf]. The so defined intervals do not have common points, i.e. they do not intersect two by two. (Without losing generality we consider that all keys are unique). We will call such set of intervals a *level of intervals*. Therefore B-tree based trees have only one level of intervals formed from their leaves (Fig. 1). The keys that are located in the internal nodes of the tree are not part of the user data and are excluded from our considerations.

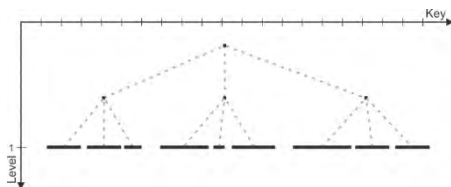


Fig. 1. Interval model of B-tree based data structure - has one level of intervals.

When random keys are entering the B-tree they end up in different intervals, i.e. different blocks will have to be accessed from the storage device. But if not all records fit into the memory, we will have to do an I/O operation to the external storage for only one record. This drastically decreases the speed of indexing. In order to reduce the negative effect from the random keys, buffers with messages are added in the internal nodes of the tree. The aim is to move messages not one by one but in groups with every I/O operation.

<sup>1</sup>Iliya Tronkov is with the Faculty of Computer Systems and Technologies at Technical University of Gabrovo, Bulgaria, E-mail: tronkov@gmail.com.

In buffered based trees (Buffer Tree [4-5], FractalIndex Tree [6], WaterfallTree [7-9]) we have several levels of intervals. In that case the leaves once again form one level of intervals. But the internal nodes also contain user data in the form of messages. Following the logic as in the leaves, the keys (of the messages) from every internal node define an interval. Therefore the buffer based trees have as many levels of intervals as their height, with the additional requirement for *subordination of the intervals* - every interval can intersect only with the intervals of its subtree and every level can have at most  $(\text{number of branches})^{\text{level number} - 1}$  intervals (Fig. 2):

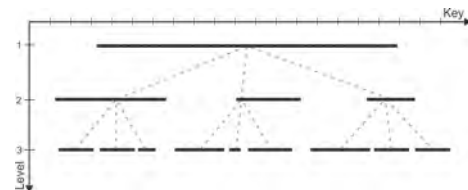


Fig. 2. Interval model of buffer tree-based data structure - has the same levels of intervals as the height of the tree with the additional subordination of the intervals.

Thanks to the buffers in buffer based trees the number of I/O operations is reduced, but in the meantime, if the speed with which the records fill a given buffer is faster than the speed in which the buffer is emptied, it becomes a tight spot for its whole subtree. This is mostly valid for the buffer in the root of the tree, because all data goes through it.

If we release ourselves from the restricting requirement of subordination of the intervals, we will derive to the idea of LSM-tree [10] (which is implemented in several systems BigTable [11], LevelDB, Apache HBase, RocksDB, etc.).

The structure of the LSM-tree is a combination of an index tree with messages into the memory (usually an AVL-tree) and several (usually two) B-tree indices on the storage device. (Most of the LSM-tree implementations have more than two levels on the disk device.) Thus, from the interval model point of view, in LSM-tree we have one level of intervals, which is defined by the AVL-tree and by a single level of intervals for every B-tree index. Every next level contains more or equal number of intervals from the previous level (Fig. 3). The logical partitioning of the AVL-tree in intervals is done in a way that every interval in it has approximately as many messages as a physical block on the storage device can store. (It is a matter of choice whether to include the intervals which are in memory into the model - we will include them for better illustration).

The elements of a given set of intervals can be permuted in a different way and in different number of levels. We will call such combination an *intervals configuration*. Thus, each of the different indexing paradigms follows a specific *interval model* – specific intervals configuration is maintained, which meets well defined restrictions (Fig. 4). From that point of view, the indexing algorithm transforms a configuration of intervals into one that has only one level of intervals.

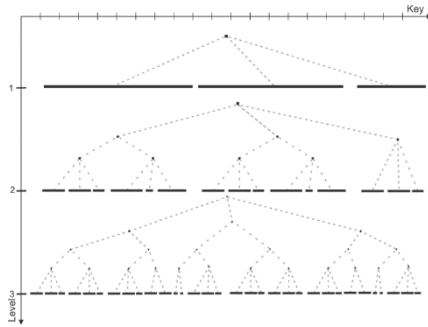


Fig. 3. Interval model of LSM-tree based data structure - has three levels of intervals without the requirement for subordination of intervals. Level 1 - AVL-tree (RAM), Level 2 - B+-tree (HDD), Level 3 - B+-tree (HDD).

As we can see, the intervals configuration of the different indexing technologies evolve towards increasing the number of levels and weakening the requirements for intersections among the intervals. But one thing remains unchanged in the current indexing algorithms – **the intervals do not change their levels and the merging of the data is always between two adjacent levels**. There are no limitations for us **to allow swapping of intervals between the levels and to merge multiple levels simultaneously** in a way that will reduce the overall number of I/O operations needed for the data indexing.

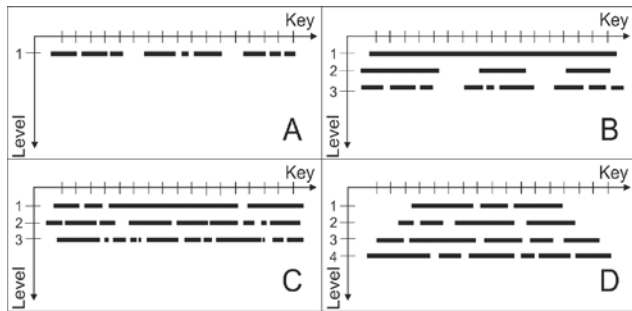


Fig. 4. Interval model of: A) B-tree – one level of intervals; B) Buffer Tree, FractalIndex Tree, WaterfallTree – subordination of the intervals; C) LSM-tree – increasing number of intervals in the levels; D) NoTree – no relations between the intervals.

## II. NOTREE

In NoTree, the input data is logged in files directly on the hard disk drive, each with approximately equal size  $B$ . The input data is logged in the form of messages. After a given file reaches its initially defined size  $B$ , the logging process continues in a new file. Parallel to the data logging, a statistic is maintained for the minimal and maximal key in every file. Thus, the data in each file defines an interval and in the process of logging the interval bounds are maintained. All intervals of the indexing structure are held into one set of intervals  $S$ . There are no connections, relations or subordination between the intervals (Fig. 4D). In that way the only limit for the incoming data is the write speed of the hard drive. The data indexing is done cyclically in two stages: *k-partitioning* and *multiway merging*:

1. **K-partitioning** – the intervals from  $S$  are distributed in sequence of groups which meet the following conditions:

a) For every group of intervals exists a configuration which has less than  $k$  levels. This condition guarantees that the intervals from a given group can be merged together with one pass when the available memory is limited to the size of  $k$  blocks ( $k \times B$ );

b) The  $k$ -partitioning should have as smaller *sum* as possible (see *K-partitioning of a set of intervals*). That sum (see Fig. 5) is the half of the number of I/O operations needed for transforming the index into one level of intervals. That follows from the definition of  $k$ -partitioning.

2. **Merging** – multiway merge of  $K$  levels of intervals in the last group of the  $k$ -partitioning.

The process continues until only one level of intervals in  $S$  remains. In the meantime, if read queries enter the NoTree, the indexing cycle is interrupted and is not executed for the entire set  $S$ , but for the subset of intervals in  $S$  defined by the queries. In that way the completed work until now is not lost and the merged intervals for the incoming queries contribute to the whole process of indexing. After execution of the queries, the indexing cycle of NoTree continues its work with the whole set of intervals  $S$ .

Essentially the indexing process is a modified external sort in two directions:

1. The  $k$  levels of intervals are chosen by the  $k$ -partitioning in such a way that the number of I/O operations when merging will be reduced. In case the index is static (no data is coming in for indexing) and the data has random keys, the order in which the intervals are chosen for merging would be irrelevant. But the structure itself is dynamic and while the existing intervals are merging, new data is coming in which forms new intervals. In that way two subsets of intervals are formed within the index – on one hand the already merged intervals which form levels of non-overlapping intervals and on the other hand newly formed intervals which form specific patterns depending on the incoming keys (in many applications the incoming data has a specific keys pattern). Thus, the  $k$ -partitioning is a significant step for determining the order for merging of intervals.

2. A significant part of the merging process is the “skipping” of intervals. The intervals that are already merged are non-overlapping and short, i.e. they have a high “density”. The newly entered intervals are long and overlap with many of the existing intervals but have a low density. That means that between two keys in the interval the probability for having intervals which can be skipped is increased with the increasing size of the index. Similar thing is happening in LSM-tree when two subsequent levels are merging – blocks for which no input keys are coming in are not loaded.

### A. K-PARTITIONING OF A SET OF INTERVALS

Let  $X$  be a set of objects for which a total (linear) order is defined [12]. Then the concept of closed interval over  $X$  is well defined. Let  $I$  be a set of intervals over  $X$ . We will mark with  $I^p$  the subset of  $I$  that contains all intervals in which the point  $p \in X$  belongs, i.e.  $I^p = \{i \in I | p \in i\}$ . We will define *depth* of  $I$  as  $\bar{I} = \max_{p \in X} |I^p|$ , i.e. the maximal number of overlapping



intervals in a point. It is obvious that no intervals configuration of  $I$  exists, with less levels of intervals than the depth of  $I$ .

Let  $I$  be a set of intervals and every interval refers a block with size  $B$ . If the available operating memory is  $M$  and  $k = M/B$ , then in every given moment we can simultaneously keep  $k$  blocks in memory. Therefore we can merge simultaneously  $k$  levels of intervals. Our goal is to re-order and merge the intervals of  $I$  with the least amount of I/O operations, in a way that at the end  $I$  will hold only one level of intervals, i.e.  $\bar{I} = 1$ .

We will call  $K$ -partitioning of  $I$  the sequence  $\Delta_k = G_1, G_2, \dots, G_n$ , from groups of intervals  $G_i \subseteq I$ , for which the following conditions are met:

1.  $G_i \neq \emptyset$ ,  $i = 1, 2, \dots, n$ , i.e. in every group there is at least one interval;
2.  $G_i \cap G_j = \emptyset$ ,  $1 \leq i, j \leq n$ ,  $i \neq j$ , i.e. the groups do not contain common intervals;
3.  $\bigcup_{i=1}^n G_i = I$ , i.e. all intervals from  $I$  participate in  $\Delta_k$ ;
4.  $|G_i| < k$ ,  $i = 1, 2, \dots, n$ , i.e. every group  $G_i$  of  $\Delta_k$  has a depth less than  $k$ , where  $k \geq 2$  is constant.

We define a *sum* of  $k$ -partitioning  $\Delta_k = G_1, G_2, \dots, G_n$  as  $S(\Delta_k) = 1|G_1| + 2|G_2| + \dots + n|G_n| = \sum_{i=1}^n i|G_i|$ , where  $|G_i|$  is the number of intervals in the group  $G_i \in \Delta_k$  (Fig. 5). Directly from the definition follows that if  $\alpha_k = A_1, A_2, \dots, A_m$  and  $\beta_k = B_1, B_2, \dots, B_n$  are two  $k$ -partitions of  $I$ , where  $m \leq n$  and  $|A_i| \leq |B_i|$  for  $i = 1, 2, \dots, m$ ,  $S(\alpha_k) \leq S(\beta_k)$ . We can see that the sum defines a partial order [32] between the  $k$ -partitions, i.e. we will say that for two  $k$ -partitions  $\alpha_k$  and  $\beta_k$ ,  $\alpha_k \leq \beta_k$ , exactly when  $S(\alpha_k) \leq S(\beta_k)$ . Therefore for every set of intervals  $I$  exist  $k$ -partitions with a minimal sum, which we will call *minimal*  $k$ -partitions.

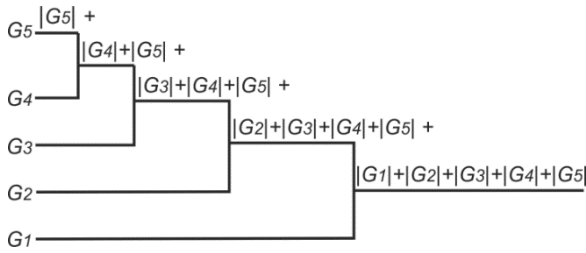


Fig. 5. Sum of  $k$ -partitioning that contains a total of 5 groups of intervals.

Exactly the minimal  $k$ -partitions are interesting to us, because the upper bound of I/O operations needed for converting a random set of intervals in a set that has one level of intervals is  $2 \times S(\Delta_k)$  (once for writing and once for reading).

After finding a  $k$ -partition, begins a process of merging the intervals of the last group. Moreover, when the last group is merged, a new level of intervals is formed which can change the  $k$ -partition in such one that has a smaller sum.

### III. TEST RESULTS

All of the tests (Fig. 6, Fig. 7) are performed with the publicly available open source tool *Database Benchmark 3.0*. The tests with random keys are conducted on the following computer configuration: CPU Intel Core i7-3770 3.4 GHz with 8MB L3 cache, RAM 32GB DDR3-1600 MHz, HDD 3TB

Western Digital Caviar Green, OS Windows 7 Professional Service pack 1 64 bit with NTFS / Ubuntu 14.04 64 bit with EXT4. The test data has the following structure: **key** – integer of type long (8 bytes); **record** – complex type (~52 bytes). The number of records for all tests is  $N = 10^9$ . A pseudo-random generator is used for generating the random keys. A random-walk algorithm is used for generating the records. The tests are conducted with the following databases: NoTree (STSDb 5 alpha), FractalIndex Tree (TokuDB 7.5.5 Community Edition for MySQL 5.5.41), LSM-tree (LevelDB 1.16).

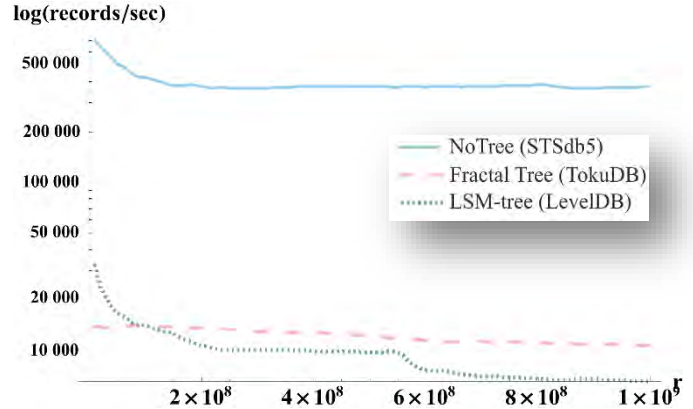


Fig. 6. Insert of  $10^9$  records with random keys. Average speeds: NoTree - 371 504 rec/s, FractalIndex Tree - 10 869 rec/s, LSM-tree - 6 592 rec/s.

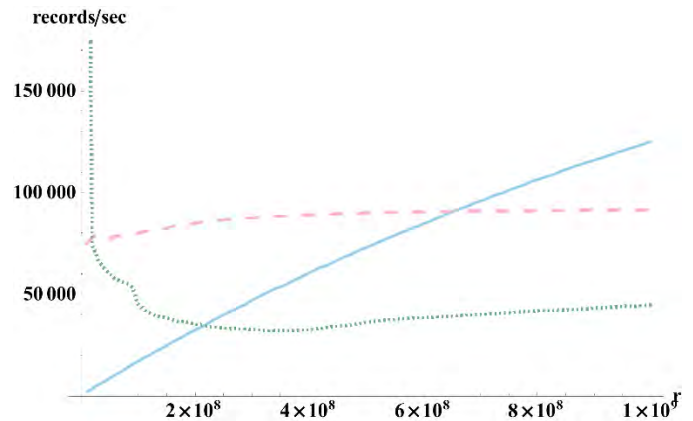


Fig. 7. Immediately full read. First sequential (in ascending order) read of all records immediately after they are inserted. Average speeds: NoTree – 125 149 rec/s, FractalIndex Tree – 91 485 rec/s, LSM-tree – 44 509 rec/s.

Total time for the whole test (Insert + Immediately full read) of  $10^9$  records: NoTree – 2h 58min=(45min + 2h 13min), FractalIndex Tree – 28h 35min=(25h 33min + 3h 2min), LSM-tree – 48h 22min=(42h 8min + 6h 14min).

### IV. RESULTS ANALYSIS

The insert test (on Fig. 6) shows that when inserting a  $10^9$  records with random keys, NoTree is 34 times faster than FractalTree and 56 times faster than LSM-tree. (The results do not include a representative of a B-tree index because its speed drops to under 500 rec/sec during the tests. After 5 days it cannot insert even half of the records.)

The high speed which the NoTree index achieves is a direct consequence of its philosophy – there are no connections, relations or subordination between the data intervals (Fig. 4D). Thus, a main limitation for the incoming data into the index is the throughput of the external storage device. In other technologies the indexing structure dominates by enforcing constraining relations between the intervals. Those relations chiefly define a *local merging strategy* whereas in NoTree we have a *global merging strategy* for the intervals, i.e. the choice which buffers to be merged is done between all buffers that are available in the index.

When there is no previous information for the distribution of the incoming keys in the index and the data is coming at a faster rate that can be indexed, one of the things that can be done is to postpone the actual ordering process of the records. That by itself leads to increase in the *levels of intervals* in LSM-tree, FractalTree and NoTree. Thus, after the insert test (in general) there will be more than one level of intervals in the indexing structures. And so the indexing task is actually brought down to effectively reducing the number of levels.

In order to make an estimate of how much time is needed for each one of the indices to convert its set of intervals into one level of intervals, it is needed all of the records to be read (in ascending order by their keys) right after the insert test completes. That forces the index – either tree based or non-tree based, to bring the structure into one that has only one level of intervals. The result from the immediate full read is shown on Fig. 7. The chart shows that the read speed of NoTree increases because as time progresses fewer levels of intervals remain.

It is clear from the total time for the complete test (*insert + immediate full read*) that the needed time for a complete execution of the whole test by NoTree is 10 times less than that of FractalTree and 16 times less than the time of LSM-tree.

A rough estimate can also be made about the time needed for the full ordering of the data. Because the indexing process in the different indexing technologies is distributed between the *Insert* and *Immediate Full Read* processes, the estimate can be made by subtracting the times for *generating, sequential write in a file* and *sequential read from the file* of all records from the total time. In our case, for  $10^9$  records and specified hardware this time (measured with the same application) is around 43 minutes. Thus, the actual time needed by NoTree for the complete ordering of  $10^9$  records is 12 times less than the time of FractalTree and 21 times less than LSM-tree.

It can also be noted that for every following data read – regardless if it is a point or range query, NoTree will increase its speed additionally compared to the other indexing technologies. NoTree will be naturally transformed into a linear structure from non-overlapping intervals after the full ordering of the data after the first read, whereas all other tree based indexing structures will have an internal height (because of their tree-like nature), regardless of the fact that they also have one level of intervals.

## I. CONCLUSION

The present article introduces a new external index data structure called NoTree, which can be defined as a next step in the development of index data structures. NoTree introduces a

new approach in terms of building the indices. Instead of building indices based on keys, NoTree builds the indices based on intervals from keys.

NoTree works by periodically performing a process of *selecting* and *merging* of data intervals, until non-overlapping intervals are left. The selecting and merging are done with the described *k-partitioning* of intervals and the subsequent *merging* of some of them. Thus, with the so described *k-partitioning* and *merging*, NoTree achieves a significant reduce of the number of I/O operations needed in the indexing process.

Apart from the more efficient indexing, NoTree also removes the need of subordination between the stored buffers. NoTree lacks a root and an idea for subordination that are common for the tree-based data structures, which limit the write and read speed. This gives possibilities for different parallelizations. In the process of working, NoTree can transform itself into an entirely linear structure of non-overlapping intervals, which increases the random and sequential read of records.

The article also proposes an *interval model* (view), which shows the conceptual differences between NoTree and the currently widespread tree-based indexing structures. The article also includes comparative tests and results.

The proposed cycle for *k-partitioning* and *merging* can also be used for effective external sort. The question for finding the minimal k-partition of a set of intervals remains open.

## REFERENCES

- [1] R. Bayer and E. McCreight, "Organization and Maintenance of Large Ordered Indexes", *Acta Informatica*, vol. 1, p. 173–189, 1972.
- [2] D. Comer, "The Ubiquitous B-Tree", *Computing Surveys*, vol. 11 No 2, June 1979.
- [3] D. E. Knuth., *Sorting and Searching*, volume 3 of *The Art of Computer Programming*, vol. 3, Massachusetts: Addison-Wesley, 1973.
- [4] L. Arge, "The Buffer Tree: A New Technique for Optimal", *Algorithms and Data Structures*, vol. 955, pp. 334-345, 1995.
- [5] L. Arge, "The buffer tree: A technique for designing", *Algorithmica*, vol. 37, pp. 1-24, September 2003.
- [6] M. A. Bender, M. Farach-Colton, J. T. Fineman, B. C. K. Y. Fogel and J. Nelson, "Cache-oblivious streaming B-trees.", *In Proceedings of the Nineteenth ACM Symposium on Parallelism in Algorithms and Architectures (SPAA)*, pp. 81-92, June 9–11 2007.
- [7] I. Tronkov, "WaterfallTree — External indexing data structure", in *Automation, Quality and Testing, Robotics, 2014 IEEE International Conference*, Cluj-Napoca, 2014.
- [8] I. Tronkov, "Mathematics behind WaterfallTree", in *International scientific conference UNITECH*, Gabrovo, 2014.
- [9] I. Tronkov, "WaterfallTree implementation details", in *International scientific conference UNITECH*, Gabrovo, 2014.
- [10] P. O'Neil, E. Cheng, D. Gawlick and E. O'Neil, "The log-structured merge-tree (LSM-tree)", *Acta Informatica*, vol. 33, pp. 351-385, June 1996.
- [11] F. Chang, J. Dean, S. Ghemawat, W. C. Hsieh, D. A. Wallach, M. Burrows, T. Chandra, A. Fikes and R. E. Gruber, "Bigtable: A distributed storage system for structured data.", *ACM Transactions on Computer Systems*, p. 26(2), June 200.
- [12] К. Куратовски, *Увод в теорията на множествата и топологията*, София: Наука и изкуство, 1979.

# Random Generation of Bent Functions on Multicore CPU Platform

Miloš Radmanović<sup>1</sup> and Radomir S. Stanković<sup>2</sup>

**Abstract** – Bent functions are an important class of Boolean functions since they are used in different kinds of cryptographic applications. Nevertheless, Bent functions are rare and difficult to determine, especially in the case of functions with many variables. During recent years, in order to improve its performance, scientific software has been ported to multicore CPU (Central Processing Unit) and GPU (Graphics Processing Unit) platforms. Thus, this paper proposes a parallel implementation of a well renowned algorithm for generation of Bent functions on multicore CPU platform using the MPI (Message Parsing Interface) framework. The algorithm is based on random discovering of Bent functions on the reduced search space in the Reed-Muller spectral domain. The experimental results show that the random generation of Bent function on multicore CPU platform is quite efficient in terms of the computation time.

**Keywords** – cryptography, Boolean function, Bent function, random generation, multicore CPU platform, MPI.

## I. INTRODUCTION

Bent functions were introduced by Rothaus in 1976 as Boolean functions with maximal nonlinearity [1]. They are important due to the applications in many areas such as combinatorics [2], coding theory [3], cryptography [4], and logic synthesis [5].

It is well known that all Walsh spectral coefficients of Bent functions have the same absolute value  $2^{n/2}$ , where  $n$  is number of variables of the function. They have the maximum possible value of nonlinearity equal to  $(2^{n-1} \pm 2^{n/2-1})$ , and they only exist for even number of variables. Extensive work on Bent functions has been done and various interesting results of researches have been brought out in respect of their generalization, construction, classification, etc.

Constructing of all Bent functions for a given number of variables is possible just for a small number of variables (less than 10) [6]. While general Bent functions are difficult to discover, specific Bent functions can be easily described [7]. It is also well known that from given Bent functions, new Bent functions with the same or greater number of variables can be constructed [8]. All known methods for constructing Bent functions are deterministic, and they do not provide any, for example cryptographic, quality to the generated function.

There are different methods for random discovering of Bent functions, and most of them are based on the Reed-Muller

expressions of Boolean functions [9]. A reason for determining Bent functions in the Reed-Muller domain is the efficiency of related algorithms in terms of time.

The algorithm for generation of Bent functions in Reed-Muller domain takes as its input the minimum and maximum number of non-zero coefficients in the Reed-Muller spectrum of orders that the Bent functions are allowed to have. Since the order of Bent functions is less or equal to  $n/2$ , where  $n$  is the number of variables, possible search space for random generation in the Reed-Muller domain is reduced.

Performing the related algorithm is a CPU time consuming task, so we have developed two independent implementations for a comparison, a single-core implementation using C++ and a multicore CPU implementation using MPI framework. Implementations on multicore CPU platforms are recognized as having the potential to considerably speedup or accelerate computing intensive algorithms over their equivalent single CPU core implementations [10]. The proposed multicore MPI implementation exploits possibilities for parallelism that can be found in the algorithm for generation of Bent functions in the Reed-Muller domain.

In the case of the random generation of Bent functions, we investigate this algorithm in the Reed-Muller domain for different minimum and maximum numbers of non-zero coefficients and for random possible orders of Bent functions. In this paper, we studied how the algorithm performances change when the multicore approach is applied. Restrictions in the Reed-Muller domain greatly influence the performance of the random generation of Bent functions. As the Boolean function size increases, the number of calculations extremely increases. For this reason we have experimented with small sizes of Boolean function with strong restrictions in the Reed-Muller domain.

The experimental results obtained on a multicore CPU platform with 8 cores show performance speedups for some benchmarks of maximum 4 times. This results confirm that the application of the proposed implementation using MPI framework leads to significant computational speedups over traditional C++ implementations processed on single CPU.

This paper is organized as follows: Section 2 shortly introduces theoretical background. In Section 3, the algorithm for discovering Bent functions in the Reed-Muller domain is discussed. Section 4 offers some details for multicore CPU implementation of this algorithm using MPI framework. The experimental results are presented and discussed in Section 5. The closing Section 6 summarizes the results of the research reported in this paper.

<sup>1</sup>Miloš Radmanović is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: milos.radmanovic@elfak.ni.ac.rs

<sup>2</sup>Radomir S. Stanković is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: radomir.stankovic@gmail.com

## II. THEORETICAL BACKGROUND

**Definition 1** For a Boolean function  $f$  defined by the truth-vector  $F = [f(0), f(1), \dots, f(2^n - 1)]^T$ , the Walsh spectrum  $S_{f,W} = [S_{f,w}(0), S_{f,w}(1), \dots, S_{f,w}(2^n - 1)]^T$  is defined as [11]:

$$S_{f,W} = W(n)F, \quad (1)$$

where,

$$W(n) = \bigotimes_{i=1}^n W(1) \quad (2)$$

where  $\otimes$  denotes the Kronecker product, and

$$W(1) = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \quad (3)$$

is the basic Walsh matrix.

The Walsh transform is a self-inverse transform up to the constant  $2^{-n}$  that is used as the normalization factor when defining the Walsh transform and its inverse. It is an important mathematical tool for the analysis of Boolean functions. It can be shown that with (1,-1) encoding of function values, the Walsh coefficients are even integers in the range  $2^{-n}$  to  $2^n$ .

**Definition 2** A Boolean function  $f$  in (1,-1) encoding is called Bent if all Walsh spectral coefficients  $S_f$  have the same absolute value  $2^{n/2}$ .

Note that a Bent functions can be characterized by the positive polarity Reed-Muller form, leading to a correspondence with the upper bound of the algebraic degree of that form.

A positive polarity Reed-Muller form (PPRM) is an exclusive-OR of AND product terms, where each variable appears uncomplemented. Any Boolean function  $f$  can be represented by the PPRM form in matrix notation defined as [11]:

$$f(x_1, x_2, \dots, x_n) = X(n)R(n)F \quad (4)$$

where

$$X(n) = \bigotimes_{i=1}^n \begin{bmatrix} 1 & x_i \end{bmatrix} \quad (5)$$

and

$$R(n) = \bigotimes_{i=1}^n \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \quad (6)$$

where addition and multiplication are modulo 2,  $R(n)$  is the positive Reed-Muller transform matrix of order  $n$ , and  $R(1)$  is the basic positive Reed-Muller transform matrix.

The PPRM spectrum  $S_{f, RM}$  is calculated as [10]:

$$S_{f, RM} = R(n)F. \quad (7)$$

The elements of  $S_{f, RM}$  are coefficients in the PPRM expressions for any Boolean function [11]:

$$f(x) = a_0 \oplus \sum_{i=1}^n a_i x_i \oplus \sum_{1 \leq i < j \leq n} a_{ij} x_i x_j \oplus \dots \oplus a_{12\dots n} x_1 x_2 \dots x_n \quad (8)$$

where  $\Sigma$  denotes modulo 2 summation.

The algebraic degree or the order of nonlinearity of a Boolean function [11]  $f$  is a maximum number of variables in a product term with non-zero coefficient  $a_k$ , where  $k$  is a subset of  $\{1, 2, 3, \dots, n\}$ . When  $k$  is an empty set, the coefficient is denoted as  $a_0$  and is called the zero order coefficient. Coefficients of order 1 are  $a_1, a_2, \dots, a_n$ , coefficients of order 2 are  $a_{12}, a_{13}, \dots, a_{(n-1)n}$ , coefficient of order  $n$  is  $a_{12\dots n}$ . The number of all coefficients of order  $i$  is  $\binom{n}{i}$ . The PPRM

coefficients are divided into order groupings according to the number of ones in the binary representation of its index in the spectrum.

Algebraic degree of Bent functions is at most  $n/2$  for  $n \geq 4$  [6]. Thus, the maximal number of PPRM coefficient of Bent functions is:  $\sum_{i=0}^{n/2} \binom{n}{i}$ .

The number of Bent functions of a given number of variables is not known. Therefore, the number of Bent functions is at most  $2^{2^{n-1} + \frac{1}{2} \binom{n}{n/2}}$  which is much less than all Boolean functions  $2^{2^n}$  [6].

For the PPRM transform, we need an inverse transform to get back from the Reed-Muller domain. Since the Reed-Muller transform matrix  $R(n)$  is a self-inverse matrix over  $GF(2)$ , the forward and inverse transform are given by the same matrix [11].

The Walsh and Reed-Muller the transform matrices, expressed in (3), and (6) respectively, can be factorized in different ways yielding different fast transform algorithms, the so-called FFT-like algorithms [11].

Figure 1 shows the elementary butterflies operations (flow-graphs) for the Reed-Muller, and the Walsh basic transform matrices, respectively.

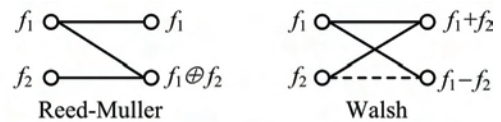


Fig. 1. The elementary butterfly operations for the basic Reed-Muller and Walsh transform matrices

The Cooley-Tukey class of algorithms is based on the Good-Thomas factorization which originates from the Kronecker product structure of the transform matrix [11].

Figure 2 shows the flow graphs of the fast Cooley-Tukey spectral transform algorithm for the computation of the Walsh spectrum of a three-variable logic function  $f$  given by the



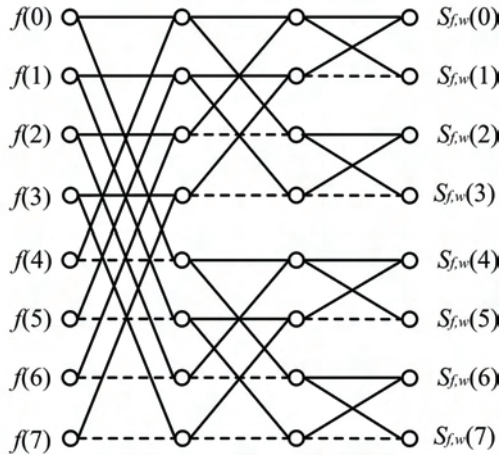


Fig. 2. The flow graphs of the Cooley-Tukey spectral transform algorithm for computing the Walsh spectrum of a tree-variable logic function.

truth- vector  $F = [f(0), f(1), \dots, f(7)]^T$ . This algorithm is highly exploited in computing the inverse PPRM transform for discovering Bent functions in the Reed-Muller domain.

Fast spectral transform algorithm reduces computational complexity of a spectral transform from  $O(N^2)$  to  $O(M \log_2 N)$ , it is an extremely effective tool in scientific computing.

### III. RANDOM GENERATION OF BENT FUNCTIONS IN REED-MULLER DOMAIN

The algorithm for the generation of Bent functions in the Reed-Muller domain takes as its input the number of function variables and the minimum and maximum number of non-zero Reed-Muller coefficients of any order that is allowed. Since the order of Bent functions is less or equal to  $n/2$ , the number of non-zero PPRM coefficients is limited and the positions of the coefficients in the PPRM spectrum are restricted. For example, Table 1 gives the limitation of number of the non-zero PPRM coefficients in relation to the total number of coefficients of Bent functions for the number of variables ranging from 8 to 14.

TABLE I  
LIMITATION OF THE NUMBER OF THE NON-ZERO PPRM COEFFICIENTS OF BENT FUNCTIONS

Num. of function variables	Limitation / total PPRM coefficients
8	163 / 256
10	638 / 1024
12	2510 / 4096
14	9908 / 16384

These restrictions ensure generation possibility, since they certainly reduce the possible search space for random generation in the Reed-Muller domain. As the Boolean

function size increases, the possible search space (see Table 1) also extremely increases.

An outline of the algorithm for the generation of Bent functions in Reed-Muller domain is given as Algorithm 1.

#### Algorithm 1

- 1: Set the number of function variables and the minimum and the maximum number of the non-zero Reed-Muller coefficients (the maximum number of coefficients should be less than limitations).
- 2: Random generation of the number of the non-zero coefficients in Reed-Muller spectrum (between minimum and maximum).
- 3: Random generation of the positions of the non-zero coefficients in Reed-Muller spectrum. Positions of coefficient are related with the order of coefficients and their generation is determined by the number of ones in the binary representation of the positions in the spectrum (the number of ones is less or equal to  $n/2$ ).
- 4: The computation of the truth-vector of a Boolean function is done by using the flow graph of the inverse fast PPRM transform of Reed-Muller spectrum.
- 5: (1,-1) encoding of a Boolean function.
- 6: Fast testing if the first Walsh coefficient has the absolute value  $2^{n/2}$ . Otherwise go to the step 2.
- 7: Fast testing if the second Walsh coefficient have the absolute value  $2^{n/2}$ . Otherwise go to the step 2.
- 8: Fast testing if the last Walsh coefficient have the absolute value  $2^{n/2}$ . Otherwise go to the step 2.
- 9: Testing if all values of Walsh spectrum have the same absolute value  $2^{n/2}$ . The computation of Walsh spectrum is using the flow graph of the fast Walsh transform. Otherwise go to the step 2.
- 10: Obtain random Bent Boolean function having extreme nonlinearity properties.

### IV. IMPLEMENTATION OF RANDOM GENERATION OF BENT FUNCTIONS ON MULTICORE CPU PLATFORM

For multi-core CPU architectures, the model of parallel processing is based on a large number of processor cores with the ability to directly address into a shared RAM memory. This organization of computations allows to have a large number of processes performing the same operations on different data simultaneously. Process of communication using a network is much slower than the process of communication using the shared memory [12].

The algorithm for random generation of Bent function in the Reed-Muller domain has a large degree of parallelism. Steps 2 to 9 of the Algorithm 1 are computationally independent and according to this, its implementation on multicore CPU platform is convenient. A fundamental step in parallelizing of this algorithm on multicore CPU is the mapping to processor cores of arrays representing Reed-Muller spectrums, that are used for random generation of functions as well as Bent detection. It should be also noticed

that random generation of Bent function can be very CPU time consuming, since the computation of the inverse fast PPRM transform (step 4 of the Algorithm 1) are exponential in the number of variables in the function.

## V. EXPERIMENTAL RESULTS

The MPI standard has become a widely used standard for parallel programming framework [13]. For comparison purposes, we developed referent single-core C++ and MPI implementation on multi-core CPU platform of algorithm for random generation of Bent function. Note that, for the largest Boolean functions, computations were not performed on multicore CPU, due to the computation time limitations of 30 minutes.

The computations are performed on an Intel i7 CPU at 3.66 GHz with 12 GBs of RAM. The quad-core CPU that is used is hyper-threading, yielding 8 logical cores with 8 MB of smart cache memory.

Table 1 shows computation performance of algorithm for random generation of Bent function using referent single-core C++ and MPI implementation on multi-core CPU platform.

The presented computational times represent average values for ten executions of implementations for each number of function variables and min. and max. of non-zero Reed-Muller coefficients. From data in Table 1, it can be seen that, on this multi-core CPU platform, for all the computations, MPI implementation of the algorithm reduces computation times when compared to the single-core C++ implementation.

## VI. CONCLUSION

Methods for generating Bent functions are deterministic, and they do not provide any, for example cryptographic, quality to the generated function. The approach for finding a non-deterministic Bent functions is most often based on random discovering Bent function in Reed-Muller domain. It should be noticed that finding the Bent function can be very CPU time consuming, since the search space in Reed-Muller domain are extremely exponential in the number of variables of the function. However, computing power can be substantially increased through the exploitation of the parallelism on multi-core CPU platform.

In this paper, we investigated parallelization of algorithm for random generation of Bent function on multi-core CPU platform. The parallel MPI implementation of this algorithm is convinient, since the algorith has a large degree of parallelism. Experimental results confirm that exploiting multi-core CPU platform can help improve the computation performances of this algorithm.

We can conclude that, when processing time is a critical parameter, the algorithm for random generation of Bent functions in Reed-Muller domain should be performed on the multi-core CPU platform. Future work will be on extension of the proposed technique to various other multi-processing platforms.

TABLE I  
COMPUTATION PERFORMANCE OF ALGORITHM FOR RANDOM  
GENERATION OF BENT FUNCTIONS USING REFERENT C++ AND MPI  
IMPLEMENTATION

Num. of function variables	Min. and max non-zero RM coefficients	Avg. computation time [s]	
		CPU	mCPU
8	1 - 100	1.443	0,288
8	1 - 163	1,792	0,509
10	1 - 100	57,201	32,175
10	1 - 200	32.103	11.108
10	1 - 300	27.883	10.002
10	1 - 400	39.751	10.145
12	1 - 100	> 30 min.	> 30 min.

## ACKNOWLEDGEMENTS

The research reported in this paper is partly supported by the Ministry of Education and Science of the Republic of Serbia, projects ON174026 (2011-2016) and III44006 (2011-2016).

## REFERENCES

- [1] O. Rothaus, "On Bent Functions", *Jourbal of Combin. Theory Ser. A*, vol. 20, pp. 300–305, 1976.
- [2] T. Helleseth and A. Kholosha, "Bent Functions and Their Connections to Combinatorics", in S. Blackburn, S. Gerke, and M. Wildon, editors, *Surveys in Combinatorics 2013*, pp. 91-126, Cambridge University Press, 2013.
- [3] O. Logachev, A. Salnikov, and V. Yashchenko, *Boolean Functions in Coding Theory and Cryptography*, American Mathematical Society, 2012.
- [4] N. Tokareva, *Bent Functions, Results and Applications to Cryptography*, Academic Press, 2015.
- [5] T. Sasao, J. Butler, and M. Thornton, *Progress in Applications of Boolean Functions*, Morgan and Claypool Publishers, 2010.
- [6] P. Langevin and G. Leander, "Counting all Bent Functions in Dimension Eight 99270589265934370305785861242880", *Designs, Codes and Cryptography*, vol. 59, pp. 193-201, 2011.
- [7] A. M. Youssef, and G. Gong, "Hyper-bent Functions", *Advances in Cryptology-EUROCRYPT 2001, Lecture Notes in Computer Science*, vol. 2045, Springer, pp. 406–419, 2001.
- [8] J. Seberry and X. Zhang, "Constructions of Bent Functions from Two Known Bent Functions", *Australasian Journal of Combinatorics*, vol. 9, pp. 21-34, 1994.
- [9] N. Y. Yu, and G. Gong, "Constructions of Quadratic Bent Functions in Polynomial Forms", *IEEE Transactions on Information Theory*, vol. 52, no. 7, pp. 3291-3299, 2006.
- [10] G. E. Karniadakis, R. M. Kirby, *Parallel Scientific Computing in C++ and MPI: A Seamless Approach to Parallel Algorithms and their Implementation*, Cambridge University Press, 2003.
- [11] T. Sasao, and M. Fujita, *Representations of Discrete Functions*, Boston: Kluwer Academic Publishers, 1996.
- [12] C. Hughes, T. Hughes, *Professional Multicore Programming: Design and Implementation for C++ Developers*, Wiley-Interscience, 2011.
- [13] W. Gropp, E. Lusk, and A. Skjellum, *Using MPI: Portable Parallel Programming with the Message-Passing Interface*, MIT Press, 1999.

# Education of Patients by using Mobile Learning System

Dusan Tatić<sup>1</sup> and Dragan Janković<sup>2</sup>

**Abstract** – *Multimedia systems are a suitable platform for various educational purposes. The paper presents a multimedia system for patient education implemented on mobile devices. The system architecture allows application of different programs for patient's education in different healthcare institutions. The universality of the system is achieved by including a Content management subsystem which allows an easy change of content was created for the intended applications which make. Mobility gives the opportunity to the patients to consult the content at any time and any place. The described mobile system is applied at the Clinic of Endocrinology, Clinical Center Nis, and shown in the paper.*

**Keywords** – Mobile health (mHealth), medical informatics, education, multimedia, content management.

## I. INTRODUCTION

It is recognized that an adequate education of patients greatly contributes to the successful treatment. The insufficient number of health professionals in healthcare institutions in Serbia, which are constantly overloaded with the number of patients they have to deal with, does not allow to paid due attention and time to education of patients.

Information technology could have a significant impact on improving patient education programs. Different systems have been developed for this purpose [1]. Many such systems are designed based on touch screen technology implemented as information kiosks [2], as well as systems that are used on mobile devices [3]. A large number of projects, initiated in terms of health promotion, are implemented on smart mobile platforms [4]. The usage of mobile devices allows patients to easily obtain the required medical information.

A previous version of this educational system has been developed for desktop machines and was intended to be installed as information kiosks for patients in healthcare institutions [5]. System was well accepted by patients who found it useful during their medical treatment. An interest to have such a system accessible more frequently and at any time was recognized. This desire was a motivation for the transformation of the system from static to mobile educational platform oriented to individual and specific usage according to a patient need. Primarily requirements of the system were modified and improved for this purpose. In this paper, we propose Mobile multimedia system solutions for patient education in the form of multimedia touch screen application that expresses a high level of adaptability for applications in various healthcare institutions. The system is designed and

implemented on an analysis based on real needs of healthcare institutions in Serbia, as well as on a study of current trends in this area abroad [6]. The primary requirement is that the system should be independent of the medical field for which the training is done. A special attention is paid to the simplicity of creating new and maintaining of the existing educational contents.

The paper has the following structure. Section II discusses functionality of the system. The architecture of the system is discussed in detail in Section III. Implementation of the system is illustrated and related details explained on the example of a system for patient education at the Clinic of Endocrinology in Section IV.

## II. FUNCTIONALITY OF THE SYSTEM

The proposed system for patient education is realized as a mobile platform for preparation (by medical doctors) and usage (by patients) of various educational contents in medicine. The system should serve for education of a wide range of population structures in order to implement different knowledge programs in health care. The system is designed to be independent of the branch of medicine where it is applied.

The proposed system for patient education consists of two main parts. The user's part is intended for providing information and guidance for patients, while the other part supports creating and management of the educational content.

Accordingly, the basic functionality is retained as in the preliminary version reported in [5], but it is enriched with additional options for individualization of the usage by taking into account specific requirements of each patient. Therefore, we will show some similarities and differences between the systems. Main differences between these two systems are given in Table I and discussed in later text.

TABLE I  
KIOSK VS MOBILE SYSTEM

Kiosk system	Mobile system
Different patient / Multiple patients	Personalized application
Static system / implemented in the waiting rooms	Mobile system / patient can bring it with him
Content management of one application	Content management of different applications
Multimedia content prepared for different patients/ large content	Material adopted to the needs of a particular patient / reduced content

<sup>1</sup> Dušan Tatić is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: [dule\\_tatic@yahoo.com](mailto:dule_tatic@yahoo.com)

<sup>2</sup> Dragan Janković is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: [dragan.jankovic@elfak.ni.ac.rs](mailto:dragan.jankovic@elfak.ni.ac.rs)



**Users part.** When exporting the former system for mobile devices, the following specific parameters were taken into account. The basic system requirements on client side are:

- The system is intended for patients of all ages and different education levels.
- The system should be easy to use, the patient without previous training can in an intuitive manner manipulate with the content.
- It should provide greater interactivity between the patient and the system. The patient according to his needs select contents of his interest.
- System must enable the viewing of multimedia educational content, including video materials, images, and text.

In comparison with the previous system which was static, positioned in the waiting rooms, system should be portable and implemented on various mobile platforms (Table I). Therefore, it is implemented on mobile devices so that can be used at any time and every place. Due to that, the usage of the education system is adapted to daily routine of patients.

Based on previous requirements the system is realized in the form of a mobile multimedia touch screen application. A user, the patient, has available a touch screen which displays educational content. Built-in interactive menu allows the patient to select by a simple touch of a finger the desired content according to his needs. The system can be used by the patients in accordance with their available time and personal interests because it is implemented on a mobile device.

**Management part.** Content management is part of the system that allows the system administrator to set and regulate the educational content. This part of the system is characterized by the following features:

Universality – the system is designed to be independent of the medical field for which the education is conducted. This allows the contents to be completely adapted to the healthcare institution which uses the system.

Adaptability – the usage of the system is entirely adapted to the needs of healthcare professionals. For them, it is not necessary to have a preliminary IT knowledge or experience in order to be able to create the contents or maintain and update the system. The system has integrated templates which are used to define the layout of the screen.

Efficiency - maintenance (add/delete/change) of the contents is simple. The concept of the system, is based on the use of databases, allows a simple and easy system maintenance. It also enables to create and display a unlimited number of educational facilities.

Main differences with previous system management concerns the content implementation (Table I). Each patient requires information according to his medical treatment. Due to that only specific material will be transferred to the patient device instead of static system that had all materials for different patients. The implementation requires smaller size according the material but many installations for the different users and their mobile platforms.

### III. SYSTEM ARCHITECTURE

The mobile system for patient education consists of three components: mobile application, external storage, administration subsystem (Fig. 1). These components are designed to work independently of each other. Thanks to this approach, where the content, design, and control are independently implemented, the universality of the system is achieved. In this way, the system can be easily adopted to specific needs of any health institution with no restrictions.

The system is realized using the Adobe Integrated Runtime (AIR), which is intended for the development of multimedia applications. The Adobe AIR is a cross platform runtime that enables to create a rich experience for Windows and Mac OS desktops as well as iOS and Android mobile devices. For the implementation of the system Adobe Flash and ActionScript 3 programming language are used.

#### A. Mobile application

The mobile application is designed to establish communication between the user and the system. This component is organized in three modules: *User interface*, *Interaction*, and *Display*.

*User interface* interprets and displays the graphics design of the application from data obtained from a database. This refers to elements such as logo, address, layout and appearance of the user menu screens.

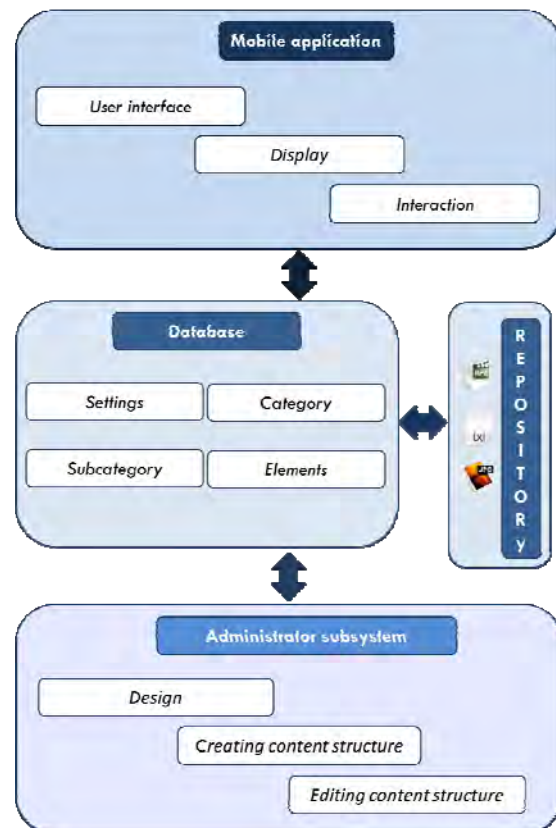


Fig 1. Architecture of the system



*Interaction module* establishes a relation of the user with the system and connection of the system to a database. Based on the user selection through the user interface, this module issues the appropriate query to the database. Result of the database query is to obtain the information about selected multimedia educational content which is stored in the *Repository*.

Components for displaying multimedia content have been implemented in the module *Display*. To view the video material, a video player is created. Video player allows displaying a variety of file formats (mp4, mpeg, flv, mov). Other information can be displayed in the form of text using html and txt format.

### *B. External storage*

External storage represents the educational material stored at the patient's mobile device. It consists of an SQL database that describe the material stored in the *Repository*.

The SQLite relational database is selected due to its simplicity and easiness of communication with the application and it is supported by the Adobe AIR software tool. The database is defined by four tables: Settings, Categories, Subcategories, and Elements. Table Settings determines the design of the application (menu appearance, name, logo and other visual elements). Tables Category, Subcategory, and Elements define the hierarchical structure of the user menu.

Multimedia contents are stored in the *Repository* (Fig. 1) which is connected to the database via the table elements. Method of forming a repository is explained in the next section.

### *C. Administrator subsystem*

Its function is to form a system and to create/maintain the educative contents. Installation is done on the administrator computer and connected to the device used for medical education. Upon establishing connection with external storage on mobile device the administrator of the system can organize and copy the content according to the patient needs. The organization of content is realized by means of the following modules: Design, Creating, and Editing the content structure. Usage of these modules is done through the Administrator Panel

**Design.** The module Design is used to define visual component of the system. By using this module, the table Settings is formed. This table contains information about the type of template that is selected, the name of the system, and logo. Thus designed table is used by the module User interface that can retrieve and interpreted data on the client side. Templates represent the layout of the menu on the user's screen. Template uses the landscape orientation of the mobile device. Currently, templates are organized for tablet devices. Selection of the templates depends on the structure of the educational content and can be changed.

**Creation.** This module allows definition of the hierarchical structure of the content that is organized in categories and subcategories. Categories can be created by giving desired

names. Each Category can be assigned to several Subcategories with their name. Categories and sub-categories represent a hierarchical structure of the system menu. They are used to get access to the desired educational content. The number of categories and subcategories is unlimited and depends of the user screen size. Each subcategory is connected with the Element that represents a specific educational content. In this way, the structure of the educational content of the system is formed.

The described activities are realized through the Administrator Panel. In this way, they form the corresponding tables in the database. Also, with these activities it is possible to remove or modify Categories, Subcategories, and Elements. All selected actions automatically update the corresponding tables in the database.

**Editing.** Through this module, the *Repository* is formed copied to the external storage of the mobile device by using the Administrator panel. The formation is performed in the way that each defined Element joins some of the multimedia content. In this manner, an unambiguous correlation is determined by all the elements in the system. Between SQL database and *Repository* elements on the external storage of mobile device.

## **IV. EDUCATION OF PATIENTS WITH DIABETES**

Implementation of the described system will be explained on the example of the Clinic for Endocrinology application. For the purposes of the Clinic for Endocrinology, Clinical Center Niš, we developed a mobile multimedia interactive system for the education of patients with diabetes.

The system is realized in the Adobe Air SDK software tool for IOS and Android mobile devices. This system is implemented and tested on the Android software platform by using Asus transformer eee pad tablet. The goal is that instead of conventional written instructions, the patient is guided by an interactive application on his mobile device. Bu using it, patients are able to follow the medical education material and raise their healthcare knowledge in a modern and simple to use way.

The educational material is prepared and determinate for each patient in cooperation with medical experts. A detailed analysis of educational materials in the context of system design is extremely important in order to define the optimal layout of the user screen. Implementation of the system consists of several steps.

### *A. Step 1*

Step 1 serves to define the visual elements of the system. Using the module Design, which is a part of the Administrator subsystem, template was selected. The choice of templates in this case was influenced by the fact that the number of titles (categories) is smaller than in the captions (sub-categories). In this manner, the basic layout of the application was determined to better arrange elements on the user screen. Also, as a part of this step, the table Settings was formed in the database.

## B. Step 2

In this step, the construction of the hierarchical structures of the system is done. The module Creation structure is used to form the structure of the system. Categories and the corresponding subcategories are defined in accordance with the educational content. Also, in this step, tables category, subcategory and elements are formed in the database.

## C Step3

Step 3 is used to connect the system structure with the educational content. Each element is linked to the corresponding multimedia content by using the Edit content module. In this way, the Repository is formed and thereby an unambiguous relationship is determined between all elements in the system.



Fig 2. User interface with categories, subcategories and elements

As a result of the action described above, the user screen is formed as shown in Fig. 2. The layout of multimedia elements is displayed on the user's screen in the form of images. These images have the appropriate title and an icon which defines the type of multimedia content (text, image, or video). To view educational content, we created a video player that enables manipulation of various formats (mp4, mpeg, flv, mov). Video is displayed in a dedicated window intended for multimedia elements (Fig. 3). In addition, it is possible to completely control the video content. To view the textual content, txt and html files are used.



Fig 3. User interface with selected video

## V. CONCLUSION

In this paper we present a mobile multimedia system for patient education. The system is implemented as an independent platform for use of educational content in medicine in form of multimedia handheld device. The hardware part of the system is based on smart phone technology for various platforms (IOs, Android, and Windows) that includes a touch screen while the software is implemented by using Adobe Flash software tool binding Adobe Integrated Runtime (AIR) application. The architecture of this system is organized into functional components which assures adaptability to different requirements for its formation. The universality of the system is achieved by implementing a content management subsystem. Educational content can be presented in various multimedia formats (video, text, images). A purposely designed video player enables viewing the content in various formats. To view the textual content, txt and html files are selected.

A pilot version of the proposed system is implemented at the Clinic for Endocrinology, Clinical Center Nis. The functionality and universality of the system is tested by implementing specific requirements specified by the medical experts in this clinic.

After simple modifications, the proposed can be applied in other spheres of education (schools, municipalities, public administration, etc.).

## ACKNOWLEDGEMENT

The work presented in this paper was supported by the Serbian Ministry of Education and Science (project III 044006).

## REFERENCES

- [1] Troshani, Indrit, and Nilmini Wickramasinghe. "e-Health trends." In *Pervasive Health Knowledge Management*, pp. 57-66. Springer New York, 2013.
- [2] A. Joshi, and K. Trout, "Heath Kiosks as an Equal Opportunity Resource for Better Health: A Systematic Review," In *GLOBAL HEALTH 2012, The First International Conference on Global Health Challenges*, pp. 66-70, 2012.
- [3] Martínez-Pérez, Borja, Isabel de la Torre-Díez, and Miguel López-Coronado. "Mobile health applications for the most prevalent conditions by the World Health Organization: review and analysis." *Journal of medical Internet research* 15, no. 6 (2013).
- [4] Malvey, Donna, and Donna J. Slovensky. "mHealth Products, Markets, and Trends." In *mHealth*, pp. 65-94. Springer US, 2014.
- [5] Tatić, Dušan, Dragan Janković, and Dragan Tatić. "Multimedia system for the education of patients", *Infoteh*, 662-665, 2014.
- [6] C. Liu, Q. Zhu, K. A. Holroyd, and E. K. Seng, "Status and trends of mobile-health applications for iOS devices: A developer's perspective," *Journal of Systems and Software*, vol. 84, no. 11, pp. 2022-2033, 2011

# Versatile Program Model of Avatar Face

Milan Adamović, Dragan Ivetić<sup>1</sup>, Aleksandar Kupusinac<sup>2</sup>

**Abstract** – Human facial models can be abstracted and coded in electronic form that is suitable for a variety of purposes. This paper describes the process of developing a program model of the human face. The model is based on a parameterized curves and topological guards. Special attention is given to the way the human mind perceives faces, as well as differences between genders, ethnic and age groups. The produced model is then implemented and tested in a facial composite modeling application used at our faculty.

**Keywords** – avatar, face, program model, facial composite.

## I. INTRODUCTION

Avatars have become commonplace in video games and on the Internet [1]. Use of avatars includes, but is not limited to identification, or, on the contrary, protection of users' identity in virtual social networks. Avatars can be made in different ways, in different styles (ranging from stylized to realistic), can represent a person's entire body, or only their face, and can be 2D or 3D. Avatars can be composed from templates and premade elements, or drawn by hand. These characteristics depend on the avatar's domain of application.

A versatile program model of the human face was developed by students of Animation in Engineering for the purpose of coding practice. Most existing program models of avatar face use simple premade shapes as building elements, which severely limits the number of available combinations. This approach ensures consistency between the avatar and other graphical interface elements in an application. While hand drawn avatars do give the final user more creative freedom, they also require certain skills to create, and may clash with the rest of the interface. The model described in this paper is vector based, and avoids the topological guard imposed by a small number of available premade shapes, but also contains safety mechanisms to prevent deformations in the final product and not burden the user with needless information.

The paper is structured as follows: In section one we introduce the subject of avatars and look over some existing areas of application, then in section two, we describe perceptual characteristics of human facial features, by the brain and the ways in which it differs from computer modeling, then we move on with abstraction of those facts into a form more suitable for a program model as well as the differences between it and , which we follow up with a example of the model's implementation into a facial

composite application in section four, and finally conclude the paper in section 5 with an evaluation of the models usability.

## II. PERCEPTUAL CHARACTERISTICS OF FACES

### A. Process of facial perception

The intuitive facial processing happens in the right hemisphere of the brain, while the conscious processing happens in the left [2]. First, the right hemisphere roughly evaluates the proportions of facial features, and compares them to faces the observer knows. Afterwards, the left hemisphere processes individual details, such as hair and eye color, scars etc. The mole on the face of Marilyn Monroe is a distinguishing feature, and it's presence or absence in a depiction on Fig. 1 determines whether the person is supposed to be Marilyn Monroe, or just any blond blue eyed woman [3].

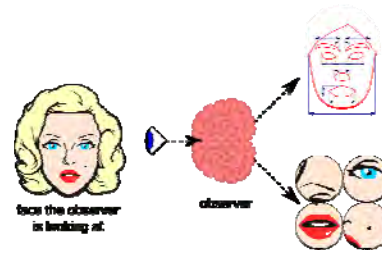


Fig 1. The face of Marilyn Monroe perceived by the brain hemispheres.

The facial recognition mechanism of the human brain is effective during tasks that include recognizing persons, but is not helpful when it comes to painting, modeling and similar tasks.. Certain individuals are suffering from a condition called prosopagnosia [4], that hinders the intuitive facial perception. Such persons can be trained to get their left brain hemisphere to completely take over the process of facial recognition. Training for persons with prosopagnosia involves observing and analyzing individual facial features and remembering certain parameters (e.g. typical chin shapes: square, round, spiked...) characteristic for certain groups of people (genders, races, age groups, etc. For example square shaped chins are more common on men than women) Persons who are trained to consciously analyze faces, but don't suffer from prosopagnosia achieve more accurate results when modeling human faces [8].

Such analytical approach to facial recognition makes it possible to consciously perceive and describe individual facial features, making them easier to model. The characteristics described in section 2.B describe facial features in conversational language, while section 3 deals with their abstraction into a program model.

<sup>1</sup> Dragan Iveticis with the Faculty of Technical Sciences at the University of Novi Sad, Trg Dositeja Obradovica 6, Novi Sad 21000, Serbia, E-mail: ivetic@uns.ac.rs

<sup>2</sup> Aleksandar Kupusinac is with the Faculty of Technical Sciences at the University of Novi Sad, Ohridski Blvd, Sofia 1000, Bulgaria.



## B. Genders, races and age groups

An important factor that contributes to easier differentiation between genders, races and age groups is neoteny [5], which is the ability of living creatures that determines what amount of youthful features adult individuals get to keep (Fig. 2, far left). Species with higher degree of neoteny tend to look younger and more feminine, while species with little to no neotenous features tend to look rugged and threatening. Most people fall into one of three basic ethnic groups (Fig. 2, center and right). Human beings are very neotenous compared to other species. Persons with neither neotenous, or non-neotenous faces are called androgynous. Gender of androgynous persons is difficult to tell from just quickly glancing at their faces.



Fig 2. Genders (far left, no color fill), three biggest ethnic groups depicted using the program model (center is Asian, followed by Caucasian and African), as well as signs of aging (far right) also depicted using the program model

Wrinkles can be easily added to the model as simple curves, as well as gray hair, while the droop of the facial muscles can be added with a topological guard (Fig. 2, far right).

he shapes described in conversational language in this chapter are abstracted and coded into a program model described in chapter 3. Differences (e.g. a round chin versus an angular chin) are broken up into sets of individual that affect certain topological areas of the face (e.g. chin width, chin height, etc.) and given descriptive names to ease the implementation for first time users.

## III. FACIAL FEATURES PARAMETRIZATION

The more control the user is given over the model, the more complex the model will have to be. Achieving a usable compromise between these two characteristics in a program model is the end goal [6]. Data in the model can be sorted into two categories:

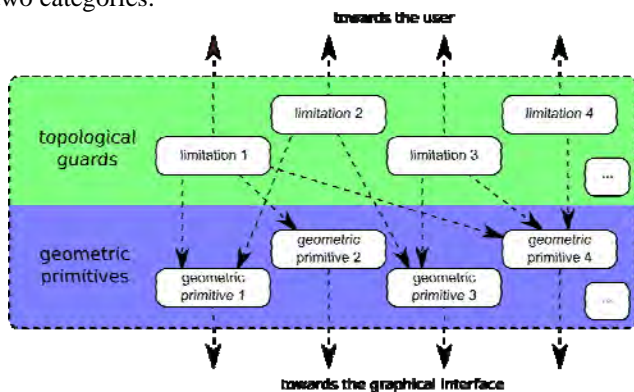


Fig 3. Hierarchy of data in the model

1. Geometric primitives that are used to model the face (2D vertex, lines, polygons, etc.) One of the goals of the model is to make manual manipulation over the geometric primitives as low as possible.

2. Topological guards that provide a mechanism for the controlled composition of avatar's face. Topological guards themselves are decimal numbers normalized to a range from 0 to 1. Topological guards make the model robust by preventing input of wrong parameters that can cause deformation of avatars' faces [7].

The graphical primitives in the model are visualized by Bezier's curves (figure 5). Green points are endpoints, while dark purple crosses and bright orange triangles are control

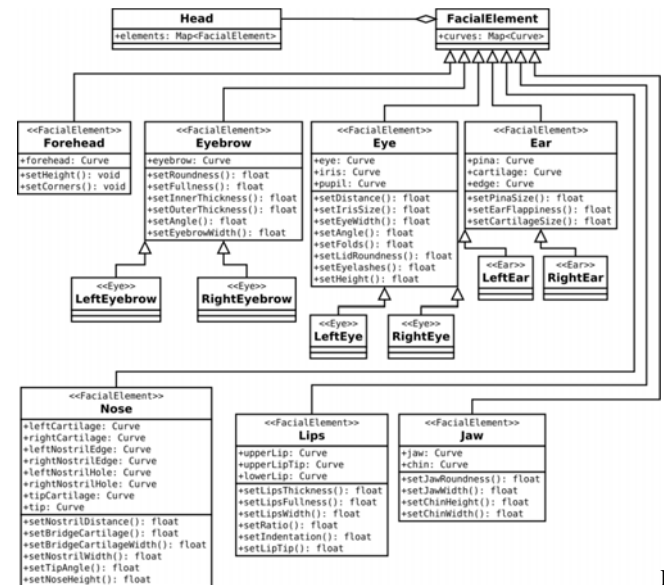


fig. 4. Class diagram of the facial elements hierarchy

points. Curves themselves are encapsulated in classes that represent individual facial features. Common parameters are position, rotation and scale.

Facial elements the model describes are Forehead, Eyebrows, Eyes, Ears, Nose, Lips, Jaw, Wrinkles, Scars, Scalp hair and Facial hair. Despite the fact that accessories (glasses, jewelry, etc.) are not a part of the face, the model includes them, as well.

The basis of the model is shown on figure 3. All facial elements inherit FacialElement that contains a collection of curves. Using a collection of curves rather than individual curves is to ensure that certain elements contain several separate lines (e.g. nostrils).



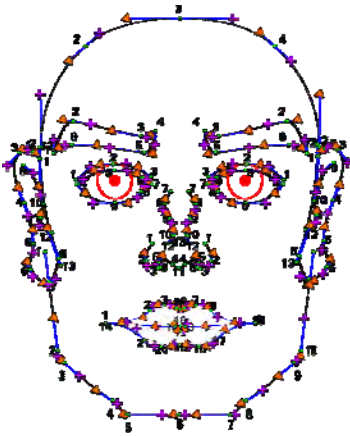


Fig 5. Face shown as a collection of Bezier curves.

Graphically, the face depicted through the use of Bezier curves (figure 4). The figure shows the face without the use of fill colors to keep focus on the shapes. Elements such as hair and accessories are not included here to avoid overcrowding. They were not coded using topological guards, but as collage elements that can be swapped in and out and replaced as needed.

Common parameters (topological guards) for all elements are: position, rotation and scale. More specific parameters can be derived from the basic three, but they only affect specific groups of nodes (Fig 6). Position, angle and scale of nodes can be affected using different mathematical equations that depend on topological guards. Figure 6. shows the process of a curve alteration when the value of a topological guard is changed from 0 to 1 (all topological guards are normalized to 0-1 range). The dark line shows the look of the curve when the limitation is set to 0, while the lightest line shows the curve when the limitation value is set to 1.

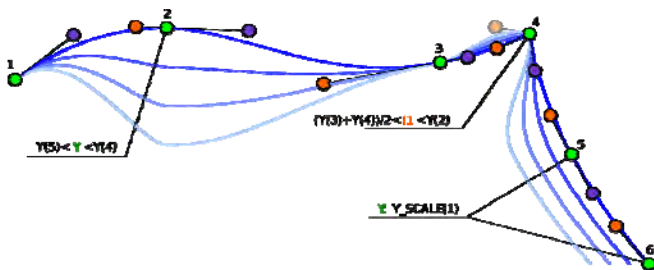


Fig 6. Example of a topological guard affecting a curve.

Colored letters can be green (nodes), orange or purple (control points). The formulas show the way the limitation affects the nodes and control points. Each topological guard represents a certain editable aspect of a facial feature. For example, adding a topological guard “chin width” would affect nodes 4, 5, 6, and 8 of the chin curve, as well as control points of nodes 3 and 9. At value zero, the chin will be the narrowest, and the nodes will be tightly packed together. As the value starts changing, the nodes will disperse horizontally, and the chin will widen. When the value of “chin width” reaches one, the chin will be at its widest. The highest and lowest possible values of a topological guard depends on a multitude of factors, including other topological guards. Chin width would certainly depends on overall head width, and possibly on jaw

width, depending on the rest of the parameters. When a topological guard changes value, values of other topological guards have to be recalculated to make sure they fit the 0 to 1 range when the change is applied.

Each topological guard represents a certain editable aspect of a facial feature. For example, adding a topological guard “chin width” would affect nodes 4, 5, 6, and 8 of the chin curve, as well as control points of nodes 3 and 9. At value zero, the chin will be the narrowest, and the nodes will be tightly packed together. As the value starts changing, the nodes will disperse horizontally, and the chin will widen. When the value of “chin width” reaches one, the chin will be at its widest. The highest and lowest possible values of a topological guard depends on a multitude of factors, including other topological guards. Chin width would certainly depends on overall head width, and possibly on jaw width, depending on the rest of the parameters. When a topological guard changes value, values of other topological guards have to be recalculated to make sure they fit the 0 to 1 range when the change is applied.

Topological guards can affect groups of other topological guards. For example, facial characteristics that are typical of men and women can be roughly quantified into a single parameter. Topological guard such as a chin width, jawline angle, eye size, eyebrows shape etc. can be connected to a single topological guard called “gender”. Individual features of a person can not be accurately modeled using only one parameter. Its primary purpose would be to get the rough shape of a person’s face, and then fine tune the features using more specific topological guards.

The guards used in the model were made using data for treatment of people with prosopagnosia, as that data readily described typical shapes of facial features in a form easily translatable into a program model.

Different geometric primitives can be used to quickly adjust the appearance of an avatar to the environment it could be used in, ranging from a realistic 3d textured environment, to an exaggerated cartoon environment. The look used in this paper is functionally realistic, as it retains realistic proportions, while also using bright cartoon-like outlines and coloring to make the shapes easier to distinguish at a low LOD.

#### IV. AN EXAMPLE OF FACIAL AVATAR PROGRAM MODEL USE

Hand made sketches of faces that faithfully represent suspects require years to train specialized staff, while training regular staff to use software specialized for modeling facial composites takes hours. While some automated facial composite applications do exist, they mostly rely on raster picture processing and achieve usable results about 40% to 60% of the time [1]. Figure 5 shows a mockup of facial composite software that implements the vector program model described in this paper.

As values are adjusted, the changes are calculated and displayed in real time.

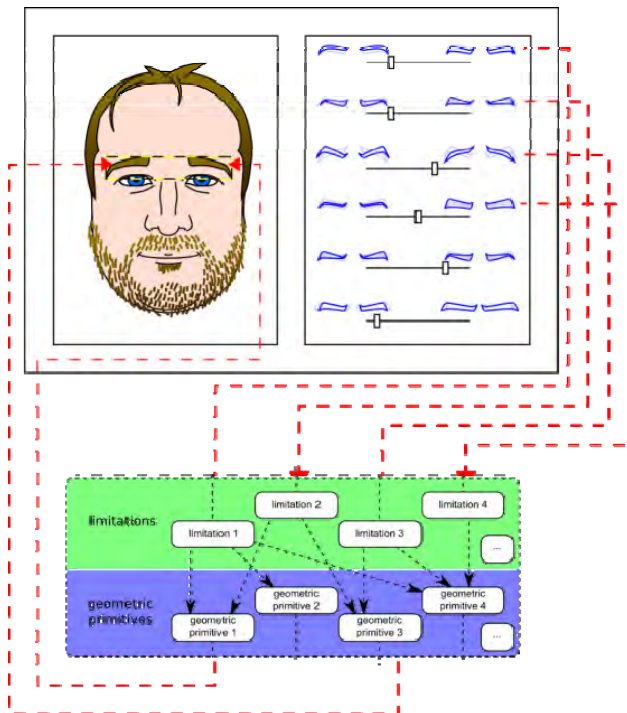


Fig 7. Implementation of the model in facial composite. The mockup interface shows eyebrow topological guards.

## REFERENCES

- [1] [1] Chi-Hyoung Rhee and Chang Ha Lee, "Cartoon-like Avatar Generation Using Facial Component Matching", School of Computer Science and Engineering, Seoul, South Korea
- [2] [2] E. De Renzi, P. Faglioni, H. Spinnler, The Performance of Patients with Unilateral Brain Damage on Face Recognition Tasks, 1st ed. 1968, pp. 17-34.
- [3] [3] Vicki Bruce, Andy Young, Understanding face recognition, 3rd ed., Vol. 77, 1986,
- [4] [4] Dr Sarah Bate, „Information about prosopagnosia“, <http://prosopagnosiaresearch.org/index/information>
- [5] [5] Science Daily team, "Neoteny", <http://www.sciencedaily.com/articles/n/neoteny.htm>
- [6] [6] Norman H. Mackworth, Anthony J. Morand, The gaze selects informative details within pictures, 1st ed, Harvard Publications, 1967, pp. 547-550
- [7] [7] "Make Human Documentation", <http://www.makehuman.org/documentation>
- [8] [8] Vicki Bruce, Andy Young, Understanding face recognition, 3rd ed., Vol. 77, 1986, pp. 305-327.

Using the avatar facial model described here would also make searching a database of existing faces far easier. Faces would be parametrized far more precisely than they are when bitmap analysis is used, but a new database would also have to be created, first.

## V. CONCLUSION

The use of procedural generation to generate graphical representations of facial features provides a practically infinite combination of different shapes, with a significant reduction of memory use, demonstrated on the facial composite application. Lip animation using the same basis as the one in model would demonstrate that it is applicable in areas it was not originally meant for. The model is expandable and can be made to support modeling of non human, but still humanoid models, as well as special cases of unusual facial structures, such as not having certain facial features. The model is promising, and could be developed to serve more purposes, rather than just having animation students practice their coding on it.

# Hadoop - Next Generation of Business Intelligence Systems

Borivoje Milosevic<sup>1</sup>, Srdjan Jovkovic<sup>2</sup> and Marko Jankovic

**Abstract** –The paper deals with the Hadoop software library which provides a framework that enables distributed processing of large data sets using computer clusters and a simple programming model. It enables scaling of applications depending on the available hardware on which it was built (on systems with one server and a large number of servers where each has a local area). The library is designed in such a way that failures can be detected and treated at the application level. So, Hadoop Framework has been developed for ad-hoc parallel processing unstructured data, using MapReduce programming model and the distributed file system HDFS (Hadoop Distributed Filesystem). This paper will explain this new technology.

**Keywords** – Big Data, Hadoop, MapReduce, Hadoop Distributed Filesystem, Java Virtual Machine.

## I. INTRODUCTION

With the development of computer technology, it is now possible to manage huge amounts of data that were previously impossible to process and that could be used only with the help of supercomputers and at great expense. System prices have fallen as a result of new techniques and distributed processing, which are currently in the focus. The real breakthrough in Big data technology occurred when companies such as Yahoo, Google, and Facebook came to the conclusion that they could process large amounts of data which their products generate. These companies have been tasked to find new technologies that will enable them to save, access, manage and analyze vast amounts of data in real time, in such a way that they can process and fairly and properly utilize the amount of data they have and include them in their networks. Their solutions that have emerged have led to changes in the market data management. In particular, the innovations that MapReduce, Hadoop and Big Table brought forth, proved to be a spark that led to a new generation of data management. In support of this technology, the Apache project is developing a number of projects related to open source Hadoop with: Ambari, Avro, HBase, Cassandra, Chukwu, Mahout, Hive, Pig, Spark. Cassandra's Hadoop is distributed "NoSQL" database open source, which is recognized as the basis for the following generation of business intelligence systems, used by Ebay, Twitter and many other companies whose characteristic is that they have large amounts of active data. The largest known Cassandra cluster has over 300 TB of data on more than 400 servers.

<sup>1</sup>Borivoje Milosevic is with the College of Applied Technical Sciences University Nis, A. Medvedeva 20, Nis 18000, Serbia, E-mail: borivojemilosevic@yahoo.com

<sup>2</sup>Srdjan Jovkovic is with the College of Applied Technical Sciences University Nis, A. Medvedeva 20, Nis 18000, Serbia, E-mail: srdjansms11@gmail.com

These technologies emphasize one of the most fundamental problems, which is the ability to process large amounts of data in an efficient and timely manner, in a manner that is cost effective and does not require large expenditures.

The scope of what is now considered as a Big Data is wide, and the definitions are unclear, even contradictory. The most widely accepted definition of the term Big Data is derived from the Meta Group analysis (now Gartner) which was conducted in 2001. According to this definition, the term Big Data stands for information resource large quantities, high speed and high diversity of data that require new and innovative methods of processing and optimization of information, improving access to the contents of the data and decision-making. This is the definition of so-called "3V Dimensions": Volume, Variety, Velocity where the amount of data goes up to the peta byte, and the variety of data to be processed in real time is tight (structured, unstructured, text, blogs, multimedia, etc.), Fig 1. According to some estimates, about 80 percent of data is not of a numeric type, but they still need to be involved in the procedure of analysis and decision-making.

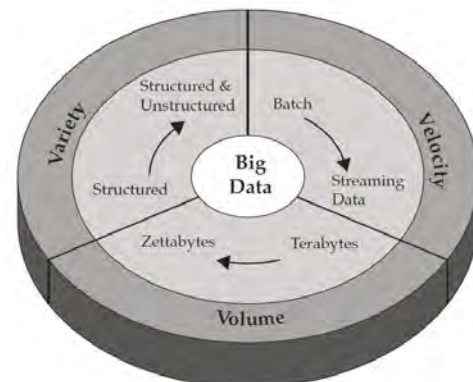


Fig. 1. 3D Big Data outlook

Also, when talking about the characteristics, it is important to note two new dimensions:

**Versatility:** How the data is subject to change? In addition to large quantities and speeds of data processing, data streams can become quite spread with time. This can be explained by a phenomenon in popular media, where one and the same information is repeated many times. Such exceptions are very difficult to handle, especially when you take into account the recent rise in popularity of social networks.

**Complexity:** How difficult is data processing? When dealing with large amounts of data, they typically come from different sources. In many cases it is a fatal pair: filtering and transforming a piece of information in any way. However, it is necessary to connect the relationships between data and



hierarchies of data, otherwise the amount of data can get out of control.

10.000 Credit card transactions are done every second in the world. 340 million tweets are sent daily, which is some 4.000 tweets per second. Facebook has more than 901 million active users who generate data daily, based on their mutual interaction. More than 5 billion people are calling, sending SMS, MMS, tweeting or surfing the Internet on mobile devices.

## II. MAP REDUCE PROGRAMMING

In contrast with traditional relational database-oriented information which organizes data into fairly rigid rows and columns that are stored in tables - MapReduce uses key/value pairs. MapReduce is a programming model for processing large data sets using parallel distributed algorithms in a cluster. MapReduce program includes Map () procedure that performs filtering and sorting (such as sorting students by name in rows, one row for each name) and Reduce () procedure that performs the operation of aggregation (for example, the number of students in each row). MapReduce system manages the distributed servers, and generally the whole process. The system performs different tasks simultaneously, manages all communications as well as the transfer of data between different parts of the system, at the same time ensuring a system against redundancy and errors. MapReduce libraries are written in different programming languages. In-Memory machine provide high-performance memory analytics processing. Free implementation is the popular Apache Hadoop organizations.

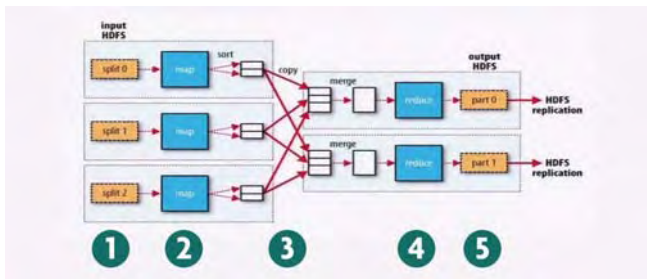


Fig. 2. MapReduce workflow

The MapReduce workflow for such a word count function would follow the steps as shown in the diagram below, Fig 2.:

1. The system takes input from a file system and splits it up across separate Map nodes
2. The Map function or code is run and generates an output for each Map node-in the word count function, every word is listed and grouped by word per node
3. This output represents a set of intermediate key-value pairs that are moved to Reduce nodes as input
4. The Reduce function or code is run and generates an output for each Reduce node-in the word count example, the reduce function sums the number of times a group of words or key occurs
5. The system takes the outputs from each node to aggregate a final view.

The reduction starts when the data is copied from the mapping phase as soon as available. Reduction phase can start only after the mapping phase is completed and the results are collected.

Reduction consists of three main phases:

1. *Shuffle*: Reduction is used for grouping the outputs from mapping phase. At this stage the system, for each node that runs the reduction, finds all relevant parts of the outputs that nodes are produced in the mapping phase. Finding is performed using HTTP.

2. *Sort*: The system groups the inputs to be reduced by using their key. This needs to be done because the different nodes that could do mapping, might produce the keys for the same node for reduction process. The steps of mixing and sorting are carried out simultaneously, that is, when acquiring the output they are joined by together.

3. *Reduce*: At this stage, the method reducer (Object, Iterable, Context) is performed, which calls for each clustered pair <key, value collections>. The output of this phase is usually entered in RecordWriter using TaskInputOutputContext.write (Object, Object).

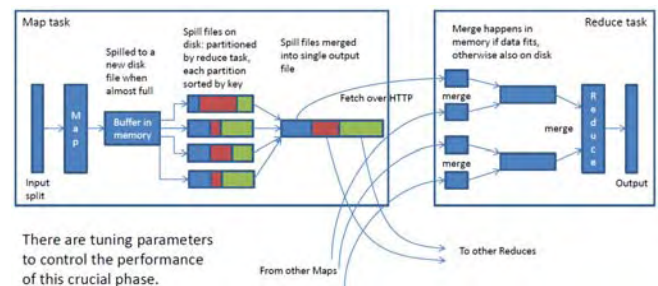


Fig. 3. MapReduce tasks

A MapReduce job is a unit of work that the client wants to be performed: it consists of the input data, the MapReduce program, and configuration information. Hadoop runs the job by dividing it into tasks, of which there are two types: map tasks and reduce tasks, Fig 3.

There are two types of nodes that control the job execution process: a job-tracker and a number of task-trackers. The job-tracker coordinates all the jobs run on the system by scheduling tasks to run on task-trackers. Task-trackers run tasks and send progress reports to the job-tracker, which keeps a record of the overall progress of each job. If a task fails, the job-tracker can reschedule it on a different task-tracker.

Hadoop divides the input to a MapReduce job into fixed-size pieces called input splits. Hadoop creates one map task for each split, which runs the user defined map function for each record in the split. Having many splits means the time taken to process each split is small compared to the time to process the whole input. So if we are processing the splits in parallel, the processing is better load-balanced if the splits are small, since a faster machine will be able to process proportionally more splits over the course of the job than a slower machine.

Example of the MapReduce process can be seen in Fig. 4.



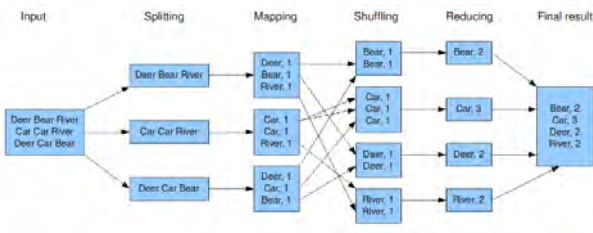


Fig. 4 MapReduce example

### III. HADOOP DISTRIBUTED FILE SYSTEM

The Hadoop distributed file system (HDFS) is a distributed, scalable, and portable file-system written in Java for the Hadoop framework. Even though the Hadoop framework is written in Java, programs for Hadoop need not to be coded in Java but can also be developed in other languages like Python, multi-threaded PHP or C++.

VM Hadoop consists of the Hadoop Common package, which provides filesystem and OS level abstractions, a MapReduce engine (either MapReduce/MR1 or YARN/MR2) and the Hadoop Distributed File System (HDFS). The Hadoop Common package contains the necessary Java ARchive (JAR) files and scripts needed to start Hadoop.

A typical Hadoop environment consists of several specialized software components: MasterNode, NameNode and Worker Node, Fig 5.

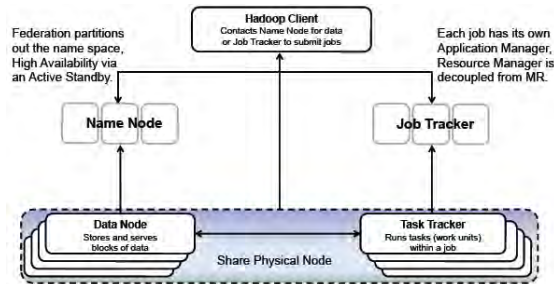


Fig. 5 Hadoop environment for better Availability/Scalability

**Master node:** The majority of Hadoop deployments consist of several master node instances. Having more than one master node helps eliminate the risk of a single point of failure. Here are major elements present in the master node:

- ✓ **JobTracker:** This process is assigned to interact with client applications. It is also responsible for distributing MapReduce tasks to particular nodes within a cluster.

- ✓ **TaskTracker:** This is a process in the cluster that is capable of receiving tasks (including Map, Reduce, and Shuffle) from a JobTracker.

- ✓ **NameNode:** These processes are charged with storing a directory tree of all files in the Hadoop Distributed File System (HDFS). They also keep track of where the file data is kept within the cluster. Client applications contact NameNodes when they need to locate a file, or add, copy, or delete a file.

**DataNodes:** The DataNode stores data in the HDFS, and is responsible for replicating data across clusters. DataNodes

interact with client applications when the NameNode has supplied the DataNode's address.

**Worker nodes:** Unlike the master node, whose numbers you can usually count on one hand, a representative Hadoop deployment consists of dozens or even hundreds of worker nodes, which provide enough processing power to analyze a few hundred terabytes all the way up to one petabyte. Each worker node includes a DataNode as well as a TaskTracker.

A Hadoop cluster has nominally a single name-node plus a cluster of data-nodes, although redundancy options are available for the name-node due to its criticality. Each data-node serves up blocks of data over the network using a block protocol specific to HDFS. The file system uses TCP/IP sockets for communication. Clients use remote procedure call (RPC) to communicate between each other. Fig. 6 shows how Hadoop runs a MapReduce job.

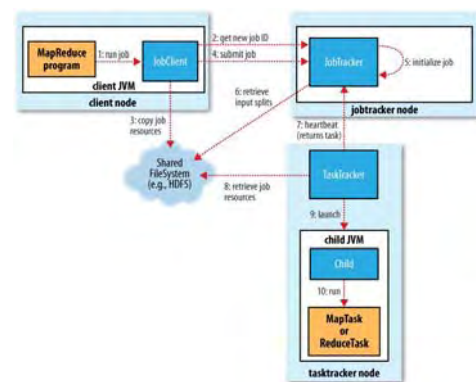


Fig. 6 How Hadoop runs a MapReduce job

The job submission process implemented by JobClient's submitJob() method does the following:

MapReduce job is sent to the execution by using the Hadoop client application JobClient. The client application requires major node (JobTracker), a new unique identifier of the transaction, and calculates the partition of the input data. Once the input data is divided into partitions and all the parameters of the transaction have been checked, copied JobClient business components in distributed file system in a directory - the same transaction identifier is generated in the first step. Components of the transaction include the JAR archive with the program itself, with configuration files and partitions of the input data.

After the components of the job became available in a distributed file system, the job is saved in the internal queue Job queue. Job Scheduler then retrieves the job and initializes the tasks required for its execution. Initialized tasks include the Map function (a Map function for each partition of the input data) and Reduce tool (Reduce the number of functions is defined in the configuration file).

Tasks performing is fully orchestrated with main nodes. Before the execution of certain tasks, JobTracker must choose which tasks belong to the work to be performed. Assumed job scheduler chooses what first arrived in the queue. Once you

choose a job, JobTracker assigned tasks that make up the selected free job creators of jobs. Tasktrackers run a simple loop that periodically sends heartbeat method calls to the jobtracker. Heartbeats tell the jobtracker that a tasktracker is alive, but they also double as a channel for messages. TaskTracker periodically reports its status main node. The state includes information on free slots for Map and Reduce tasks. Important optimization occurs in assignment to the Map tasks. Map tasks are trying to allocate TaskTracker nodes on which there are data processed by the administration assigned task, thus avoiding the costly network communication, since the data are in the local Map task. In order to optimize the overlap of reading and data processing, framework Hadoop runs more Map and Reduce tasks competitively at the nodes workers.

After the TaskTracker node assigned to the task, it retrieves the JAR archive with the program and launches a separate instance of the Virtual Java Machine (JVM) for performing the assigned task. MapReduce programs can take hours, so workers nodes periodically provide information on the progress of the execution. The job is finished when a TaskTracker which executes the last task in the business told the head node to end the execution of the assigned task.

Performing tasks is subject to hardware or software errors that can manifest in various ways. One advantage of Hadoop framework is to monitor the status of jobs and tasks, handling errors and delays in the work and enabling execution of the transaction in a precarious computer online environment. A mistake by the worker node can cause an exception during the execution of Map / Reduce job or some other error in the JVM system. A common case is the phenomenon of slow workers or workers at a standstill. When the main node receives periodic status message alert workers with errors, it restarts the failed task (avoiding the task start on the same node worker where the error occurred). The master node receives periodic general messages about the status of work (tasktrackers pool). In addition, the availability of funds in the system is very high using a distributed file system replication blocks (eg. level of replication funds jobs is 10). In this way, apart from the obvious benefits of reducing network communication during the execution of tasks, system reliability and data redundancy is increased. Hadoop is not currently considered a mistake in the head node that represents a single point of system failure (single point of failure). One possibility is to introduce redundancy protection using the system to manage the ZooKeeper Internet nodes and determine the primary (main) node.

HDFS stores large files (typically in the range of gigabytes to terabytes) across multiple machines. It achieves reliability by replicating the data across multiple hosts. With the default replication value, 3, data is stored on three nodes: two on the same rack, and one on a different rack. Data nodes can talk to each other to rebalance data, to move copies around, and to keep the replication of data high. HDFS is not fully POSIX-compliant, because the requirements for a POSIX file-system differ from the target goals for a Hadoop application.

HDFS has a master/slave architecture. An HDFS cluster consists of a single NameNode, a master server that manages the file system namespace and regulates access to files by

clients. In addition, there are a number of DataNodes, usually one per node in the cluster, which manage storage attached to the nodes that they run on. HDFS exposes a file system namespace and allows user data to be stored in files. Internally, a file is split into one or more blocks and these blocks are stored in a set of DataNodes. The NameNode executes file system namespace operations like opening, closing, and renaming files and directories. It also determines the mapping of blocks to DataNodes. The DataNodes are responsible for serving read and write requests from the file system's clients. The DataNodes also perform block creation, deletion, and replication upon instruction from the NameNode.

Hadoop does its best to run the map task on a node where the input data resides in HDFS. This is called the data locality optimization. It should now be clear why the optimal split size is the same as the block size: it is the largest size of input that can be guaranteed to be stored on a single node. If the split spanned two blocks, it would be unlikely that any HDFS node stored both blocks, so some of the split would have to be transferred across the network to the node running the map task, which is clearly less efficient than running the whole map task using local data.

#### IV. CONCLUSION

Unlike Relational database management system - RDBMS, Hadoop MapReduce model exhibits a proportional increase because the Map and Reduce functions do not depend on the size of the input data set, nor the size of the Internet PC network to which the system is running. MapReduce model processes the entire set of data during the execution of queries, while RDBMS systems typically maintains additional data structures ( B tree ), that speeds up executing queries or updating small amounts of records, but significantly slows down the update of most records in the database. In addition, the MapReduce programming model is designed to handle unstructured (or semi-structured) data such as text, multimedia or binary data.

#### REFERENCES

- [1] CS246: Mining Massive Datasets, Hadoop Tutorial, January 12, 2016, New York.
- [2] A. Hammad, A. García, Hadoop tutorial, Karlsruhe Institut of Technology, SSC, | September, 2014, Germany.
- [3] Dhruba Borthakur, HDFS Architecture Guide, Copyright © 2008 The Apache Software Foundation.
- [4] Amr Awadallah, Introducing Apache Hadoop: The Modern Data Operating System, Stanford EE380 Computer Systems, 2011, California, USA.
- [5] Diana MacLean for CS448G, 2011, A Very Brief Introduction to MapReduce, <http://labs.google.com/papers/mapreduce.html>
- [6] MapReduce by examples, <https://github.com/andreaiacono/MapReduce>
- [7] Ivan Validžić, Primena paralelne obrade u analizi društvenih mreža, Fakultet elektronike i računarstva, Zagreb, Croatia, 2015.
- [8] Borivoje Milošević, Danica Milošević, In Memory baze podataka, Zbornik Radova, Visoka Tehnička Škola Strukovnih Studija - NIŠ, 2015.

# Formal Methods for Data Analysis

Blagovest Kirilov<sup>1</sup>

**Abstract** – In this paper formal methods for data analytics will be discussed upon. The main goal is to calculate an analytical system that can formalise methods for data analysis.

**Keywords** – *Data Analytics, Formal methods*

## I. INTRODUCTION

Data analytics is getting more and more important for business, science as well as ordinary people. Because of the abundant sources of information that flood the world with it through various channels this leads to an ever-growing demand for analyzing this data and filtering it for effect. Still, although, there exist numerous methods that can execute this task, there is no formal method of defining and using them. This means that the approach to each analysis is different and thus can also lead to a result of wrong data or increased time and resource costs in order to receive the results.

## II. DATA ANALYTICS METHODS

Information is the core of all success stories as well as failures in the current age of technology. In the ever-changing daily drive there is a requirement for secure information sources, as well as fast analysis of provided information. In the last ten years a new term has been established, namely Big Data. This is the description of techniques and technologies that require new integration methods, in order to find hidden values and data in complex and massive sets of data. They not only analyze and find those, but also sort and order them for best effect.

This information management systems were and are still seen by many businesses as a hindrance rather than a supporting function that gives them new opportunities. Surprising is the fact that even today some businesses refuse to switch to such systems, in the process wasting more than half of their IT budgets just to keep the old systems alive and running, albeit at high cost and high impact. The rising flow of data that flow between interconnected companies, the internet, as well as the public domain, coupled with official state prescriptions, led to companies using such systems for cost reduction on data analytics. In order for this to happen, existing systems and methods are “fixed” to work with the new systems and methods, but without any underlying architecture, process or outcome expectation set-up.

In order to understand the key of this publication on a better scale, basic methods for data analytics and their development

in the last ten years is reviewed, as well as a comparison of their key features.

### A. Big Data

Even ten years ago big data was in the background of all future-talk on data analytics. The then used method of choice was Business Intelligence, where Business Intelligence was finding trends through descriptive analysis of big data sets with high data density. On the other hand, Big Data is using inductive statistical dependencies for data with lower data density. In addition, Big Data creates forecasts for output data as well as behavior models. Base descriptors of Big Data are the Mass (big datasets), Diversity (data in various formats), Velocity (dataflow) and accuracy (confidence in correct data).

“Actual forecasts say that the global data volume will double itself each year.” [1]

This trend, coupled with advances in hardware technology, shows clearly, that data must be handled faster, because their volume is rising constantly. Software and the methods for this are getting more and more crucial in order to achieve a constant flow and analysis result as well as to be scalable. The final goal is to have a way to analyze and present Data in real-time.

### B. Base methods for data analysis

One of the widely used methods is the so-called linkage analysis – this method analyses data for opportunities on linking it together, combining it in the so-called life-cycles, which are returned as an output of categorized data.

Another method is Data Mining. Data Mining searches for re-occurring events in data-sets of structured data. Those can also be combined in new patterns.

The last observed method is the predictive analysis. For this method the main difference in comparison to the previous ones is that the data is not structured and coming in big volumes. This input happens mainly in real-time, and the method looks for patterns in the different data-sets and their data. The end-result is prepared data for further analysis which is combined in separate entities based on the preliminary analysis.

All methods serve their purpose in providing data analysis to different data sets with different results, although there is no clear distinction on when each one of them should be used in reality. As they treat data differently, this also leads to

<sup>1</sup>Blagovest Kirilov is with the Faculty for German Engineer and Business Management Education at the Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria.

different outputs and time effort required for each one of them.

### C. Comparison of the methods

We can split the methods in three types – descriptive, predictive and prescriptive.

The descriptive methods, such as the linkage analysis, rely on already existing data in order to use milestones for certain events, thus splitting the data in smaller units, which are linked logically to each other, leading to clear outputs. Those methods rely on input data, which are available at an earlier stage, in order to give an output. This is getting harder with the complex flow of input data rising in real-time, because there can be new data types that are created during this real-time flow, who make previous analyses irrelevant. With this, the time required for the execution of those analyzes is directly linked to the amount of incoming data. In addition, if the dataflow is really high, the analysis needs to be repeated and re-arranged every time, which also has a direct effect on the effort. Because of the constant need for state-of-art data-sets the accuracy of the data is also not guaranteed, which has a negative effect on the end-result. Descriptive methods also have a relatively low impact on the optimization of data, because a data reduction is only possible if the only data that is used is historical one, which is not refreshed in real-time.

Predictive methods, for example Predictive Analysis and Data Mining, analyze historical as well as new data-sets, in order to find patterns, that can be used as predictions for future events. Those methods can not say what will happen in the future, they only point at what could happen. This is done by deriving data, which is not gathered, from the one already available. For those methods the time for execution is similar to the one of the descriptive methods, because predictive analyses that are done for future data are not required to be executed again if the future data matches the prediction. Predictive methods are generally used for data reduction purposes, because from them the data can be derived that will flow in in the future, allowing for models for their analysis to be provided, thus having a faster analysis time.

The last type of methods, in this case specific examples of Predictive Analysis and Data Mining, named prescriptive methods, are seen just now as being fully functional. Their goal is to not only provided an analysis on given input data, but also to go a step beyond by presenting possible actions and their effects. The prescriptive methods are a further development of the predictive methods, by saying what exactly needs to be done in order to reach a set goal. They also don't give just one way of achieving this, but multiple branches with different actions, depending on the future data coming in, thus providing actions for every situation that can arise. In order to execute a prescriptive method, a predictive method is required as well as two additional parameters – data, for which actions need to be put in place to influence the

results, as well as a system which is documenting the influence of taken actions on the data.

For those systems the data mass is determined not only by the input, but also from the different branches, actions and resulting events, because all of them need to be analyzed in parallel. Thus, it is very important to set up a high data reduction, which can be controlled by the incoming feedback, in order to prevent non-critical and unneeded data set amounts. Furthermore, the accuracy of data is extremely important for such methods, which also leads to a longer execution time compared to the other methods. Still, this is a pointer in the direction of intelligent systems for data analysis giving proven and accurate data as outputs, which also increases their usability. The dataflow doesn't influence the method as far as the rest, because here the future flow can be predicted according to the actions, which is not always the case for say predictive methods.

The following table shows the summary of the most important parameters in relation to the three types of methods for data analysis described prior. It also shows their influence on accuracy and data reduction, their execution time as well as the influence of dataflow, data variance and data quantity on the methods. This calculations also apply for all types of methods mentioned within the prior chapter.

TABLE I  
BASE DATA ANALYSIS METHOD COMPARISON

	Execution time	Influence on	
		Accuracy	Data-reduction
<b>Descriptive Methods</b>	Very high	Medium	Low
<b>Predictive Methods</b>	High	High	High
<b>Prescriptive Methods</b>	Very high	Very High	Very High
	Influenced by		
	Dataflow	Variance	Data quantity
<b>Descriptive Methods</b>	Low	High	Very High
<b>Predictive Methods</b>	Very High	High	Very High
<b>Prescriptive Methods</b>	High	High	High



As can be seen from the table, there is no clear “winner” from the methods, but each one of them has it’s negative and positive sides. The user needs to decide which one to use. The challenge arises when the user is not aware how to proceed and thus can not rely on a formal method for data analysis in order to cover his needs. This can lead either to increased cost because of additional method set-up or repeated method execution, as well as to wrong data accuracy if the correct method to be used is not chosen. In addition, routine maintenance and future-proofing of such methods bound into client systems is not guaranteed, leading to additional costs and missed opportunities for using the data as a competitive advantage. This, coupled with the complex business environment and customer expectations, can be the decisive factor between staying in business or losing everything. Thus, the need for a formal method for data analysis becomes more and more apparent. It should not only reflect different data inputs, but also differentiate between them and update the output as necessary in order to reflect the best accuracy and dataflow process continuity.

### III. FORMAL METHOD DESCRIPTION

#### A. Assumptions

In order for the methods to be formalized, they need to be having a couple of base information descriptors provided. The first one is data flow, followed by data quantity as well as data mass. Those are basic definitions of the data inputs, and most methods existing today are also using those descriptors in order to be differentiated between each other[1].

In addition, due to the nature of the methods, they affect the data accuracy, data reduction and execution time of the given input set of data.

#### B. Method approach

Different methods that can be included in the formalisation are descriptive, predictive and prescriptive methods, as they all can be used upon the same base input data. The important part is the interaction between the data, methods and the resulting output, in order to have formalised rules for the whole process. Thus, an interface will be created based on an use-case that can define the formalisation process, accepting data from different entities in different types and forms, and applying the same basic calculations for the different types of data analysis methods.

The output should be based on this process and always give a result to the users on the receiving side, thus completing the formalisation steps set-up.

The approach taken for the method is described in the next point, where a use-case definition is set-up as a basis for the further rollout and description of the method.

#### C. Use-case definition

The following use-case is set forward for the method:

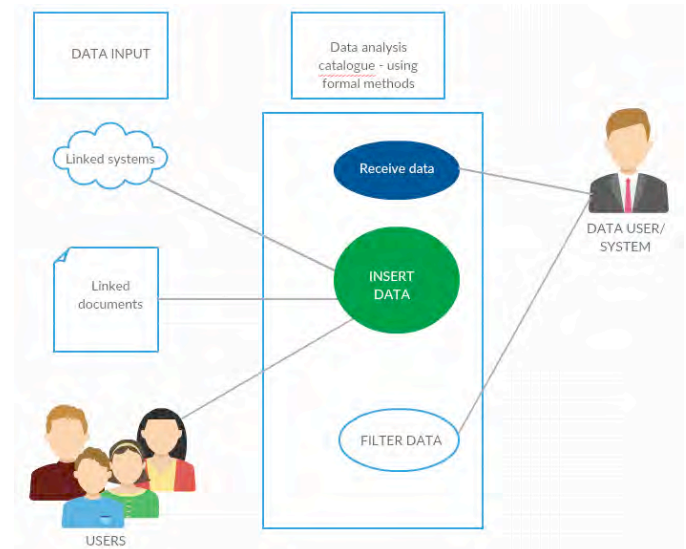


Fig. 1 Use-case for method description

In the above use-case scenario (Fig. 1), all types of data can be collected, regardless of their origin and type. Thus, this ensures the use-case is valid for all scenarios of data analysis.

The data analysis catalogue will pursue the following steps:

- Collect data from the input sources
- Analyze data for base descriptors[2]
- Apply data analysis to collected and pre-analyzed data
- Create output data and publish to end-users

Each of those steps is highlighted in the points below.

#### D. Data collection

Data collection should work regardless of input method. This is why the interface for the formalization will have input ports accepting all protocols and inputs. Error rates on data recognition should be non-existent, going through several checks to ensure data integrity. Data can be real-time or historical, structured or non-structured, or any combination of the ones above.

#### E. Data analysis for base descriptors

After collecting the input data, the interface will analyze the data for base characteristics and descriptors to match further algorithms. Thus, data can be sorted and prepared prior to the next step. Here, useless data pieces will be discarded, and the remaining data clustered into different category layers, thus producing a base process for identification, data quantity description, as well as highlighting criteria such as accuracy that can be obtained, mass of data and others.[3]

#### F. Data analysis application

Following the previous step, the data analysis application will have the following process (including previous steps)[4]:

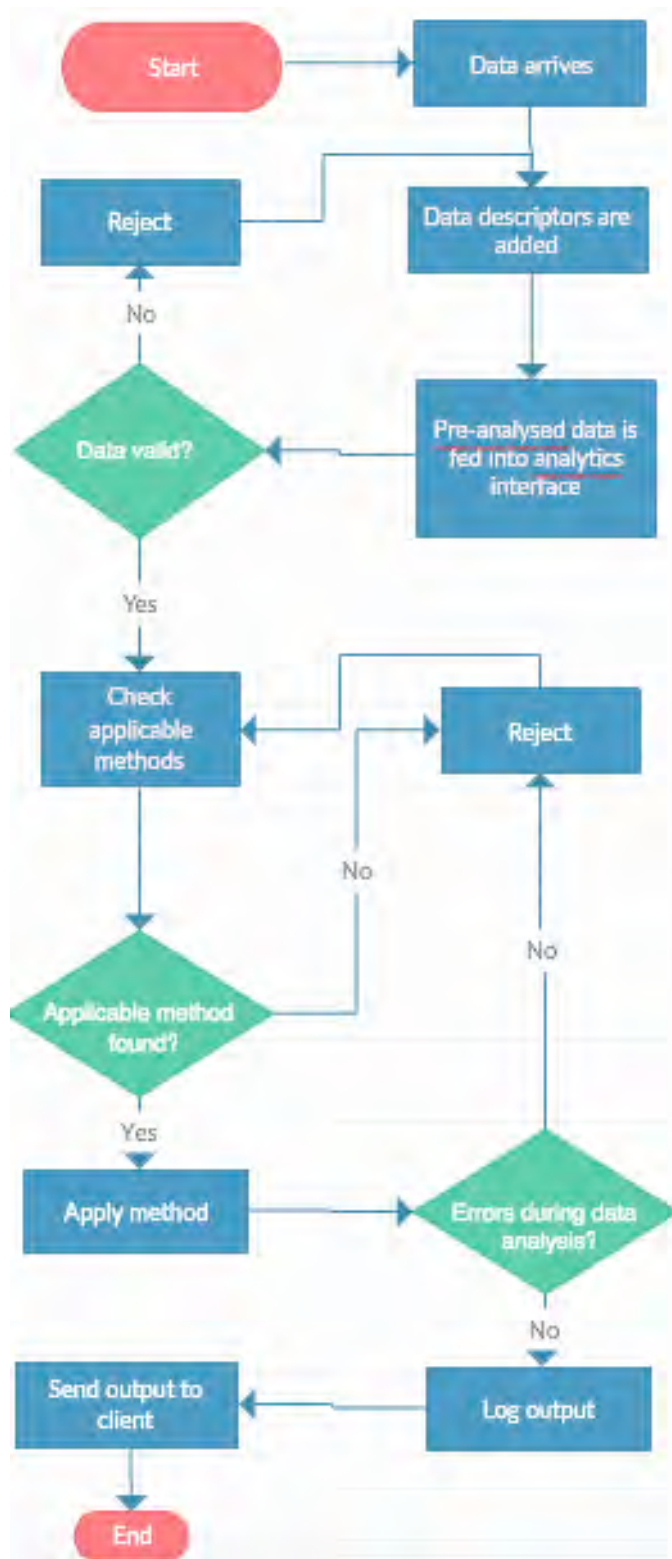


Fig. 2 Process flow for method description

In order for the process to have additional checkpoints, underlying documents for all checks should be created through testing and implementation. Each of those steps should be amended to the process flow as notes and additional documentation steps. In addition, the processing time, requirements as well as error rates should be measurable in order to provide inputs for accuracy and a proof of concept. The interface should be also created in a software implementation that can differentiate and show the process flow in action with base input data.

#### G. Create output data and publish to end-customers

After the analysis is done and the formal method applied, the resulting data with the analysis should be packaged in different formats, so as to be input in different kinds of databases, linked to existing data for matching and further analysis or saved as a documentation for end-users and data operators. This ensures the process flow described in Fig. 2 will have an end-to-end approach, going from the input of the base data through the analysis and ending with the output of the results as per requirements set-up. This is an additional step after the formal method for data analysis is applied, but ensures practical usability as well as a well designed interface that is created during the previous steps.

## IV. CONCLUSION

In this paper an introduction on methods used to formalize data analysis is laid out on the basis of use-cases[5]. A draft process flow model is outline that gives the base process steps for the formalization. Overlying on this, an interface should be created that will enable the rollout of the formalization for all types of data input, as well as all methods to be used. This model will feature end-to-end functionality, going from the pre-analysis and integrated descriptor checks of input data through applying data analysis methods through formalization and ending in user output. Given the development and importance of the topic for all industries and public areas of importance, this would give a solid foundation for a set-up of formalization to various methods, so as to have scalability, cost-effectiveness and future-proofing going ahead[6].

## REFERENCES

- [1] IBM - Analytics: Big Data in production
- [2] International Data Corporation (IDC) (2013), From: February 2013
- [3] BUBECK, Dr., Uwe; KLEINE BÜNING, Prof. Dr., Hans (2014): Maschinelles Lernen.
- [4] Webster, John. "MapReduce: Simplified Data Processing on Large Clusters", "Search Storage", 2004
- [5] Hu, Han; Wen, Yonggang; Chua, Tat-Seng; Li, Xuelong (2014). "Towards scalable systems for big data analytics: a technology tutorial". IEEE Access 2: 652–687
- [6] From Data Mining to Big Data and Beyond; By Gregory Piatetsky-Shapiro April 18, 2012;

# An Approach in Spatial Databases

Mariana Stoeva<sup>1</sup> and Violeta Bozhikova<sup>2</sup>

**Abstract** –This paper presents our work in the area of image retrieval in Image Databases for images saved by spatial similarity of extended objects location. We propose a new approach to description of the spatial location of extended objects. The approach is based on a geometric approximation. By the development of the proposed approach we enrich our former efforts for creation of effective method for image storage and retrieval from Spatial Databases.

**Keywords** –Image databases, spatial image databases, symbol description of extended objects, spatial similarity retrieval and spatial query processing.

## I. INTRODUCTION

The aim of this paper is to achieve a new response to the raised requirements of different applications to Spatial Image Databases (SIDB), namely attainment of invariance of query processing results with respect to arbitrary compositions of transformations. Our attention is directed to the object description methods by the spatial location of the extended objects contained in it. The image description in SIDB is in conformity with the next image retrieval from it by spatial query.

In real-world database applications, the rotation invariance is a basic issue because each image is captured and stored in agreement with a viewpoint which is implicitly dependent on an outside viewer who establishes a fixed scanning direction. Any translation of this viewpoint and/or any rotation of the image affect the direction relations between each pair of objects. In the recent literature, several approaches can be found whose aim is to provide a solution to the rotation invariance of the conventional indexing methodologies based on symbolic projections [2], [3], [4]. In [7] and [8] are presented approaches for speeding-up the time responses in databases which uses Spatial Access Methods (SAMs) to treat image content in conjunction with two well-known image matching methods, namely the editing distance on Attributed Relational Graphs (ARGs) and the Hungarian method for graph matching. It provides index support for the two most common types of similarity queries, referred to as range and nearest neighbour queries and has many desirable properties. However, in real application, it would also be able to find the images in the database that present a given pattern, even if it appears mirror reflected. An iconic indexing methodology is present in [6] which guarantees the rotation and reflection invariance of the image visual content, where it is described

by direction relationships, which are view-based. This methodology does not recognize the similarity between two images when their corresponding symbolic descriptions (R-virtual images) have been extracted with respect to different rotation centres. In [9] we have proposed a geometric based structure which makes possible the extraction of spatial relations among domain objects that are invariant with respect to transformation such as translation, rotation, scaling. In [10] we propose an approach to symbol description of extended objects in SIDB and now we present its further development. By the proposed approach to spatial information description in an image we aim to achieve an effective storage of the image description information, effective query to SIDB processing, and efficient invariant with respect to transformations image retrieval from SIDB by spatial similarity.

By the development of the proposed approach for image spatial features description we enrich our former efforts to create an effective method for image storage and retrieval from SIDB through “query by image example” that gives an account of the spatial location of image domain objects and is invariant with respect to transformation compositions.

The solution of the task for image storage and retrieval from SIDB by spatial similarity lies on the following basic sub-solutions: consistency of original approximations of image objects area; determination of the spatial relations among domain objects, that are invariant with respect to transformations in images; determination of the similarity between spatial relations; determination of the measure of spatial similarity between relations of two images, spatial similarity algorithm whose results are determining for the spatial query processing.

## II. APPROXIMATION OF THE EXTENDED IMAGE OBJECT AREA

The requirement for storage and spatial investigation efficiency imposes the condition for minimal dimension of the shape description of extended image object area. The objects are two-dimensional areas and the search for their minimal description is as a description of their approximation with suitable for a following processing rectangle form. The requirements for invariance with respect to transformation impose the perceived approximation to be close to the object area, as well as to be simply determined with respect to transformations. This means that an object is approximated with rectangle, whose sides concur with the same points of its approximations of the image objects area.

**Definition 1.** From all rectangles that include the area of an image object, MAR is the one whose area is minimal. If there are more than one such rectangles with minimal area, for approximating MAR is accepted this one, whose sides form smallest angle with the axis orientation of the object domain.

<sup>1</sup>Mariana Stoeva is with the Department of Computer Science and Technologies, Varna Technical University, Bulgaria, E-mail: mariana\_stoeva@abv.bg

<sup>2</sup>Violeta Bozhikova is with the Department of Computer Science and Technologies, Varna Technical University, Bulgaria, E-mail: vbjickova2000@yahoo.com



This definition insures that for the approximating MAR determination are taken the coordinates of the same points of the area external contour independently from the possible transformations. The points from the contour of such determined approximating MAR of an object from an image keep their relative location towards its area centroid. MAR avoids the popular in the literature “diagonal imprecation” in two-dimensional area approximation by Minimal Boundary Rectangle (MBR). MAR is calculated by using object rotation around it’s centroid. The spatial data for each object include the absolute Decart coordinates of these 4 characterizing the object points (angle points of MAR) and of the object centroid, obtained from the points of its external contour.

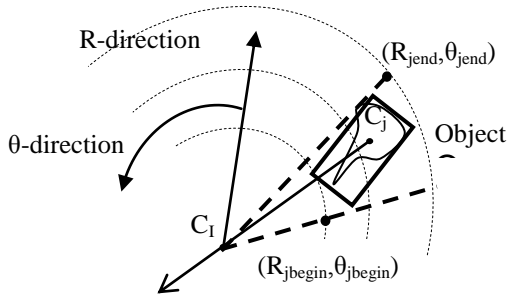


Fig. 1. Illustration of the approximations used

This information that is saved for each object, allows us to use ring sectors (MRS) determined by concentric circles for approximating the objects’ shape extent. Figure 1 illustrates the approximations used. By using this representation we determine, analogically to the orthogonal models, the well-known 13 relations in one linear (R) direction and one specifically used circle ( $\theta$ ) direction. We provide the transformation invariance of determination the atomic relations between two objects by utilizing the properties of the object and image centroids.

### III. IMAGE CONTENT DESCRIPTION

By image symbol description we avoid the need of their repeated understanding. The memory that is necessary for the storage of the performed symbol image is insignificant in comparison with its physical performance. The query processing for image retrieval from SIDB uses only stored information for the image. The stored in SIDB information for the domain objects of each image contains as attributes symbol object names and 5 pairs of Decart coordinates that describe the spatial data of the object. The first one is the pair of object centroid coordinates, obtained from the coordinates of its area external contour. The next 4 are the coordinate pairs of the angle points of the approximating the object area MAR. The storage of absolute location of each object’s centroid is indispensable due to the fact that the relative spatial location of the stored information for each object’s area is preserved towards it. The image centroid, towards which the image objects preserve their relative spatial location, can be obtained from the stored centroids of an objects set from an image.

We accept that the domain objects describing symbol information is stored and arranged in alphabetic order of objects names. We accept that domain objects are named consistently through all images in DB and the images contain numerous examples of object type. The chosen SIDB model is object-relational and results in a general structure with general type of relational diagrams of two relations: “Images” and “Image objects”.

### IV. SPATIAL SIMILARITY IMAGE RETRIEVAL IN IMAGE DATABASE

This information that is saved for each object, allows us to use MRS determined by concentric circles for approximating the objects’ shape extent [8]. MRS of an object is obtained by the cross of concentric circles with center – the image centroid point, with passing through it strait lines. MRS minimally encloses the object area describing MAR and is determined by the minimums and maximums of its polar coordinates.

The spatial relations between each pair domain objects are determined in sense of conditions over the initial and final points of their MRS in two directions of a polar coordinate system [8]. By using this representation we determine, analogically to the orthogonal models [1], [4], the well-known 13 relations in one linear direction and one specifically used circle direction.

**Definition 2.** A triple like  $(O_j \gamma O_i)$  that is called an atomic spatial relation, where  $O_j, O_i \in O_1$  are object names and  $\gamma \in \{<, >, |, :, =, [, (, ), /, \backslash, \%, \#\}$  is an operator. We use the notation  $(O_j \gamma_R O_i)$  to indicate that the pair of objects  $(O_j, O_i)$  belongs to relation  $\gamma$  in R-direction and  $(O_j \gamma_\theta O_i)$  to indicate that the pair of objects  $(O_j, O_i)$  belongs to relation  $\gamma$  in  $\theta$ -direction.

The defining of binary spatial relations is identical to those in [1] adapted to a polar system and it is presented in [8]. Applied for both directions here arise generally 169 spatial relations between two objects in two dimensions, by which the spatial content can be suitably presented. We provide the transformation invariance of determination the atomic relations between two objects by utilizing the properties of the object and image centroids.

The defining of binary spatial relations is identical to those, defined in the orthogonal models [8], adapted to a polar system, and it is presented in Table 1. The 7 symbols for spatial operators, used by Lee and Hsu in [5], are used here also, and 6 new symbol marks for operators indication  $(>, :, |, (, ), \#)$  are inserted.

Table 1: Determination of R- $\theta$  symbol spatial operators

Spatial relations		Conditions
Before	$A < B$	$\text{end}(A) < \text{begin}(B)$
After	$A > B$	$\text{begin}(A) > \text{end}(B)$
Touch	$A   B$	$\text{end}(A) = \text{begin}(B)$
Touch inversely	$A : B$	$\text{Begin}(A) = \text{end}(B)$
Overlap	$A / B$	$\text{begin}(A) < \text{begin}(B) < \text{end}(A) < \text{end}(B)$
Overlap inversely	$A \backslash B$	$\text{begin}(B) < \text{begin}(A) < \text{end}(B) < \text{end}(A)$



Equally $A=B$	$\text{begin}(A) = \text{begin}(B);$ $\text{end}(A) = \text{end}(B)$
Common start $A[B$	$\text{begin}(A) = \text{begin}(B);$ $\text{end}(A) > \text{end}(B)$
Common inverse start $A(B$	$\text{begin}(A) = \text{begin}(B);$ $\text{end}(A) < \text{end}(B)$
Common end $A]B$	$\text{begin}(A) < \text{begin}(B);$ $\text{end}(A) = \text{end}(B)$
Common inverse end $A]B$	$\text{begin}(A) > \text{begin}(B);$ $\text{end}(A) = \text{end}(B)$
Include $A\%B$	$\text{begin}(A) < \text{begin}(B);$ $\text{end}(A) > \text{end}(B)$
Include inversely $A\#B$	$\text{begin}(A) > \text{begin}(B);$ $\text{end}(A) < \text{end}(B)$

Similarity distance is defined on the number of objects from the query, the common for both images objects, as well as on the similarity of their corresponding atomic relations and equations. Its values are in the diapason [0,1], where 1 is valid for image that completely satisfies the query and the value is 0 if they have no common objects.

According to our understanding for similarity and our desire the method to be independent of subjective interpretation, we put forward a formula for similarity evaluation that assesses the similarity between the common for both images objects and their corresponding atomic relations.

**Definition 3.** Let the query image is  $Q$  and the Image Database image is  $I$ . We define the similarity distance  $\text{sim}(Q,I)$  between  $Q$  and  $I$  by Eq.1 and  $\text{sim}(Q,I) \in [0,1]$ ,

$$\text{sim}(Q,I) = \frac{1}{m} \left[ C_o n + C_s \frac{1}{2n} \sum_{j=1}^n (\text{sim}_{ji}(\gamma_R, \gamma_R') + \max \left\{ \begin{array}{l} \text{sim}_{ji}(\gamma_\theta, \gamma_\theta') \\ \text{sim}_{ji}(\gamma_\theta^\circ, \gamma_\theta^\circ') \end{array} \right\}) \right] \text{ for } \forall n \neq 0$$

$$\text{sim}(Q,I) = 0 \text{ for } \forall n = 0 \quad (1)$$

where where:  $m = \|O_Q\|$  is the number of objects in the query image,  $n = \|O_Q \cap O_I\|$  is the number of the common for both images objects,  $\text{sim}_{ji}(\gamma_R, \gamma_R')$  is the spatial similarity between the images  $Q$  and  $I$  for the object match  $(O_j, O_i)$  in  $R$ -direction, and  $\text{sim}_{ji}(\gamma_\theta, \gamma_\theta')$  is the spatial similarity between the images  $Q$  and  $I$  for the object match  $(O_j, O_i)$  in  $\theta$ -direction.  $\text{sim}_{ji}(\gamma_\theta^\circ, \gamma_\theta^\circ')$  is similarity between the spatial relations of image  $I$  and reflection image of the query  $Q$  for the object match  $(O_j, O_i)$  in  $\theta$ -direction; and  $C_o, C_s$  are coefficients ( $C_o + C_s = 1$ ).

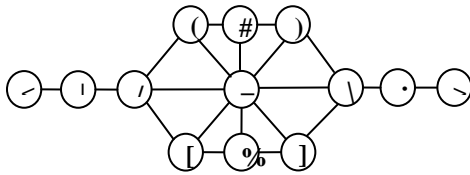


Fig. 2. Interval neighbor graph of the projection relations

This definition is different from the corresponding similarity distance from [8] by the including of the similarity between the spatial relations of image  $I$  and reflection image of the query  $Q$  in  $\theta$ -direction ( $\text{sim}_{ji}(\gamma_\theta^\circ, \gamma_\theta^\circ')$ ) and the coefficients ( $C_o, C_s$ ). The processing of the image query to SIDB includes a filter that detaches as candidates only those images from SIDB that include objects from the query. The

images – candidates are evaluated by spatial similarity for their proximity to the query by similarity algorithm  $\text{SIM}_{R\theta}$ .

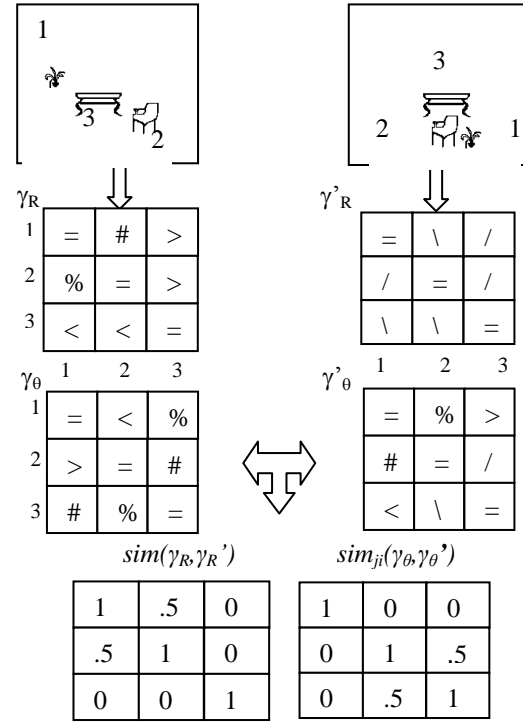


Fig. 3. An example, illustrating the obtaining of the spatial relations similarity, got between the three objects of two images

The similarity value  $\text{sim}_{ji}(\gamma_1, \gamma_2)$  is defined for each pair  $(\gamma_1, \gamma_2)$  of spatial relations with equation:

$$\text{sim}(\gamma_1, \gamma_2) = 1 - d(\gamma_1, \gamma_2) / 4 \quad (2)$$

where the distance between two projection relations  $d(\gamma_1, \gamma_2)$  is determined as the length of the shortest way between their corresponding knots in the graph (Fig. 2.). The maximum value of the so determined distance in the graph is 6, and the minimum is 0. We accept that some degree of similarity exists if the distance is smaller than or equal to the average possible value of the distance.

The relations similarity in the two images  $Q$  and  $I$  for the pair of objects  $(O_j, O_i)$  in  $R$ -direction is indicated as  $\text{sim}_{ji}(\gamma_R, \gamma_R')$ , and as  $\text{sim}_{ji}(\gamma_\theta, \gamma_\theta')$  in  $\theta$ -direction.

An example, illustrating the obtaining of the spatial relations similarity, got between the three objects of two images, is presented by Fig. 3.

## V. EXPERIMENTS

The experiments that improve invariance to transformation compositions are published in [10]. We present hear the following experiment that investigates how the similarity distance detects smaller and bigger spatial differences between the images and how it reads the target understanding for spatial similarity. For these aims a test group 1 is used, which is presented in Fig. 4. The images in it are synthesized, and in one original image 6.1. smaller and bigger object and spatial differences are

inserted as disposed, missing and alien to the query objects. The images include part of generally 24 named regional objects, whose iconic sketches are taken from the TESSA collection, or are created for variety of the test data. The chosen 6 extended objects in the original image 4.1 allow the generations of not big number of image variants differ by with gradually increasing intentionally inserted differences, for which similarity distances in possibly larger range shall be obtained. Images 1.2, 1.3, and 1.4 differ from the query by the disposed spatial position of one object. This object changes its relative spatial location in the consecutive variant. Images 1.5, 1.6, 1.7 and 1.8 differ by disposed two objects, and images 1.9, 1.10, 1.11, and 1.12 differ by the spatial position of three disposed objects. Images 1.4, 1.8, and 1.12 each have as additional difference one missing and one alien to the query object. Images 1.7 and 1.11 are transformation variants with composition (scaling, translation, rotation) respectively to 1.6 and 1.9.

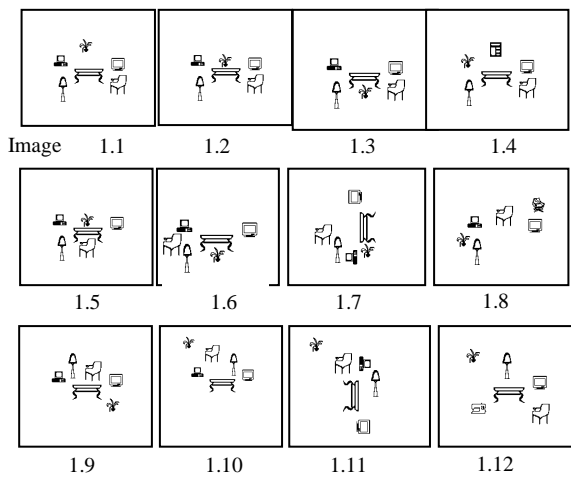


Fig. 4. Test images of Group 1, when image 1.1 is query by example

For the image query 4.1 and parameters  $C_o=C_s=0.5$  the similarity ordering of the images in the group is as follows: {4.1, 4.2, 4.3, 4.5, 4.6, 4.7, 4.4, 4.9, 4.10, 4.8, 4.11, 4.12}. The obtained ordering corresponds completely to the expected one on the base of the introduced spatial differences in the group images. The values of the similarity measure change in the range [1; 0.6750] and decrease by the increase of the spatial differences (in disposed, missing and additional objects) between the images and the query. If the images are grouped by the number of the object differences, close, though different, values of the similarity distance are obtained for each object sub-group. The results, obtained for the similarity distance for the images of the sub-groups show strong dependence of the retrieval value on the number of differences, inserted in the image variant.

## VI. CONCLUSION

The image spatial properties describing approach, proposed in this paper, implemented in a method for storage and retrieval of images from SIDB by spatial

similarity of their domain objects achieves stability with respect to arbitrary compositions of transformations, shows completeness, correctness and sensitivity of the results. The method achieves very good effectiveness of information storage in SIDB and good efficiency of query processing. The experiments investigate it in details by use of various evaluations of the results and their comparison with other methods results. The experiments result demonstrates that the proposed approach is invariant with respect to transformations including reflection and its evaluation has stable behavior when enriching and detailing spatial relations among objects.

The main contributions are:

Utilization of new approximations, that provides short symbol description for storage in SIDB. This short information allows achieving invariance from transformations when determining the spatial relations in images.

Spatial similarity approach and distance measure for image retrieval from SIDB that recognize the shape, measures and mutual location of their objects. They detect transformed images and sub-images.

## REFERENCES

- [1] J. F., Allen "Maintaining Knowledge about temporal Intervals", Comm. ACM, vol. 26 pp.832-843, 1983
- [2] F. Gudivada, V. Raghavan, "Design and Evaluation of Algorithms for Image Retrieval by Spatial Similarity", ACM Trans. Information Systems, vol. 13(1) pp. 115-144, 1995
- [3] F. Gudivada - □□-strib: A geometry-based representation for efficient and Effective Retrieval of Images by Spatial Similarity, IEEE Trans. Knowledge and data engineering, vol. 10, no.3, pp.504-511, 1998
- [4] E. Jungert, "Rotation Invariance in Symbolic Slope Projection as a Means to Support Spatial Reasoning", Proc. Third Int'l Conf. Automation, Robotics and Computer Vision "ICARCV'95", pp. 364-368, 1994
- [5] Y. Lee, F. J. Hsu, Spatial Reasoning and Similarity Retrieval of image Using 2D C-string Knowledge Representation: Pattern Recognition, vol.25, no.10, 1992, pp.305-318
- [6] G. Petraglia, G. Sebillio, M., Tucci G. Tortora, "Virtual Images for Similarity Retrieval in Image Databases", IEEE Trans. on knowledge and data engineering vol. 13(6) pp. 951-967, 2001
- [7] E.G.M. Petrakis., Fast Retrieval by Spatial Structure in Image Databases, Journal of Visual Languages and Computing, Vol. 13, No. 5, pp. 545-569, October 2002
- [8] E.G.M. Petrakis, "Content- Based Retrieval of Medical Images", International Journal of Computer Research, Volume 11, pp. 171-182, Number 2, 2002
- [9] M. Stoeva, B. Rachev, "Retrieval by Spatial Similarity in Image Databases", Proc. of 7th East European Conference, ADBIS 2003, pp.48-59, Dresden, Germany, 2003
- [10] M. Stoeva, V.Bozhikova, A New Approach to Symbol Description of the Spatial Location of Extended Objects in Spatial Databases, Proc. of XLII International Scientific Conference of Information Communication and Energy Systems and Technologies (ICEST 2007), Macedonia, Ohrid, vol.2, pp. 865-868, 2007

# Studying the process of software development through intensive use of software tools

Violeta Bozhikova<sup>1</sup>, Mariana Stoeva<sup>2</sup>, Dimitrichka Nikolaeva<sup>3</sup>, Zlatka Mateva<sup>4</sup>

**Abstract** – The paper focuses on the further development of our approach to teach the discipline of Software Engineering (SE) which is based on intensive use of software tools. The use of software tools in the process of software development is a way to increase both the effectiveness of teaching SE and the effectiveness of the software engineer's work. This approach was used during the "Software Engineering" course labs at the Technical University – Varna. The approach is intended to focus students' attention on both the use of software tools in all phases of software development as well on the tools development. An open for future development package of software tools is developed for this case; its current structure is discussed in the paper.

**Keywords** – Software Engineering, Software Engineering education, tools for Software Engineering education, tools supporting software development, etc.

## I. INTRODUCTION

"Software engineering" (SE) is a basic course for bachelor degree students in "Computer Science". It examines the different models of software life cycle and covers all phases and activities of the software production. The course is difficult for the students because of its highly theoretical nature. The popular SE books also do not easily attract the students' attention, because of their theoretical-style approach. As a result of everything mentioned above, students lose interest in the Software Engineering discipline.

Based on our long teaching experience in the field of Computer Sciences in the Technical University in Varna, we believe that teamwork and tool supported project-oriented exercises help students easily understand the theory included in SE textbooks.

This paper is about the further development of our approach to teach Software engineering discipline, which is based on intensive use of software tools ([3] and [5]). An open for

future development package of software tools ([1] ÷ [5]) is in the core of this approach; its current state is discussed in the paper.

## II. OUR APPROACH

In this section we discuss our approach to teach the discipline of Software Engineering which is based on intensive use of self developed software tools.

The discipline of Software Engineering is conducted in the third year of the student's curriculum. It brings 6 credits and includes lectures, practical labs and course project. The students work in teams of maximum 3 students, each student has his different role in the team.

In the first week, all teams receive by the teacher a list with practical software problems to resolve. The last week, each team must submit implementation and documentation of the entire list of software problems.

The student semester includes 15 weeks. The software tools used in the different phases of the software life cycle by weeks of semester are shown below in the Table 1:

TABLE I  
THE SOFTWARE TOOLS USED IN THE DIFFERENT PHASES OF THE SOFTWARE LIFE CYCLE BY WEEKS OF SEMESTER

Tool	SE Phase	Weeks
SCE	Requirement Analysis and Specification	Week 1 ÷ Week 2
SR&A	Operation & Maintenance	Week 13 ÷ Week 14
VLT	Design, Coding and Testing	Week 6 ÷ Week 12
PMS	Software Project Management	Week 1 ÷ Week 15
DPAP-T	Design, Coding and Testing	Week 6 ÷ Week 12
UML-T	Requirement Analysis and Specification	Week 3 ÷ Week 5

The teams work is based on evolutionary model for software development. The package of software tools (fig.1) used in solving the software problems is extensible and currently consists of six tools, supporting different software development activities (fig.2):

<sup>1</sup>Violeta Bozhikova is with the Department of Computer Sciences and Engineering, Technical University of Varna, 1 Studentska Str., Varna 9010, Bulgaria, E-mail: vbozhikova2000@yahoo.com.

<sup>2</sup>Mariana Stoeva is with the Department of Computer Sciences and Engineering, Technical University of Varna, 1 Studentska Str., Varna 9010, Bulgaria, E-mail: mariana\_stoeva@abv.bg.

<sup>3</sup>Dimitrichka Nikolaeva is a PhD student in the Department of Computer Sciences and Engineering, Technical University of Varna, 1 Studentska Str., Varna 9010, Bulgaria, E-mail: dima.nikolaeva@abv.bg

<sup>4</sup>Zlatka Mateva is with the Department of Computer Sciences and Engineering, Technical University of Varna, 1 Studentska Str., Varna 9010, Bulgaria, E-mail: ziz@abv.bg



Fig 1. The package of software tools

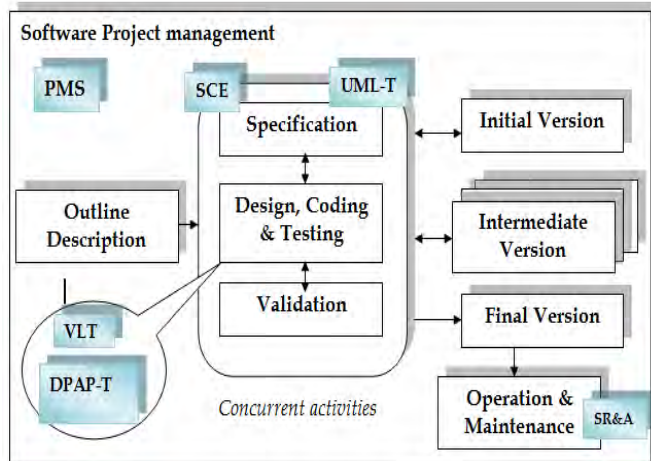


Fig. 2. Software tools in different phases of the software life cycle

- A tool that supports Software Project Management, named PMS.

This tool was in details discussed in ([3] and [5]). Project management software (PMS) is a term covering many types of software, including scheduling, cost control and budget management, resource allocation, collaboration software, communication, quality management and documentation or administration systems, which are used to deal with the complexity of large projects.

- A tool (VLT) with lecture notes, exercises and tests on programming.

VLT ([3] and [5]) is a case of educational software that aims at helping the student study the relevant Microsoft.Net programming language. Before doing the current laboratory exercises, the students could read the language materials. Choosing “Tests” pane the students could make the corresponding quiz and evaluate their learning progress. The teacher also can analyse the students' results, and get an overall view for the progress of the students. The architecture of the developed web-based tool provides instructors a possibility to easily create and manage different course materials. Although the tool doesn't impose limitation about the program language materials, the students are suggested to use a .Net language.

- A tool (SCE) that supports the software cost estimation ([3], [4] and [5]).

This tool implements a hybrid approach for software cost estimation [4], which is based on the classical models for software cost estimation: Basic COCOMO, COCOMOII and Function Points Analysis. Software estimation is the part of project planning aimed at size estimation, effort, time, people

required etc. Software cost estimation process gives the information needed to develop a software project's schedule, budget and assignment of personnel and resources.

- A tool for Software Re-structuring and Analysis (SR&A) is integrated in the package ([2], [3] and [5]).

This tool could support mainly the maintenance activity. Using this tool the students could analyze and re-structure the software with the idea to improve its structure. The software must be formally presented as a weighted oriented graph  $G=(X, U)$ , where the set of weighted nodes  $X$  ( $N=|X|$ ) models the software's components (classes, modules, files, packages, etc.) and  $U$  (the set of graph's edges) presents the dependencies between the components (e.g., procedural invocation, variable access, etc.).

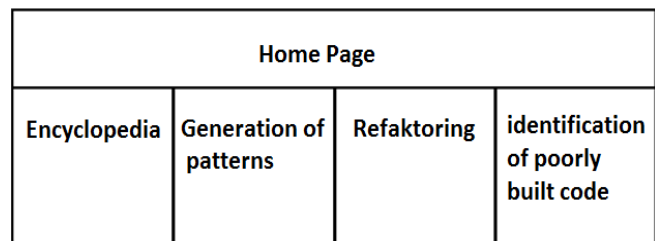


Fig. 3. The structure of DPAP-T

- A tool for software anti-pattern identification, named DPAP-T.

DPAP-T is a web-based application that contains home page and several sections representing different aspects of design patterns and anti-pattern support (Fig. 3).

Home page aims to present the different sections of the system with a short description and redirect the user to any of them.

Encyclopedia: it is a section of the system, which aims to provide information about design patterns and anti-patterns. This section describes the problems that each design pattern resolves; it describes the benefits of the use of design patterns and the situations in which it could be used. Anti-patterns are common solutions to problems that lead to negative consequences. As a result, they produce a poor code structure. The structure of the code is an important factor in maintaining and adding new functionality in one system.

Generation of patterns: this section provides functionality for design pattern generation (fig 4). An example of design pattern (abstract factory) generation is presented in figure 5.

Section “Refactoring” provides methods for automatic code refactoring. Refactoring is a reconstruction of the code with the goal to improve its properties. The code is supplied as input to the refactoring method. The method performs the appropriate changes and returns as a result the modified code. 8 refactoring methods are provided by the tool: “Extract Method”, “Inline method”, “Replace Temp with Query”, “Encapsulate Field”, “Replace Magic Number with Symbolic Constant”, “Replace Constructor with Factory Method” and “Self Encapsulate Field”.

Identification of poorly built code: this section offers methods for analysis of the code. It is supplied as input of a



particular method which identifies the poorly constructed code sections. The poorly build code could be rewritten to be more easily readable and to allow easy maintenance. 3 methods (fig.3) for poorly built code identification are provided by the tool: “Duplicated code”, “Too many parameters in a method”, “Complicated If’s”. Several algorithms that improve the software readability, software maintenance and the ability to add new functionality are provided, but still are waiting to be realized.

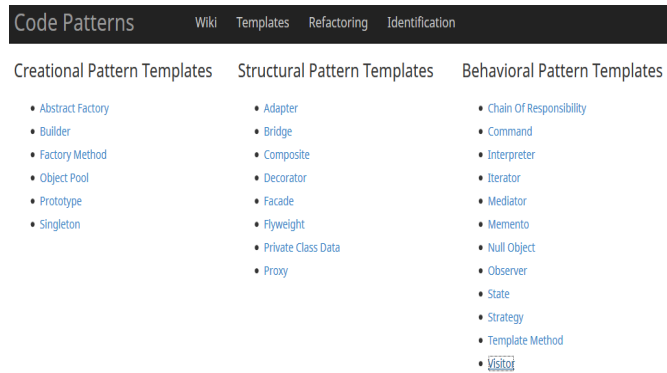


Fig. 4. Section “Generation of patterns”



Fig. 5. An example of design pattern (abstract factory) generation

- A tool for UML diagrams training and drawing, named UML-T.

This software tool (UML-T) provides theoretical and practical training for Unified Modeling Language (UML). It contains 2 modules: “UML textbook” and “UML editor” - (fig.6 and fig.7).

UML is a "universal" to the process of software development language for visual modeling. The visual modeling (visual modeling) is a method that:

- Use graphic (often based on graphs) model for software visualization;
- Offers subject area modeling from different viewpoints (perspectives);

- Supports almost all phases and activities of the software development process; it can be used for both: the development of new software and the evolution of the existing software.

The diagrams are the most important building elements in UML which can be used for the purposes of the:

- Specification of the system requirements and the system architecture;
- System documentation;
- Model simulation (with automatically generated code);
- Model testing and validation;
- Communication between project participants, etc.

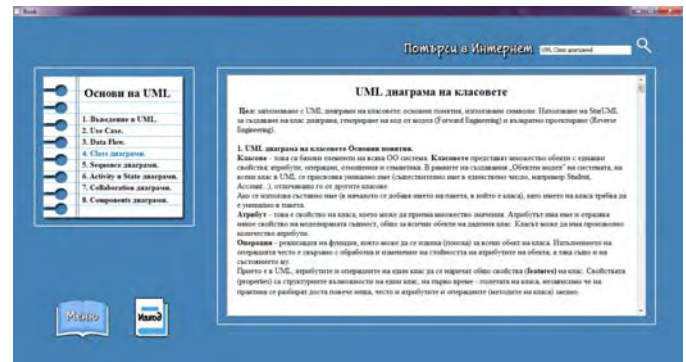


Fig. 6. UML textbook

UML-T supports the drawing of the basic UML diagrams through “UML editor” module. It provides an intuitive and very simple interface. Compared with the existing commercial and free UML tools, UML-T provides a training subsystem (something as e-book for UML), supports the modeling of Data Flow diagrams, but it is still in its initial phase of development and does not yet support the entire set of charts and Reverse/Forward Engineering options. As a conclusion, many things in the current realization of UML-T wait to be improved.

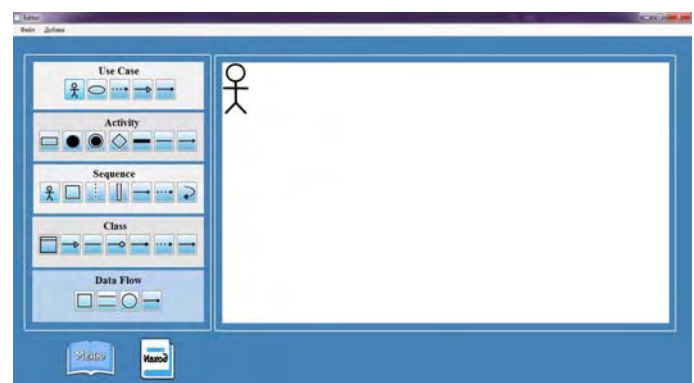


Fig. 7. UML editor

### III. CONCLUSION

The paper focuses on the further development of our approach to teach the discipline of Software Engineering which is based on intensive use of software tools. An open for future development package of software tools is in the core of this approach; its further development is discussed in this paper.

Our observation is that the effect of the approach application for SE educational purposes is positive. The student's interest in SE discipline has grown. It is intended to focus students' attention on both the use of software tools in all phases of software development as well on the tools development. The students acquire useful theoretical and practical knowledge in software analysis and specifications, software cost estimation, software maintenance, software development and so on.

Further work is needed in order to improve and to enrich the teaching package: the existing tools could be improved and unified; new functionalities could be added; new tools could be included in the package.

### ACKNOWLEDGEMENT

The authors express their gratitude to the students Bozhidar Georgiev and Nely Evgenieva, who participated actively in the development of the package resources. The paper was sponsored by two research projects in TU Varna: HPI8 and ПД6.

### REFERENCES

- [1] Violeta Bozhikova, Mariana Stoeva, Krasimir Tsonev, A practical approach for software project management, CompSysTech Conference 2009, Russe (to appear)
- [2] V.Bozhikova, M. Stoeva, A. Antonov, V. Nikolov, Software Re-structuring (An architecture-Based Tool), ICSOFT 2008, Third Int'l Conference on Software and data Technologies, pp.269-273, Porto, Portugal, 2008.
- [3] Bozhikova V., N. Ruskova – “A Computer-Based Approach for Software Engineering Teaching”, Proceedings of the International Conference “E-Learning and the Knowledge Society”, Berlin, GERMANY, 2009, pp. 161-166, ISBN 1313-9207.
- [4] Bozhikova V., M. Stoeva – “An Approach for Software Cost Estimation”, Proc. of the International conference on computer systems and technologies (CompSysTech'2010), International Conference Proceedings series - V.471, София, Bulgaria, 2010, pp.119 -124, ACM ISBN 978-1-4503-0243-2.
- [5] Bozhikova V., M. Stoeva, N. Ruskova – “A tool package for software engineering teaching”, Материали XI Международная конференция „Стратегия качества в промышленности и образовании”, Том 1, Варна, 2015, pp. 313-318.

# Statistical Analysis of Text Encryption Software

Darko Brodić<sup>1</sup>, Ivo R. Draganov<sup>2</sup> and Zoran N. Milivojević<sup>3</sup>

**Abstract** – The paper describes the problem related to the efficacy of different text encryption programs. Many of the known programs uses similar techniques along with different keys to encrypt and decode the encrypted text. Still, many of them have different efficacy in the relation between output and input data, i.e. encrypted and original data. Also, some of them are more adjusted to different text document, which includes various languages and scripts as well as multilingual documents, too. This study statistically analyses aforementioned problems on the example of 5 different text encryption programs. The obtained results show meaningful differences between different encryption programs, which were under research process.

**Keywords** – Algorithm, Cipher, Cryptography, Encryption, Software, Statistical Analysis.

## I. INTRODUCTION

Encryption represents a process of transforming accessible data into an unreadable code that cannot be understood by normal means [1]. The encryption process uses a key as well as an algorithm to turn the accessible data into an encoded data. To decode encoded data or to convert encoded data to the original form, the ciphering algorithm and a secret key are needed. Hence, the encryption software includes the algorithm for encoding data in order that unauthorized users cannot access it [2]. These data can be accessed only if we have an appropriate key for decoding. Typically, the encryption process is used to protect our personal or business sensitive data. In this way, encryption software comprises different methods that may encrypt files, directories and drives to ensure data privacy. Currently, the application of this software is mainly in the area of archiving a large amounts of data or secure private communication over the Internet.

A whole group of different encryption software solution has been developed. Many of them differ according to their simplicity or complexity of included algorithms. Some of them includes less or more complex keys for coding and decoding. In the recent times, more and more solutions represent a freeware software ones.

The base of this study incorporates the research of different encryption programs and their implication on data that were subject to the encryption. In this way, the inputs to the encryption programs represent text documents given in

different languages and scripts. The outputs represent encrypted data. The study statistically analyzes the output-input relationship. The conclusions are drawn according to the efficacy of the encryption level and suitability of using documents written in different languages and scripts.

The paper is organized in the following manner. Section 2 describes the encryption software under the research. Section 3 defines the experiment as well as the statistical measures used for comparison. Section 4 presents the results and gives the discussion. Section 5 draws conclusions.

## II. TEXT ENCRYPTION PROGRAMS

In this study, the following encryption programs were under consideration and evaluation: (i) Arcanum Editor, (ii) BCTextEncoder, (iii) TheLetterEncrypter, (iv) Simple Text Encryptor, and (v) CryptX.

Arcanum Editor is a text encryption software. It is one of the best tools for encryption and decryption. It has multiple methods to encrypt your data. The various supported encryption algorithms are AES, Base64, Bytes, 1337 speak and Rot13. You can set an encryption password when using AES algorithm, while others simply encrypt/decrypt your text. It also has a Hash feature.

BCTextEncoder is a free text encryption software. It can encrypt typed text and text files. The encrypted text can be saved in a text file or copied from the clipboard. It has two types of encryption methods; one is password based encryption and the second is public key based encryption. It doesn't require installation, just download it and run it.

TheLetterEncrypter is a free text encryption software. You can encrypt or decrypt your text using any password of your choice. You can also import text file to encrypt and export the encrypted text to a file. It uses strong Rijndael Encryption algorithm to encrypt text. You can copy encrypted text to your email message to send. It requires no installation.

Simple Text Encryptor is a free and simple program that encrypts and decrypts text using 128-bit AES encryption method. It comes in two versions to download; one is an installer and other is stand alone zipped version. Zip version doesn't require installation and you can use it from any portable drive also. You can set or generate a random key to encrypt your text easily. It has very simple interface and is easy to use.

CryptX is a free text encryption software. You can encrypt and decrypt text files with it easily. It supports three types of encryption methods: Binary, Base64 and Base64 without password. Only Base64 is password supported encryption method. It saves the encrypted file in the normal text file. It has different tabs for encryption and decryption. It also has a feature of computing hash and calculate MD5 checksum.

<sup>1</sup>Darko Brodić is with the Technical Faculty in Bor, University of Belgrade, Vojske Jugoslavije 12, 19210 Bor, Serbia, E-mail: [dbrodic@tf.bor.ac.rs](mailto:dbrodic@tf.bor.ac.rs)

<sup>2</sup>Ivo R. Draganov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: [idraganov@tu-sofia.bg](mailto:idraganov@tu-sofia.bg)

<sup>3</sup>Zoran N. Milivojević is with the College of Applied Technical Sciences, Aleksandra Medvedeva 20, 18000 Niš, Serbia, E-mail: [Zoran.milivojevic@vtsnis.edu.rs](mailto:Zoran.milivojevic@vtsnis.edu.rs)

### III. EXPERIMENT

The experiment is performed by encrypting a small database of text documents by all five chosen encryption programs. The database consists of 75 text documents, i.e. 25 English text documents, 25 Serbian text documents written in Latin script and 25 Serbian text documents written in Cyrillic script. The example of the text document samples is given in Fig. 1.

But in the ancient way of Himalayan travel, they're still 20 switchback roads and several torturous hours away.

(a)

U ostvarivanju ovog projekta, američki internet pretraživač se udružio sa šest fotografa, skijaša i planinara koji su sa kamerama u rancu prošli najpoznatijim stazama po letnjem i zimskom vremenu.

(b)

У остваривању овог пројекта, амерички интернет претраживач се удружио са шест фотографа, скијаша и планинара који су са камерама у ранцу прошли најпознатијим стазима по летњем и зимском времену.

(c)

Fig. 1. The example of the text document samples from the database: (a) English document, (b) Serbian Latin document, and (c) Serbian Cyrillic document

All documents included in the database represent a small text document. English documents consist of 75 to 176 characters, while Serbian Latin documents consist of 74 to 196 characters and Serbian Cyrillic documents consist of 73 to 196 characters. As a result, the input text is coded. The sample of encrypted documents is given in Fig. 2.

Lt6nrSTxksmvuc7I61ROMtnx9liVi8dRaq5P0Pe9ZKjnkag7D  
F2Q5ck5pOR9wPuk0btVctdu3wXvalvGoPr7ayp977aUAGT3  
MYNon4H/K2gYEZNVcuuky+OK7Y2dVRioL4Xj5WGRqu  
RL/bdxz/AIqkyHUvftOQGr

(a)

XUAReMDMKUJQj9SWWV/N4w6z7t/EDBfxxcc15ma6Dzs  
Y0L6xoCGdidZL2a51vAbQHH7vJBHWFN5vjjQ09cxlTLj4q  
EkfrtXqvuoGo/ZXfF2JI9UsBAWAfVqCPgnvGfXSgziF7EA  
aw3ILrnKr9W1imUBUG+ksXO2qXK3JovPwU4Q8DohyeD  
n3h9F0ZZ7Wpf7Ydpg2UB3q/ZL66hqWz1ZnazJHKm1meD  
7Nt91fiJCLWjp7G183kHuEYcHirkdztn5vExI+KtWug=

(b)

0l6b8krM8vETLxnOcCgh7qGbQ2IezvTxyqLRgas/nrYTLxn  
OcCgh7vvWqxnvl5/1sPf4NIJYW6Wg0Mi7uRt8TRMvGc5w  
KCHuGQGx7v6e0kSZsFnr3+9/1eiR160rKsRLn3ihAvp3gE  
A/bQ95+yzsB9VOScas7f22sPf4NIJYW6WfeKEC+neAQIQ  
HYoSNOIhFsPf4NIJYW6UTLxnOcCgh7hMvGc5wKCHuh  
AdihI04iEXSxpvySszy8dJem/JKzPLx/8jTYS9QtbU=

(c)

Fig. 2. Examples of the encrypted text document samples from the database: (a) encrypted English document, (b) encrypted Serbian Latin document, and (c) encrypted Serbian Cyrillic document

The output, which represents the coded text is statistically analyzed by typical statistical measures like mean, standard deviation and correlation coefficients. These measures are defined as follows, respectively:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i, \quad (1)$$

$$s = SD = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}, \quad (2)$$

$$R = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}}. \quad (3)$$

The correlation coefficient  $R$  expresses the strength and direction of a linear relationship between two variables  $x$  (input) and  $y$  (output), where  $x_i$  is the value of the variable  $x$  for the instance  $i$ ,  $\bar{x}$  is the mean of  $x$  value,  $y_i$  the value of the variable  $y$  for the instance  $i$ , and  $\bar{y}$  is the mean of  $y$  value [4]. Consequently,  $R$  measures the strength and direction of a linear relationship between two variables on a scatterplot [5]. It can receive the value from -1 to +1. The larger the absolute value of the coefficient, the stronger linear relationship between the variables (input and output). An absolute value of 1 indicates a perfect linear relationship. If  $R$  is positive, then the two variables tend to increase or decrease together. In contrast, if  $R$  is negative, then one variable increases as the other decreases.

### IV. RESULTS AND DISCUSSION

Table I shows the statistical values obtained after encryption of Serbian Latin documents by all five encryption programs (Output1 - Arcanum Editor, Output2 - BCTextEncoder, Output3 - TheLetterEncrypter, Output4 - Simple Text Encryptor, Output5 - CryptX).

TABLE I  
STATISTICAL MEASURES OF ENCRYPTED SERBIAN LATIN DOCUMENTS

	Output1	Output2	Output3	Output4	Output5
Min	108	236	108	160	108
Max	272	356	280	384	268
Mean	185.08	297.44	193.92	283.2	186.56
SD	43.69	28.78	43.60	60.21	41.80
R	0.9944	0.9768	0.9812	0.9814	0.9874

Table II shows the statistical values obtained after encryption of Serbian Cyrillic documents by all five encryption programs (Output1 - Arcanum Editor, Output2 -



BCTextEncoder, Output3 - TheLetterEncrypter , Output4 - Simple Text Encryptor, Output5 – CryptX).

TABLE II  
STATISTICAL MEASURES OF ENCRYPTED SERBIAN CYRILLIC DOCUMENTS

	Output1	Output2	Output3	Output4	Output5
Min	176	284	192	160	128
Max	476	420	492	416	268
Mean	324.8	349.76	337.44	282.88	197.44
SD	79.57	34.56	80.14	63.79	39.76
R	0.9992	0.9892	0.9952	0.9902	0.9894

Table III shows the statistical values obtained after encryption of English documents by all five encryption programs (Output1 - Arcanum Editor, Output2 - BCTextEncoder, Output3 - TheLetterEncrypter, Output4 - Simple Text Encryptor, Output5 – CryptX).

TABLE III  
STATISTICAL MEASURES OF ENCRYPTED ENGLISH DOCUMENTS

	Output1	Output2	Output3	Output4	Output5
Min	100	232	108	160	108
Max	236	324	256	384	248
Mean	158.12	273.92	170.68	248.32	168.48
SD	35.38	23.35	36.59	56.00	36.47
R	0.9992	0.9849	0.9835	0.9855	0.9957

From the results given in Tables I-III, it is quite clear that all encryption programs due to the high value of correlation coefficient  $R$  create linearly dependent output-input correlation. It means that the longer input document text will tend to create a longer coded output text. Furthermore, the analysis will be continued by comparison of the statistical analysis of documents written in different languages and/or scripts.

Fig. 3 shows the output-input relation in accordance to the number of characters obtained by the encryption program Arcanum Editor.

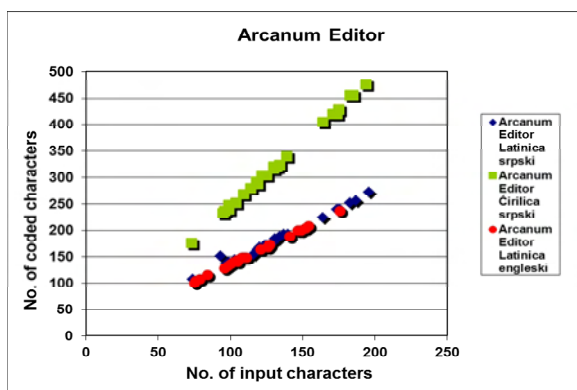


Fig. 3. Coded vs. input text obtained by Arcanum Editor

From Fig. 3, it is an obvious that the encryption of English and Serbian Latin documents has similar efficacy. However, the coding of Serbian Cyrillic documents creates larger coded

documents. Hence, the program is more appropriate for the Latin based languages.

Fig. 4 shows the output-input relation in accordance to the number of characters obtained by the encryption program BCTextEncoder.

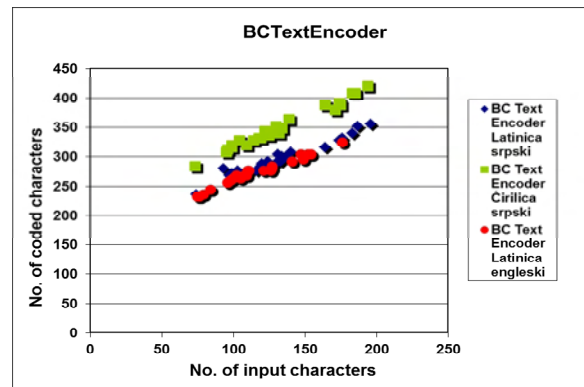


Fig. 4. Coded vs. input text obtained by BCTextEncoder

BCTextEncoder also depends on the type of the input scripts. Hence, English and Serbian Latin documents are similarly coded. As a consequence, Serbian Cyrillic documents are coded by creating larger documents.

Fig. 5 shows the output-input relation in accordance to the number of characters obtained by the encryption program TheLetterEncrypter.

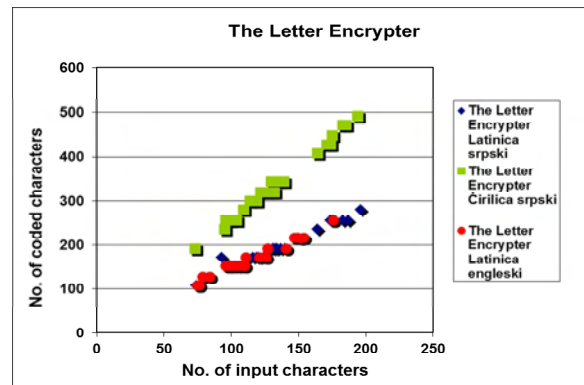


Fig. 5. Coded vs. input text obtained by TheLetterEncrypter

The situation with TheLetterEncrypter program is similar to previously analyzed programs.

Fig. 6 shows the output-input relation in accordance to the number of characters obtained by the encryption program Simple Text Encryptor.

Simple Text Encryptor similarly coded documents written by different languages/scripts. Hence, it has similar efficacy of encrypting document given by different languages and scripts.

Fig. 7 shows the output-input relation in accordance to the number of characters obtained by the encryption program CryptX.

CryptX program has similar way of encrypting documents in spite of their language or script. In this way, it is similar to the Simple Text Encryptor.

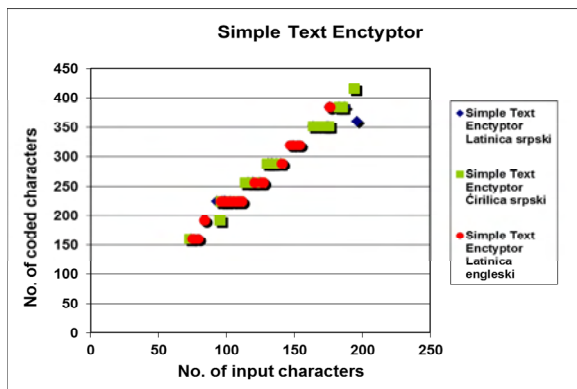


Fig. 6. Coded vs. input text obtained by Simple Text Encryptor

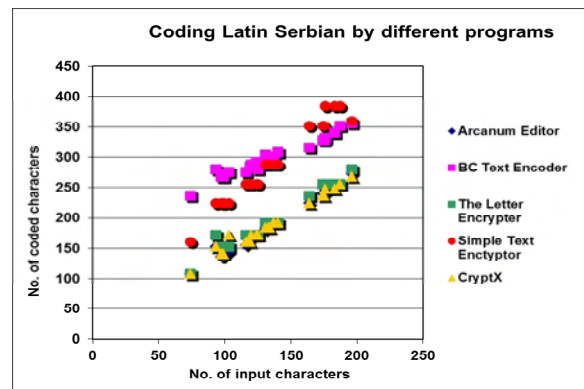


Fig. 9. Coded vs. input Serbian Latin text

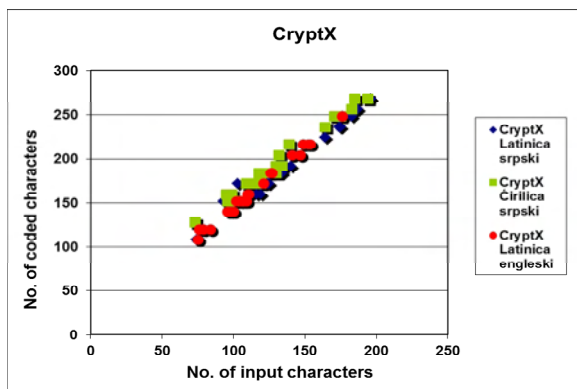


Fig. 7. Coded vs. input text obtained by CryptX

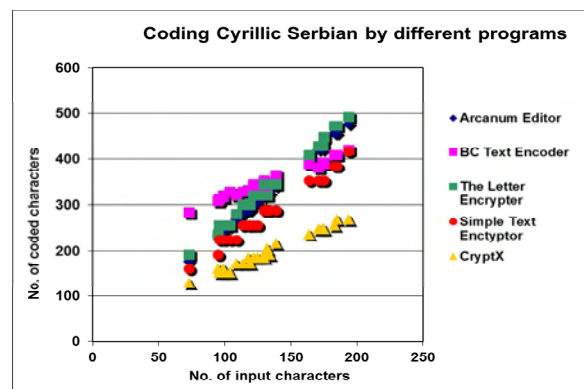


Fig. 10. Coded vs. input Serbian Cyrillic text

At the end, the comparison of different programs is established by giving their efficacy of encrypting documents in certain languages or scripts. Fig. 8 shows coded vs. input for English based documents.

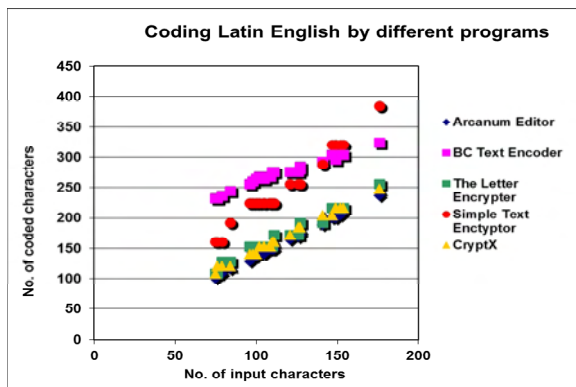


Fig. 8. Coded vs. input English text

Fig. 9 shows coded vs. input for Serbian Latin based documents, while Fig. 10 shows coded vs. input for Serbian Cyrillic based documents.

## V. CONCLUSION

This paper analyzed different types of encryption programs. Their efficacy of coding was in the focus. The statistical analysis showed similarities as well as dissimilarities between their efficacy of coding. Many of the programs are primarily adjusted to the Latin based languages. However, some of them like Simple Text Encryptor and CryptX are almost completely independent of the type of document that they are coding. However, CryptX program created smaller coded documents achieving better efficacy.

Future research will include wider range of encryption programs as well as more complex database, which will include Chinese and Indian texts.

## REFERENCES

- [1] A.J. Menezes, P. C. van Oorschot, S. A. Vanstone, *Handbook of Applied Cryptography*, Boca Raton, CRC Press, 1996.
- [2] P. C. Paar, J. Pelzl, *Understanding Cryptography*, New York, Springer, 2010.
- [3] <http://listoffreeeware.com/list-best-free-text-encryption-software/>
- [4] J. Higgins, *The Radical Statistician: A Beginners Guide to Unleashing the Power of Applied Statistics in The Real World* (5th Ed.), Jim Higgins Publishing, 2006.
- [5] Chi Yau, *R Tutorial with Bayesian Statistics Using OpenBUGS*, Kindle Book, 2015.

# Electronics





# Analytical Modeling of a ZVS Bidirectional Series Resonant DC-DC Converter

Aleksandar Vuchev<sup>1</sup>, Nikolay Bankov<sup>2</sup>, Yassen Madankov<sup>3</sup> and Angel Lichev<sup>4</sup>

**Abstract** – The present article considers a bidirectional series resonant DC-DC converter, operating above the resonant frequency. The conditions of operation with ZVS are discussed. The processes in the converter resonant tank circuit are presented on the basis of well-known analytical models and their coefficients are defined. The obtained theoretical results are compared with those from simulation in the environment of OrCAD PSpice.

**Keywords** – Bidirectional Series Resonant DC-DC Converter, ZVS, Phase-Shift Control.

## I. INTRODUCTION

The bidirectional DC-DC converters are suitable for different applications where it is necessary to control the electrical energy transmission [3], such as in motor drives [4, 8], grids with renewable energy sources [6, 9], etc. They are realized on the base of two inverter stages. For connection between the inverters, a single inductance can be used [5, 9].

It is long known that series resonant DC-DC converters can also be used for bidirectional energy transfer [1, 3, 4]. They have various advantages, one of which is the possibility of significant switching loss reduction. When the converters operate above the resonant frequency, it is really possible that their power devices switch at zero voltage (ZVS – Zero Voltage Switching) [2]. This makes them a preferred solution for different applications.

There is a plenty of investigations [3, 4, 7, 10] of bidirectional series resonant DC-DC converters operating at higher than their resonant frequency. They show that the characteristics of the considered solutions depend on the applied control technique to a significant extend. It is known that these characteristics can be obtained as a result of modelling of the converter resonant tank circuit processes [2]. In general, this is the description of the variation of the

inductor current and the capacitor voltage. These models are well-known, but their coefficients depend on the converter circuit and the used control technique.

The current work presents analytical modelling of the resonant tank circuit processes of bidirectional series resonant DC-DC converter operating above the resonant frequency. For this purpose, well-known analytical models are used and their coefficients are determined for the considered circuit.

## II. PRINCIPLE OF THE CONVERTER OPERATION

Circuit of the examined converter is presented in Fig. 1. It consists of two identical bridge inverter stages, resonant tank circuit ( $L$ ,  $C$ ), matching transformer  $Tr$ , capacitive input and output filters ( $C_d$  и  $C_0$ ). Fig. 1 also presents the snubber capacitors  $C_1 \div C_8$  by which a zero voltage switching is obtained.

A voltage  $U_d$  is applied to the DC terminals of the „input” inverter stage (transistors  $Q_1 \div Q_4$  with freewheeling diodes  $D_1 \div D_4$ ), and a voltage  $U_0$  – to those of the „output” stage (transistors  $Q_5 \div Q_8$  with freewheeling diodes  $D_5 \div D_8$ ).

Depending on the energy transfer direction, two operating modes are possible for the converter. The first of them is called **DIRECT MODE**. In this mode, it is assumed that energy flows from the „input” to the „output”, i.e. from the source of voltage  $U_d$  to the one of voltage  $U_0$ . The second is **REVERSE MODE**. In this mode, the energy flows from the

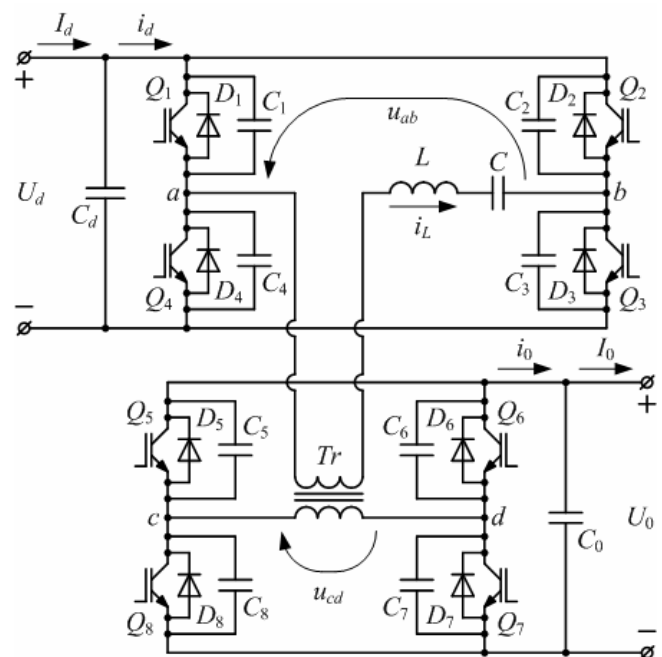


Fig. 1. Circuit of the Bidirectional Resonant DC/DC Converter

<sup>1</sup>Aleksandar Vuchev is with the Department of Electrical Engineering and Electronics, Technical Faculty, University of Food Technologies, 26 Maritza Blvd., 4002 Plovdiv, Bulgaria, E-mail: avuchev@yahoo.com

<sup>2</sup>Nikolay Bankov is with the Department of Electrical Engineering and Electronics, Technical Faculty, University of Food Technologies, 26 Maritza Blvd., 4002 Plovdiv, Bulgaria, E-mail: nikolay\_bankov@yahoo.com

<sup>3</sup>Yassen Madankov is with the Department of Electrical Engineering and Electronics, Technical Faculty, University of Food Technologies, 26 Maritza Blvd., 4002 Plovdiv, Bulgaria, E-mail: yassen.madankov@gmail.com

<sup>4</sup>Angel Lichev is with the Department of Electrical Engineering and Electronics, Technical Faculty, University of Food Technologies, 26 Maritza Blvd., 4002 Plovdiv, Bulgaria, E-mail: angel\_lichev@abv.bg

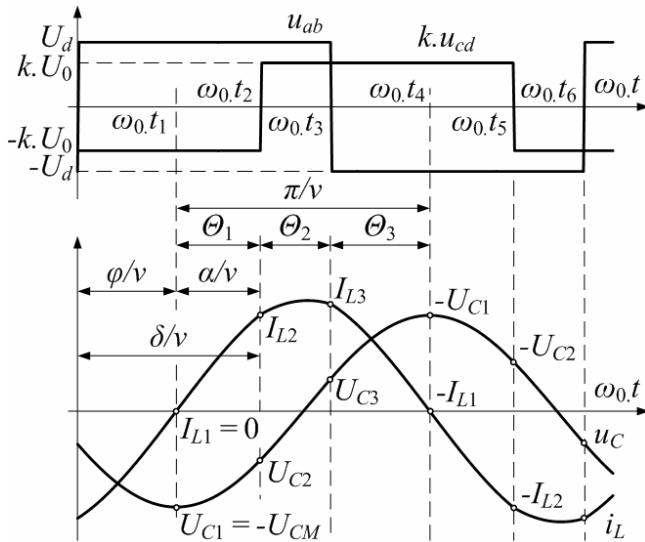


Fig. 2. Waveforms at **DIRECT MODE**

„output” to the „input” (from  $U_0$  to  $U_d$ ).

The waveforms, shown in Fig. 2, illustrate the converter operation in **DIRECT MODE**, and those in Fig. 3 – in **REVERSE MODE**.

Independently from the mode, the converter operates at constant frequency  $\omega_s$ , which is higher than the resonant  $\omega_0$ . Therefore, the transistor pairs of the „input” stage ( $Q_1, Q_3$  or  $Q_2, Q_4$ ) operate at ZVS. They are controlled in such a way that the voltage  $u_{ab}$  has almost square-wave form and amplitude  $U_d$ . Along with this, the resonant current  $i$  falls behind the voltage  $u_{ab}$  at an angle  $\varphi$ . The transistors of the „output” stage also operate at ZVS. Therefore, when the current  $i$  passes through zero, the corresponding pair ( $Q_5, Q_7$  or  $Q_6, Q_8$ ) begins conducting. This pair switches off after time, corresponding to an angle  $\delta$ , accounted to the moment of switch-off of the „input” stage transistors that conducted in the same half cycle. The voltage  $u_{cd}$ , which also has almost square-wave form and amplitude  $U_0$ , is shifted in time from  $u_{ab}$ . In this way, the output power control is obtained by phase-shifting – the variation of the angle  $\delta$ .

Angle  $\varphi$  corresponds to the conduction time of the „input” stage freewheeling diodes, and angle  $\alpha$  – to the conduction time of the „output” stage transistors. When  $\varphi < \pi/2$  and  $\alpha < \pi/2$ , energy is transferred in „forward” direction, and when  $\varphi > \pi/2$  and  $\alpha > \pi/2$  – in „reverse” direction. From Figs. 2 and 3, it can be observed that the following equation is valid:  $\varphi + \alpha = \delta$ . Therefore, when  $\delta > \pi$ , **REVERSE MODE** is observed, and when  $\delta < \pi$  – **DIRECT MODE**.

Angles  $\varphi$ ,  $\alpha$  and  $\delta$  are related to the operating frequency  $\omega_s$ .

### III. MODELING OF THE CONVERTER OPERATION

For the purposes of the analysis, the following assumptions are made: the matching transformer is ideal with a transformation ratio  $k$ , all the circuit elements are ideal, the influence of the snubber capacitors  $C_1 \div C_8$  and the ripples of the „input” voltage  $U_d$  and the „output” voltage  $U_0$  are neglected, i.e. voltages  $u_{ab}$  and  $u_{cd}$  have rectangular shape.

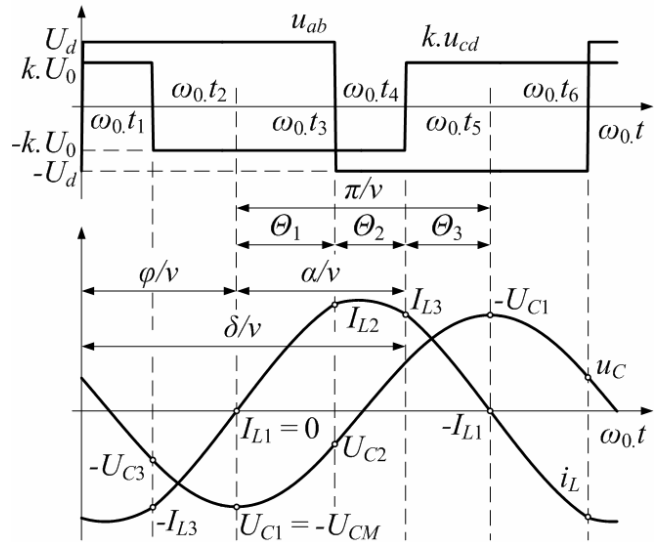


Fig. 3. Waveforms at **REVERSE MODE**

The waveforms (Figs. 2 and 3) show that, independently from the converter mode, any of the half cycles can be divided into three intervals. For each of them, an equivalent DC voltage  $U_{EQ}$ , defined by the values of  $u_{ab}$  and  $u_{cd}$ , is applied to the resonant tank circuit. This fact allows only the resonant tank circuit processes to be investigated. Therefore, apart from the waveforms of the current  $i_L$  through the inductor  $L$  and the voltage  $u_C$  across the capacitor  $C$ , Figs. 2 and 3 show the initial values ( $I_{L1} \div I_{L3}$ ,  $U_{C1} \div U_{C3}$ ) for each of the mentioned intervals.

In accordance with the assumptions made, the resonant frequency, the characteristic impedance and the frequency detuning are:

$$\omega_0 = 1/\sqrt{LC}; \rho_0 = \sqrt{L/C}; v = \omega_s / \omega_0 \quad (1)$$

Because of the similarity, the modelling of the resonant tank circuit processes is realized in the same way as in [2]. For each of the half cycle intervals, the current  $i_L$  and the voltage  $u_C$  values are determined as follows:

$$\begin{aligned} i_{Lj}(\theta) &= I_{Lj} \cos \theta - \frac{U_{Cj} - U_{EQj}}{\rho_0} \sin \theta \\ u_{Cj}(\theta) &= \rho_0 I_{Lj} \sin \theta + (U_{Cj} - U_{EQj}) \cos \theta + U_{EQj} \end{aligned} \quad (2)$$

where  $j$  is the number of the considered interval;  $I_{Lj}$  and  $U_{Cj}$  are the inductor current and the capacitor voltage values at the beginning of the interval;  $\theta = 0 \div \theta_j$ ;  $\theta_j$  – angle of the interval for the resonant frequency  $\omega_0$ ;  $U_{EQj}$  – the value of the equivalent DC voltage applied to the resonant tank circuit.

Actually, the parameters  $I_{Lj}$ ,  $U_{Cj}$  и  $U_{EQj}$  represent coefficients of the modelling Eqs. (2).

In order to obtain generalized results, all the quantities are normalized as follows: the voltages according to  $U_d$ , and the currents – according to  $U_d/\rho_0$ . Then, Eqs. (2) are transformed and in normalized form are:

$$\begin{aligned} i'_{Lj}(\theta) &= I'_{Lj} \cos \theta - (U'_{Cj} - U'_{EQj}) \sin \theta \\ u'_{Cj}(\theta) &= I'_{Lj} \sin \theta + (U'_{Cj} - U'_{EQj}) \cos \theta + U'_{EQj} \end{aligned} \quad (3)$$

where the normalized values of the corresponding quantities are marked with the prime symbol.

From Figs. 2 and 3 it is observed that the value of  $i_L$  at the end of given interval appears to be initial value for the following one. The same applies to the voltage  $u_C$ . Therefore:

$$\begin{aligned} I'_{Lj+1} &= I'_{Lj} \cos \Theta_j - (U'_{Cj} - U'_{EQj}) \sin \Theta_j \\ U'_{Cj+1} &= I'_{Lj} \sin \Theta_j + (U'_{Cj} - U'_{EQj}) \cos \Theta_j + U'_{EQj} \end{aligned} \quad (4)$$

Taking into consideration that  $I_{L3} = -I_{L1}$  and  $U_{C3} = -U_{C1}$ , after recursion of the Eqs. (4) the following system is obtained for the three consecutive intervals in a half cycle:

$$\begin{aligned} I'_{L1} \left( 1 + \cos \frac{\pi}{v} \right) - U'_{C1} \sin \frac{\pi}{v} \\ = \sum_{j=1}^3 U'_{EQj} \left[ \sin \left( \frac{\pi}{v} - \sum_{i=1}^j \Theta_i \right) - \sin \left( \sum_{i=j}^3 \Theta_i \right) \right] \end{aligned} \quad (5)$$

$$\begin{aligned} I'_{L1} \sin \frac{\pi}{v} + U'_{C1} \left( 1 + \cos \frac{\pi}{v} \right) \\ = \sum_{j=1}^3 U'_{EQj} \left[ \cos \left( \sum_{i=j}^3 \Theta_i \right) - \cos \left( \frac{\pi}{v} - \sum_{i=1}^j \Theta_i \right) \right] \end{aligned} \quad (6)$$

It is convenient the interval at which the initial values are  $I_{L1} = 0$  and  $U_{C1} = -U_{CM}$  to be chosen as first. In this case, the parameters  $\Theta_j$  and  $U'_{EQj}$  are defined on the base of the waveforms (Figs. 2 and 3). Their values are presented in Table I for each of the two operating modes.

TABLE I

MODE	Parameter	Number of interval		
		1	2	3
DIRECT	$\Theta_j$	$\frac{\delta - \varphi}{v}$	$\frac{\pi - \delta}{v}$	$\frac{\varphi}{v}$
	$U'_{EQj}$	$1 + kU'_0$	$1 - kU'_0$	$-1 - kU'_0$
REVERCE	$\Theta_j$	$\frac{\pi - \varphi}{v}$	$\frac{\delta - \pi}{v}$	$\frac{\pi - \delta + \varphi}{v}$
	$U'_{EQj}$	$1 + kU'_0$	$-1 + kU'_0$	$-1 - kU'_0$

As  $I'_{L1} = 0$ , after substitution of the actual values for  $U'_{EQj}$  in Eqs. (5) and (6), expressions for the voltages  $U'_0$  and  $U'_{CM}$  are obtained. Thus, for the **DIRECT MODE** the following normalized dependencies are obtained:

$$U'_0 = \frac{1}{k} \frac{\sin(\Theta_1 + \Theta_2) - \sin(\Theta_3)}{\sin(\Theta_2 + \Theta_3) - \sin(\Theta_1)} \quad (7)$$

$$U'_{CM} = 2 \frac{\sin(\Theta_3) + U'_0 \sin(\Theta_2 + \Theta_3)}{\sin(\pi/v)} - (1 + kU'_0) \quad (8)$$

Respectively, for the **REVERCE MODE** the following is obtained:

$$U'_0 = \frac{1}{k} \frac{\sin(\Theta_1) - \sin(\Theta_2 + \Theta_3)}{\sin(\Theta_3) - \sin(\Theta_1 + \Theta_2)} \quad (9)$$

$$U'_{CM} = 2 \frac{\sin(\Theta_2 + \Theta_3) + U'_0 \sin(\Theta_3)}{\sin(\pi/v)} - (1 + kU'_0) \quad (10)$$

When the values for the angles  $\Theta_j$  are substituted in the above equations, identical expressions are obtained for the two operating modes:

$$U'_0 = \frac{1}{k} \frac{\sin\left(\frac{\pi - \varphi}{v}\right) - \sin\left(\frac{\varphi}{v}\right)}{\sin\left(\frac{\pi - \delta + \varphi}{v}\right) - \sin\left(\frac{\delta - \varphi}{v}\right)} \quad (7)$$

$$U'_{CM} = 2 \frac{\sin\left(\frac{\varphi}{v}\right) + U'_0 \sin\left(\frac{\pi - \delta + \varphi}{v}\right)}{\sin\left(\frac{\pi}{v}\right)} - (1 + kU'_0) \quad (8)$$

Now, after calculation of the  $U'_0$  and  $U'_{CM}$  values, the other coefficients of the model – the initial values of the current  $i'_L$  through the inductor and the voltage  $u'_C$  across the capacitor, can also be determined for the second and the third intervals. For this purpose, Eqs. (4) are used.

Analyzing the obtained dependencies, it can be observed that all the model coefficients are expressed by two angles. One of them is the control angle  $\delta$ , which represents a parameter. The other angle is  $\varphi$ . Its value depends on the converter load. In other words, the model coefficients are determined in accordance with the value of the control parameter  $\delta$  for a determined load expressed by the angle  $\varphi$ .

TABLE II

MODE	Angle $\delta$	Angle $\varphi$
DIRECT	$0 \div \pi/2$	$0 \div \delta$
	$\pi/2 \div \pi$	$0 \div \pi/2$
REVERCE	$\pi \div 3\pi/2$	$\pi/2 \div \pi$
	$3\pi/2 \div 2\pi$	$\delta - \pi \div \pi$

During the calculations, the limits of the angle  $\varphi$  variation must be taken into consideration. They depend on the operating mode, as well as, on the value of angle  $\delta$ . The possible limits are presented in Table II.

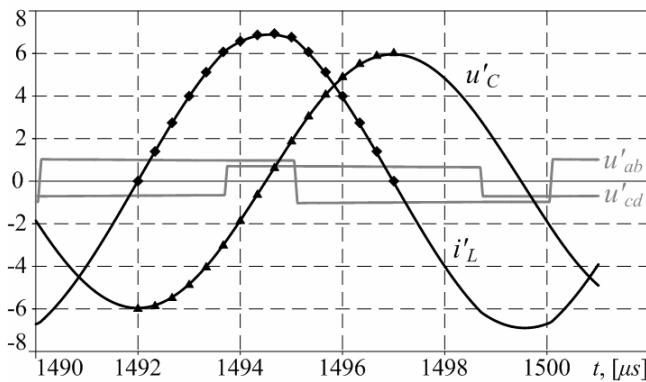


Fig. 4. Comparison at **DIRECT MODE**

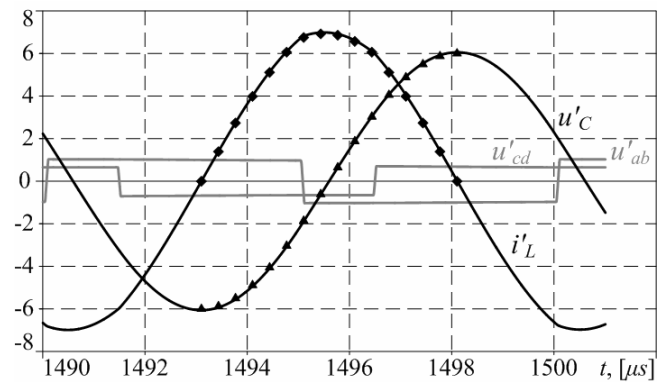


Fig. 5. Comparison at **REVERSE MODE**

#### IV. RELIABILITY OF THE RESULTS

For verification of their authenticity, the results from the accomplished modeling are compared to ones obtained from computer simulations. For this purpose, a model of a bidirectional series resonant DC-DC converter operating above the resonant frequency is realized in the environment of OrCAD PSpice. The simulation model has the following main parameters: Input source voltage –  $U_d = 200\text{V}$ ; Resonant tank circuit inductance –  $L = 223,332\mu\text{H}$ ; Resonant tank circuit capacitance –  $C = 15\text{nF}$ ; Operating frequency –  $f_s = 100\text{kHz}$ ; Frequency detuning –  $\nu = 1,15$ ; Transformation ratio  $k = 1$ .

The obtained simulation results are normalized in the same way as the values from the above presented modelling. Comparison in graphic form for the **DIRECT MODE** is presented in Fig. 4, and for the **REVERSE MODE** – in Fig. 5. The waveforms of the voltages  $u_{ab}$ ,  $u_{cd}$ ,  $u_C$  and the current  $i_L$  obtained from the computer simulations are shown with thick lines. The calculation results are presented by means of symbols (♦ – for the current  $i_L$  and ▲ – for the voltage  $u_C$ ). The examinations for the **DIRECT MODE** are accomplished at  $\delta = 0,722\pi$  and  $\varphi = 0,389\pi$  ( $U'_0 = 0,67972$ ), and for the **REVERSE MODE** –  $\delta = 1,278\pi$  and  $\varphi = 0,611\pi$  ( $U'_0 = 0,67972$ ).

A very good coincidence between the results obtained from the simulations and from the calculations is observed from the charts. This is also confirmed by the accomplished examinations for other converter operating points.

#### V. CONCLUSION

A modelling of the resonant tank circuit processes of bidirectional series resonant DC-DC converter operating above the resonant frequency is executed. For this purposes, well-known analytical models are used with their coefficients being determined for the considered circuit.

Computer simulations are accomplished, the results of which show very good coincidence with those from the analytical modelling.

The obtained models can be used for future investigation and design of such converters.

#### REFERENCES

- [1] Y. Cheron, H. Foch, J. Roux, "Power Transfer Control Methods in High Frequency Resonant Converters", PCI Proceedings, Munich, June 1986, pp. 92-103.
- [2] K. Alhaddad, Y. Cheron, H. Foch, V. Rajagopalan, "Static and Dynamic Analysis of a Series-Resonant Converter Operating above its Resonant Frequency", PCI Proceedings, Boston, October 1986, pp. 55-68
- [3] Patent № 4717990, U.S. H04M 7/00; H02J 3/38: Cheron Y., P. Jacob, J. Salesse; "Static Device for Control of Energy-Exchange between Electrical Generating and/or Receiving Systems", January 1988.
- [4] Dixneuf Daniel, "Etud d'un variateur de vitesse à résonance pour machine asynchrone triphasée", Thèse, 1988.
- [5] H.L. Chan, K.W.E. Cheng, D. Sutanto, "A Phase-Shift Controlled Bi-directional DC-DC Converter", 42<sup>nd</sup> Midwest Symposium on Circuits and Systems, Vol. 2, pp. 723-726, 1999.
- [6] S. Jalbrzykowski, T. Citko, "A bidirectional DC-DC converter for renewable energy systems", Bulletin of the Polish Academy of Sciences, Technical Sciences, Vol. 57, No. 4, pp. 363-368, 2009.
- [7] K. Ramya, R. Sundaramoorthi, "Bidirectional Power Flow Control using CLLC Resonant Converter for DC Distribution System", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 3, Iss. 4, pp. 8716-8725, April 2014, ISSN: 2278 – 8875
- [8] Ancy A.M, Siby C Arjun, "A Resonant Converter Topology for Bidirectional DC DC Converter" International Journal of Engineering Research & Technology, Vol. 3, Iss. 9, pp. 469-473, September 2014, ISSN: 2278-0181
- [9] D. Dharuman, C. Thulasiyammal, "Power Flow Control Using Bidirectional Dc/Dc Converter for Grid Connected Photovoltaic Power System", International Journal of Innovative Research in Electronics and Communications, Vol. 1, Iss. 8, pp. 13-24, November 2014, ISSN 2349-4050
- [10] Y. Jang, M. M. Jovanović, J. M. Ruiz, G. Liu, "Series-Resonant Converter with Reduced-Frequency-Range Control", Applied Power Electronics Conference and Exposition (APEC), 2015 IEEE, pp. 1453-1460, 15-19 March 2015, Charlotte, NC



# Load and Control Characteristics of a ZVS Bidirectional Series Resonant DC-DC Converter

Aleksandar Vuchev<sup>1</sup>, Nikolay Bankov<sup>2</sup>, Yassen Madankov<sup>3</sup> and Angel Lichev<sup>4</sup>

**Abstract** – The present paper considers a bidirectional series resonant DC-DC converter, operating above the resonant frequency. On the basis of a well-known analytical modelling of the resonant tank circuit processes, a sequence for determination of the main converter quantities is proposed. As a result, its normalized load and control characteristics are built.

**Keywords** – Bidirectional Series Resonant DC-DC Converter, ZVS, Phase-Shift Control.

## I. INTRODUCTION

The well-known series resonant DC-DC converter operating at frequency higher than the resonant one [2] has a significant disadvantage. It does not allow energy to be transferred back to the power supply source, which is necessary for a significant number of applications. The most common solution to this problem is the use of controllable rectifier [1] combined with a phase-shifting control. In this way, the converter becomes bidirectional with preservation of the option for ZVS (Zero Voltage Switching).

It is known that the characteristics of each converter are to a significant extent defined from the used control method.

In [3], a time-domain analysis of bidirectional series resonant DC-DC converter operating above the resonant frequency is presented. As a result, modelling of the converter resonant tank circuit processes is accomplished and expressions for the model coefficients are obtained.

The current paper presents sequel of the theoretical examinations achieved in [3]. Its purpose is load and control characteristics of the bidirectional series resonant DC-DC converter to be obtained with phase-shift control and operation at constant frequency above the resonant one.

<sup>1</sup>Aleksandar Vuchev is with the Department of Electrical Engineering and Electronics, Technical Faculty, University of Food Technologies, 26 Maritza Blvd., 4002 Plovdiv, Bulgaria, E-mail: avuchev@yahoo.com

<sup>2</sup>Nikolay Bankov is with the Department of Electrical Engineering and Electronics, Technical Faculty, University of Food Technologies, 26 Maritza Blvd., 4002 Plovdiv, Bulgaria, E-mail: nikolay\_bankov@yahoo.com

<sup>3</sup>Yassen Madankov is with the Department of Electrical Engineering and Electronics, Technical Faculty, University of Food Technologies, 26 Maritza Blvd., 4002 Plovdiv, Bulgaria, E-mail: yassen.madankov@gmail.com

<sup>4</sup>Angel Lichev is with the Department of Electrical Engineering and Electronics, Technical Faculty, University of Food Technologies, 26 Maritza Blvd., 4002 Plovdiv, Bulgaria, E-mail: angel\_lichev@abv.bg

## II. PRINCIPLE OF THE CONVERTER OPERATION

Circuit of the examined converter is presented in Fig.1. It consists of two identical bridge inverter stages, resonant tank circuit ( $L$ ,  $C$ ), matching transformer  $Tr$ , capacitive input and output filters ( $C_d$  и  $C_0$ ). Fig.1 also presents the snubber capacitors  $C_1 \div C_8$  by which a zero voltage switching is obtained.

A voltage  $U_d$  is applied to the DC terminals of the „input” inverter stage (transistors  $Q_1 \div Q_4$  with freewheeling diodes  $D_1 \div D_4$ ), and a voltage  $U_0$  – to those of the „output” stage (transistors  $Q_5 \div Q_8$  with freewheeling diodes  $D_5 \div D_8$ ).

The converter operation is discussed in details in [3] with the possible operating modes being defined. The first of them is called **DIRECT MODE**. In this mode, it is assumed that energy is transmitted from the source with voltage  $U_d$  to the one with voltage  $U_0$ . In the second mode – **REVERSE MODE**, energy flows in reverse direction – from  $U_0$  to  $U_d$ .

The presented in Fig. 2 waveforms illustrate the converter operation in **DIRECT MODE**, and those in Fig. 3 – in **REVERSE MODE**.

The „input” stage generates the voltage  $u_{ab}$ , which has square-wave form and amplitude  $U_d$ . Because of the fact that the converter operates at constant frequency  $\omega_s$ , which is higher than the resonant  $\omega_0$ , the current  $i_L$  falls behind the voltage  $u_{ab}$  at an angle  $\varphi$ . The „output” stage generates the

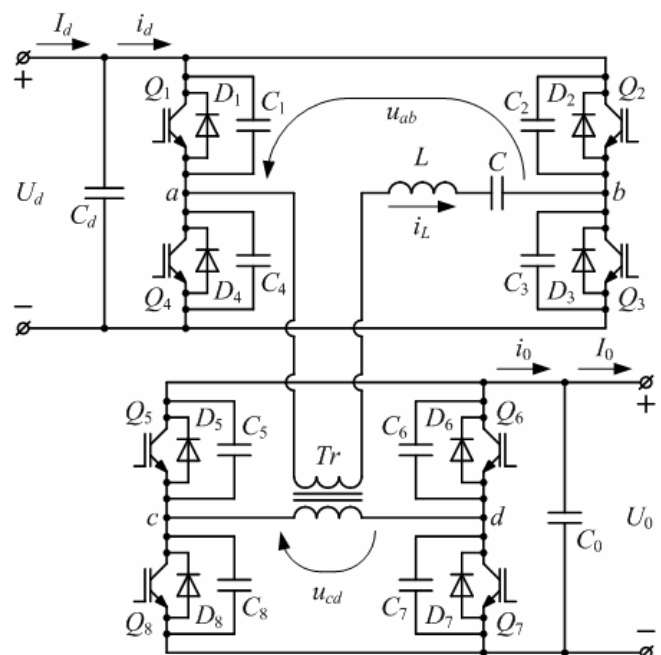


Fig. 1. Circuit of the Bidirectional Resonant DC/DC Converter

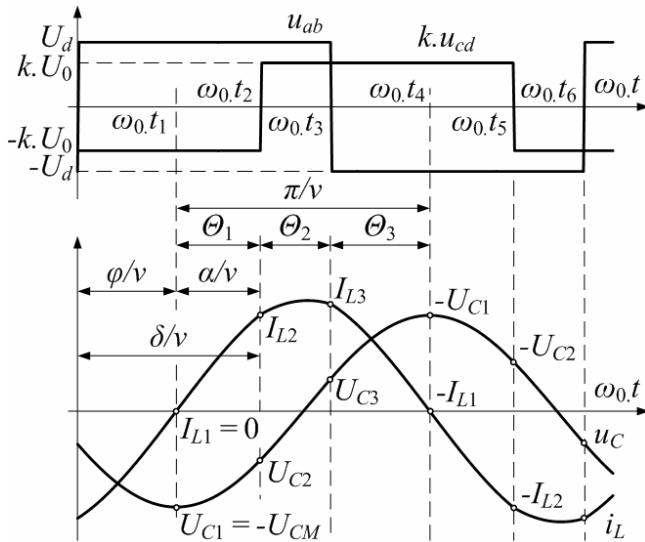


Fig. 2. Waveforms at **DIRECT MODE**

voltage  $u_{cd}$ , which also has square-wave form and amplitude  $U_0$ . This voltage is shifted from  $u_{ab}$  at an angle  $\delta$ . When  $\delta < \pi$ , **DIRECT MODE** is observed, and when  $\delta > \pi$  – **REVERSE MODE** is observed. In this way, the output power control is achieved by the variation of angle  $\delta$ .

Angle  $\varphi$  corresponds to the conduction time of the „input” stage freewheeling diodes, and angle  $\alpha$  – to the conduction time of the „output” stage transistors.

Angles  $\varphi$ ,  $\alpha$  and  $\delta$  are related to the operating frequency  $\omega_s$ .

### III. MODELING OF THE CONVERTER OPERATION

For the purposes of the analysis, the following assumptions are made: all the circuit elements are ideal, the transformer  $Tr$  has a transformation ratio  $k$ , the commutations are instantaneous, and the ripples of the voltages  $U_d$  and  $U_0$  are negligible.

The waveforms (Figs. 2 and 3) show that any of the half cycles can be divided into three intervals. For each of them, an equivalent DC voltage  $U_{EQ}$  is applied to the resonant tank circuit.

In accordance with the assumptions made, the resonant frequency, the characteristic impedance and the frequency detuning are:

$$\omega_0 = 1/\sqrt{LC}; \quad \rho_0 = \sqrt{L/C}; \quad v = \omega_s / \omega_0 \quad (1)$$

In order to obtain generalized results, all the quantities are normalized as follows: the voltages according to  $U_d$ ; and the currents – according to  $U_d/\rho_0$ . For each of the mentioned intervals, the normalized values of the current  $i_L$  through the inductor  $L$  and the voltage  $u_C$  across the capacitor  $C$  are determined as follows:

$$\begin{aligned} i'_{Lj}(\theta) &= I'_{Lj} \cos \theta - (U'_{Cj} - U'_{EQj}) \sin \theta \\ u'_{Cj}(\theta) &= I'_{Lj} \sin \theta + (U'_{Cj} - U'_{EQj}) \cos \theta + U'_{EQj} \end{aligned} \quad (2)$$

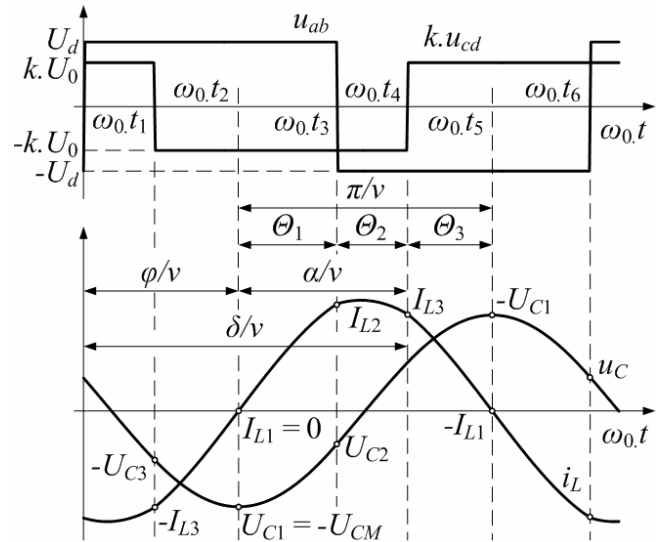


Fig. 3. Waveforms at **REVERSE MODE**

where  $j$  is the interval number;  $I'_{Lj}$  and  $U'_{Cj}$  are normalized values of the inductor current and the capacitor voltage at the beginning of the interval;  $\theta = 0 \div \Theta_j$ ;  $\Theta_j$  – angle of the interval for the resonant frequency  $\omega_0$ ;  $U'_{EQj}$  – normalized value of the voltage applied to the resonant tank circuit during the interval.

From Figs.2 and 3, it is observed that the value of  $i'_L$  at the end of given interval appears to be initial value for the following one. The same applies to the voltage  $u'_C$ . Therefore:

$$\begin{aligned} I'_{Lj+1} &= I'_{Lj} \cos \Theta_j - (U'_{Cj} - U'_{EQj}) \sin \Theta_j \\ U'_{Cj+1} &= I'_{Lj} \sin \Theta_j + (U'_{Cj} - U'_{EQj}) \cos \Theta_j + U'_{EQj} \end{aligned} \quad (3)$$

It is convenient the interval at which the initial values are  $I'_{L1} = 0$  and  $U'_{C1} = -U'_{CM}$  to be chosen as first. Then, for the second and the third intervals, the initial values of the current ( $I'_{L2}$  and  $I'_{L3}$ ) and the voltage ( $U'_{C2}$  and  $U'_{C3}$ ) are calculated on the base of Eqs. (3). For one half cycle the necessary for the purpose parameters  $\Theta_j$  and  $U'_{EQj}$  are pointed out in Table I.

TABLE I

MODE	Parameter	Number of interval		
		1	2	3
DIRECT	$\Theta_j$	$\frac{\delta - \varphi}{v}$	$\frac{\pi - \delta}{v}$	$\frac{\varphi}{v}$
	$U'_{EQj}$	$1 + kU'_0$	$1 - kU'_0$	$-1 - kU'_0$
REVERSE	$\Theta_j$	$\frac{\pi - \varphi}{v}$	$\frac{\delta - \pi}{v}$	$\frac{\pi - \delta + \varphi}{v}$
	$U'_{EQj}$	$1 + kU'_0$	$-1 + kU'_0$	$-1 - kU'_0$

In [3], analytical expressions for determination of the normalized values of the voltages  $U'_0$  and  $U'_{CM}$  are obtained. Like the initial values of the current  $i'_L$  and the voltage  $u'_C$ , they also represent function of the control parameter  $\delta$  and the angle  $\varphi$ :

$$U'_0 = \frac{1}{k} \frac{\sin\left(\frac{\pi - \varphi}{v}\right) - \sin\left(\frac{\varphi}{v}\right)}{\sin\left(\frac{\pi - \delta + \varphi}{v}\right) - \sin\left(\frac{\delta - \varphi}{v}\right)} \quad (4)$$

$$U'_{CM} = 2 \frac{\sin\left(\frac{\varphi}{v}\right) + U'_0 \sin\left(\frac{\pi - \delta + \varphi}{v}\right)}{\sin\left(\frac{\pi}{v}\right)} - (1 + kU'_0) \quad (5)$$

#### IV. LOAD AND CONTROL CHARACTERISTICS

For determination of the converter capabilities and for the purposes of its design, it is necessary the average values of the currents through the particular elements to be determined. For convenience, the following substitution is used:

$$I'_{AVj} = \frac{v}{2\pi} \int_0^{\Theta_j} i'_{Lj}(\theta) d\theta$$

$$= \frac{v}{2\pi} \left[ I'_{Lj} \sin \Theta_j - (U'_{Cj} - U'_{EQj}) (1 - \cos \Theta_j) \right] \quad (6)$$

The converter power devices conduct only within a single half cycle. The normalized average values of the currents through them for the **DIRECT MODE** are determined by expressions (7a) and (8a), and for the **REVERSE MODE** – by expressions (7b) and (8b):

$$I'_{QI} = I'_{AV1} + I'_{AV2}; \quad I'_{DI} = I'_{AV3} \quad (7a)$$

$$I'_{QI} = I'_{AV1}; \quad I'_{DI} = I'_{AV2} + I'_{AV3} \quad (7b)$$

$$I'_{QR} = kI'_{AV1}; \quad I'_{DR} = k(I'_{AV2} + I'_{AV3}) \quad (8a)$$

$$I'_{QR} = k(I'_{AV1} + I'_{AV2}); \quad I'_{DR} = kI'_{AV3} \quad (8b)$$

where  $I'_{QI}$  and  $I'_{DI}$  are the normalized average values of the currents through the transistors and the freewheeling diodes of the „input” stage;  $I'_{QR}$  and  $I'_{DR}$  – through the transistors and the freewheeling diodes of the „output” stage.

Independently from the operating mode, the normalized average values of the „input” current  $I'_d$  and the „output” current  $I'_0$  are determined as follows:

$$I'_d = I'_{QI} - I'_{DI} \quad (9)$$

$$I'_0 = I'_{DR} - I'_{QR} \quad (10)$$

By means of expressions (3) ÷ (10), for a fixed value of the control parameter  $\delta$  and with variation of the angle  $\varphi$ , values for the important converter quantities  $U'_0$ ,  $I'_{QI}$ ,  $I'_{DI}$ ,  $I'_{QR}$ ,  $I'_{DR}$  and  $I'_0$  can be calculated. On the base of these values, different load characteristics of the converter are easily built. For the purposes of the analytical examination, such characteristics

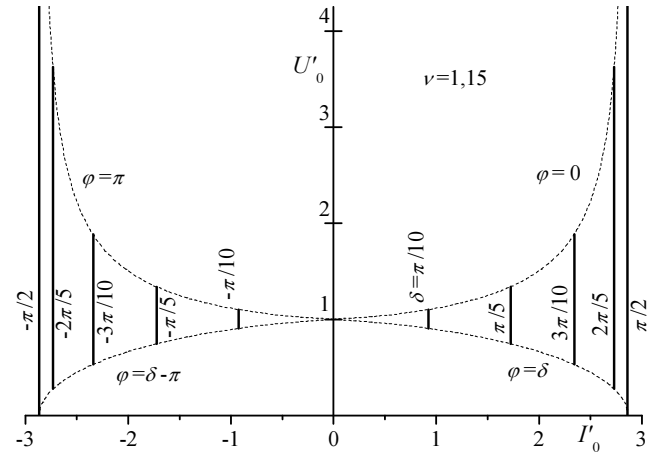


Fig. 4. Normalized output characteristics at  $-\pi/2 \leq \delta \leq \pi/2$

are obtained at frequency detuning  $v=1,15$  and transformation ratio  $k=1$ .

Fig. 4 presents the normalized output characteristics of the converter for values of the control parameter in the range  $-\pi/2 \leq \delta \leq \pi/2$ . The dependencies of the output voltage  $U'_0$  from the output current  $I'_0$  are shown with thick lines. With dotted lines are presented boundary curves determining the converter operating area at ZVS. It is observed that this area is strongly restricted. Even the no-load mode ( $I'_0 = 0$ ) is possible only with  $kU'_0 = U_d$ . Apparently, this range of variation of the control parameter  $\delta$  is not recommended for operation of the converter.

Fig. 5 presents the normalized output characteristics of the converter for values of the control parameter in the range  $\pi/2 \leq \delta \leq 3\pi/2$ . In this case, limitations for the converter operation at ZVS are not observed.

The output characteristics (Figs. 4 and 5) show that, independently from the range of variation of the control parameter, the output voltage does not depend on the output current, i.e. the examined converter behaves as an ideal current source. Moreover, the output voltage does not change its polarity and, independently from the energy transfer direction, can significantly exceed the input one.

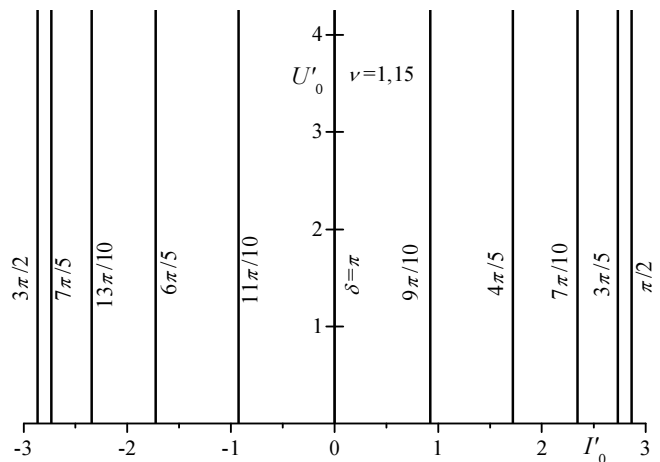


Fig. 5. Normalized output characteristics at  $\pi/2 \leq \delta \leq 3\pi/2$

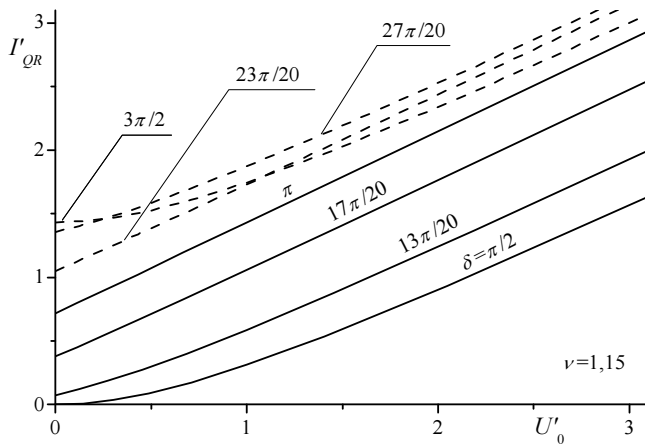


Fig. 6. Normalized average current through the „output” stage transistors –  $I'_{QR}$

Fig.6 presents the normalized dependencies of the average value of the current  $I'_{QR}$  through the „output” inverter stage transistors from the output voltage  $U'_0$ . Similar dependencies of the average value of the current  $I'_{DR}$  through the „output” inverter stage freewheeling diodes are shown in Fig. 7. The characteristics corresponding to the **DIRECT MODE** are presented with thick lines, and those to the **REVERSE MODE** – with dotted lines. They are obtained for variation of the control parameter in the range  $\pi/2 \leq \delta \leq 3\pi/2$ . For both modes, monotonous rise of the currents  $I'_{QR}$  and  $I'_{DR}$  is observed with the increase of  $U'_0$ . This shows that, independently from the energy transfer direction, the converter power devices stress is bigger for greater values of the output voltage.

Fig. 8 presents normalized control characteristics of the examined converter for several values of the frequency detuning ( $\nu = 1,08; 1,10; 1,15; 1,20; 1,30$ ). According to the fact that the output voltage does not depend on the output current, they are easily obtained with the substitution  $\phi = \delta/2$  ( $U'_0 = 0$ ). In the range  $\pi/2 \leq \delta \leq 3\pi/2$ , the dependencies variate monotonously, and a significant linear section can be observed.

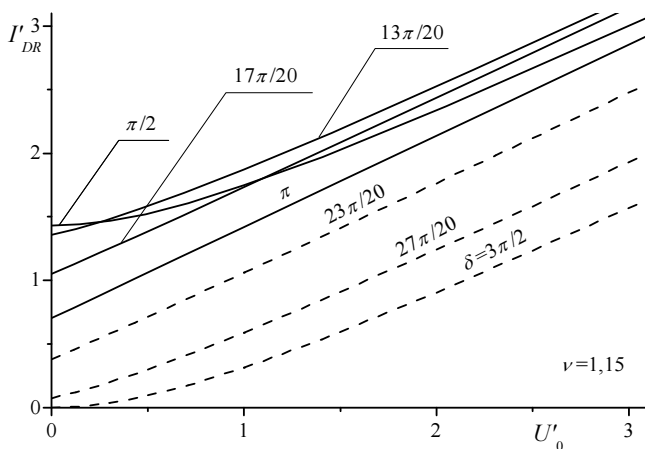


Fig. 7. Normalized average current through the „output” stage freewheeling diodes –  $I'_{DR}$

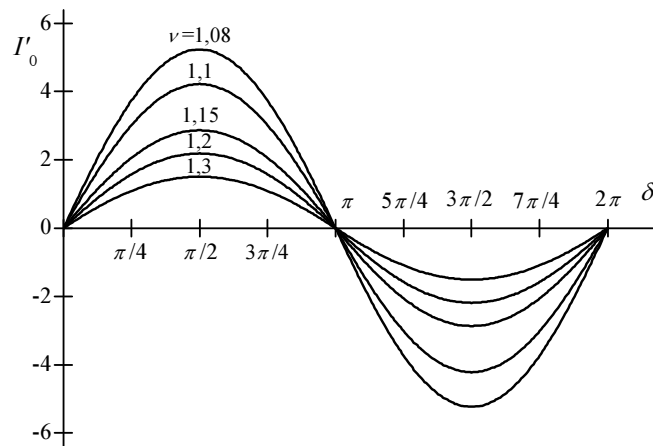


Fig. 8. Normalized control characteristics of the converter

## V. CONCLUSION

A bidirectional series resonant DC-DC converter operating above the resonant frequency is examined. A phase-shift control technique is used. As a consequence, its normalized load and control characteristics are built.

On the base of the obtained results, the following can be concluded:

- The considered bidirectional series resonant DC-DC converter can operate without violation of the ZVS conditions in wide range of variation of the control parameter. Furthermore, a significant linear section of the control characteristic is observed.
- With the chosen control technique, the converter behaves as a current source. Its output voltage can exceed the input one independently from the energy transfer direction.
- The converter power devices stress grows up with the output voltage increase.

The obtained results can be used for further examination and design of such converters.

## REFERENCES

- [1] Y. Cheron, H. Foch, J. Roux, "Power Transfer Control Methods in High Frequency Resonant Converters", PCI Proceedings, Munich, June 1986, pp. 92-103.
- [2] K. Alhaddad, Y. Cheron, H. Foch, V. Rajagopalan, "Static and Dynamic Analysis of a Series-Resonant Converter Operating above its Resonant Frequency", PCI Proceedings, Boston, October 1986, pp. 55-68
- [3] A.Vuchev, N. Bankov, Y. Madankov, A. Lichev, Analytical Modeling of a ZVS Bidirectional Series Resonant DC-DC Converter, LI International Scientific Conference on Information, Communication and Energy Systems and Technologies ICEST 2016, 28-30 June 2016, Ohrid, Macedonia.



# Two-Phase Switching-Mode Converter with Zero Voltage Switching Designed on CMOS 0.35 $\mu\text{m}$ Technology

Tihomir Brusev<sup>1</sup>, Georgi Kunov<sup>2</sup> and Elissaveta Gadjeva<sup>3</sup>

**Abstract** –A two-phase switching-mode dc-dc converter with Zero Voltage Switching (ZVS) is presented in this paper designed on CMOS 0.35  $\mu\text{m}$  technology with Cadence. The losses in the main power transistors are evaluated. The investigation results show that when ZVS technique is used, the efficiency of the whole system could be increased by about 3% compared to the standard two-phase dc-dc converter architecture.

**Keywords** – Two-phase switching-mode converter, Zero Voltage Switching (ZVS), Efficiency, CMOS technology, Cadence.

## I. INTRODUCTION

Today smart phones can transfer large data packages in a real time, thanks to the fourth generation Long-Term Evolution (4G LTE) wireless communication standard. The great functionality of modern telecommunication devices is due to the use of OFDM (Orthogonal Frequency-Division Multiplexing) modulation. The signal is transferred by several sub-carrier frequencies, which are summed at the output of modulator. Thus the spectrum in 4G LTE standard is used much more effectively [1].

On the other hand the sizes of the new mobile phones became larger, which allows increasing of the battery dimensions. Nevertheless, the time between two consecutive charges is small. The efforts of the designers are focused over the increasing of power supplies efficiency. Those circuits have to ensure the desire energy of transmitter's power amplifier (PA). The standard switching-mode dc-dc converter cannot fulfill the requirements to deliver appropriate fast dynamically changeable output voltage to drain or collector, respectively of MOS or BJT RF transistor of PA.

The power supply circuit, which provides the desire energy to transmitter's PA is called envelope amplifier. The block diagram of envelope tracking power amplifier's system is shown in Fig. 1. Envelope amplifier delivers the drain or collector supply voltage of PA's transistors [2]. It dynamically changes this supply voltage according to the variations of the PA input signal. The envelope amplifier tracks the PA input signal and controls the PA supply voltage according to the

envelope of this signal. The efficiency  $\eta_{ETPA}$  of the envelope tracking power amplifier's system is equal to [3]:

$$\eta_{ETPA} = \eta_{EA} \eta_{PA}, \quad (1)$$

where  $\eta_{EA}$  is the efficiency of the envelope amplifier;  $\eta_{PA}$  is respectively the efficiency of the PA. The run-time of wireless communication devices could be increased if envelope amplifier with high efficiency is designed.

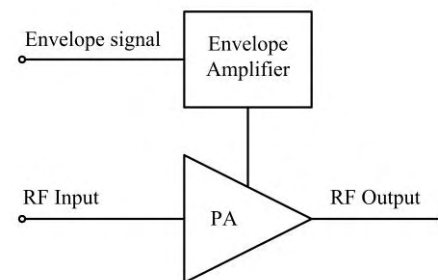


Fig. 1. Envelope tracking power amplifier's system

Most of the envelope amplifier's architectures include parallel or series connected linear amplifier and switching-mode dc-dc converter. Switching-mode converter ensures between 70% and 80% energy delivered to PA [2]. Therefore, high efficient dc-dc converter could improve significantly the efficiency of envelope amplifier. One of the features of switching-mode converters is that they are good choice only when relatively low-data rate communication signal is transferred. The tracking speed of the switching-mode converter could be increased if the single phase dc-dc converter is replaced by two-phase converter. Thus the portion of energy distributed from low efficient linear amplifier in envelope amplifier will be smaller, compared to the case when the switching-mode amplifier is a single phase dc-dc converter. This leads to improving of the overall efficiency of envelope tracking power amplifier system.

The tracking speed possibilities of standard synchronous single phase dc-dc converter and two-phase interleaved buck converter are discussed in Section II A of this paper. In Section II B, the Zero Voltage Switching (ZVS) technique is considered, which helps to reduce power losses in the main transistor of switching-mode dc-dc converters. Two-phase dc-dc converter with ZVS is designed with Cadence on CMOS 0.35  $\mu\text{m}$  process. Power losses in the main power MOS transistors are considered and evaluated. The efficiency results of two-phase dc-dc converter with ZVS are evaluated and compared to the efficiencies of the standard two-phase interleaved dc-dc converter. The received results are presented in Section III.

<sup>1</sup>Tihomir Brusev is with the Faculty of Telecommunications, Technical University of Sofia, Kl. Ohridski 8, 1797 Sofia, Bulgaria, E-mail: brusev@acad.tu-sofia.bg.

<sup>2</sup>Georgi Kunov is with the Faculty of Electronic Engineering and Technologies, Technical University of Sofia, Kl. Ohridski 8, 1797 Sofia, Bulgaria, E-mail: gkunov@tu-sofia.bg.

<sup>3</sup>Elissaveta Gadjeva is with the Faculty of Electronic Engineering and Technologies, Technical University of Sofia, Kl. Ohridski 8, 1797 Sofia, Bulgaria, E-mail: egadjeva@tu-sofia.bg.

## II. SWITCHING-MODE CONVERTERS

### A. Single phase and two-phase switching-mode converters

The standard synchronous single phase dc-dc converters, shown in Fig. 2, indicate high power losses in the inductor at large values of the inductor current ripple  $\Delta i_L$ . The two-phase converter structure helps to reduce this negative effect [4].

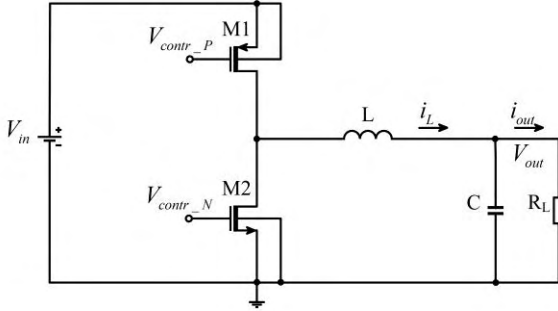


Fig. 2. Single phase dc-dc converter

The two-phase interleaved dc-dc converter architecture, presented in Fig. 3, leads to reduction of the output current ripple  $\Delta i_{out}$  of the circuit. The reason is that the phase sifted inductor current ripples respectively of the first and second sub-converter stage  $\Delta i_{L1}$  and  $\Delta i_{L2}$  are summed at the output.

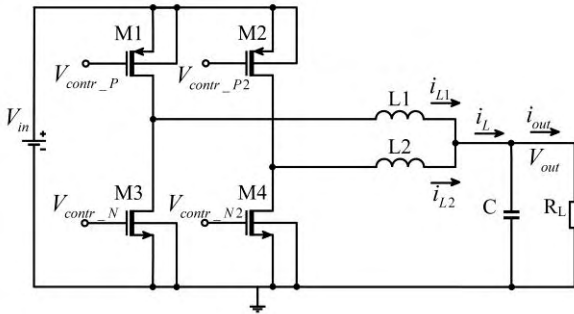


Fig. 3. Two-phase interleaved dc-dc converter

The output current ripple  $\Delta i_{out}$  of the two-phase interleaved buck converter with non-coupled inductors can be expressed in the form [5]:

$$\Delta i_{out} = \frac{V_{out}}{L} (1 - 2D) T_s, \quad (2)$$

where  $T_s$  is the switching period of converter,  $L$  is the value of filter inductors (if  $L1=L2$ , which is the case of the investigated dc-dc converter architecture).

Minimum values of the output current ripple  $\Delta i_{out}$  can be received if the duty cycle of the converter  $D$  is close to 0.5. The inductor current ripples  $\Delta i_L$  of the single phase buck dc-dc converter and two-phase interleaved buck converter with non-coupled inductors have equal values, and can be expressed in the form [5]:

$$\Delta i_L = \frac{V_{out}}{L} (1 - D) T_s. \quad (3)$$

In two-phase interleaved dc-dc converters architectures the same output current ripples as those of single-phase dc-dc converters could be established with smaller values of output filter inductors respectively  $L1=L2$ . These phenomena could be very useful for LTE applications power supply circuit, when envelope amplifier has to be fast in order to track high frequency envelope signal. The two-phase dc-dc converter could replace the single phase switching-mode regulator in parallel hybrid envelope amplifier structure. Thus most of the energy delivered to power amplifier can be ensured from fast and high efficient switching-mode multiphase dc-dc converter. The portion distributed from low efficient linear amplifier will be smaller, compared to the case when switching-mode amplifier is a single phase dc-dc converter, improving the overall efficiency of envelope tracking power amplifier system.

### B. Zero Voltage Switching

The circuit of synchronous single phase switching-mode buck dc-dc converter with ZVS is shown in Fig. 4. The advantage of those types of circuits is that the main power transistor  $M1$  can be switched-on and switched-off respectively at zero voltage [6].

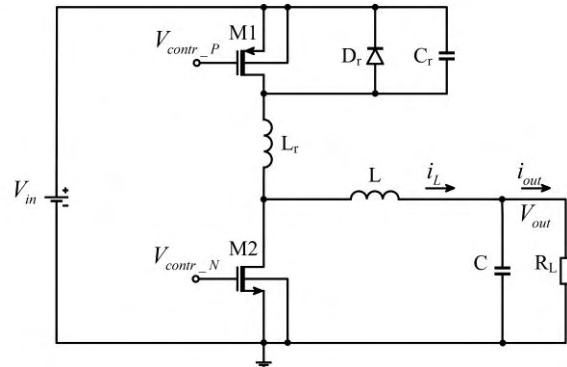


Fig. 4. Single phase dc-dc converter with ZVS

The zero voltage switch-off of the main PMOS transistor  $M1$  of dc-dc converter is because of the capacitor  $C_r$ . The Zero voltage switching-on state of the PMOS transistor  $M1$  is ensured by the diode  $D_r$ . The function of this component is to clamp to zero capacitor voltage  $V_c$ , when transistor  $M1$  is at switch-off state [6].

The effect of ZVS will lead to zero switching power losses of main PMOS transistor  $M1$ . Thus the total power losses in the MOS transistors of synchronous dc-dc converter could be decreased. They are equal to [7]:

$$P_{tot,MOS} = a \sqrt{\left( I^2 + \frac{\Delta i_L^2}{3} \right) f_s}, \quad (4)$$

where  $\Delta i_L$  is the inductor current ripple,  $I$  is a dc current

supplied to the load, and  $a$  is a coefficient depending on the equivalent series resistance of the transistors, the input total capacitance of the MOS transistors  $C_{tot}$ , and the power supply  $V_{in}$ . The decreasing of total power losses in MOS transistors of dc-dc converter will improve the efficiency of the circuit. The efficiency of the buck dc-dc converter can be expressed by:

$$\eta = \frac{P_{out}}{P_{out} + P_{losses}}, \quad (5)$$

where  $P_{out}$  is the output power of the dc-dc converter;  $P_{losses}$  are overall power losses in the dc-dc converter.

### III. INVESTIGATION OF TWO-PHASE SWITCHING-MODE CONVERTER WITH ZVS

The two-phase switching-mode converter with ZVS is designed with Cadence on AMS CMOS 0.35  $\mu\text{m}$  4-metal technology. The block circuit of whole buck dc-dc converter system is presented in Fig. 5. It consists of two power buck stages, error amplifier, ramp generator, comparator and buffer stage. In the designed two-phase switching-mode converter, Pulse-Width Modulation (PWM) control technique is used.

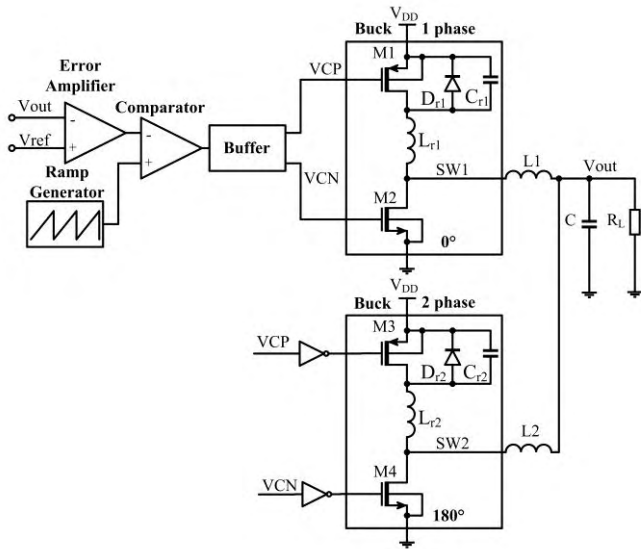


Fig. 5. Two-phase switching-mode converter with ZVS

The control signals, which regulate the main power MOS transistors of the both power buck stages, are 180° phase shifted. The supply voltage  $V_{DD}$  is equal to 3.6 V, which is a standard output voltage of lithium-ion battery. The average value of the output voltage  $V_{out(av)}$  of the converter is regulated to be equal to 1.2 V. The output filter inductors  $L1$  and  $L2$  of the both power buck stages are equal to 125 nH. The filter capacitor  $C$  is equal to 400 pF. The switching frequency  $f_s$  of the two-phase dc-dc converter with ZVS is equal to 76 MHz. The PMOS transistors  $M1$  and  $M3$  in both buck power stages are represented by 6 equal “modprf” transistors connected in parallel. Their sizes (W/L) are respectively 150/0.35  $\mu\text{m}$ . For NMOS transistors  $M2$  and  $M4$ , which replace the diode in the standard buck dc-dc converter circuit, 4 equal connected

in parallel “modnrf” transistors are used. Their sizes (W/L) are respectively 200/0.35  $\mu\text{m}$ . The values of resonant inductors  $L_{r1}$  and  $L_{r2}$  are equal to 10 nH. The values of resonant inductors  $C_{r1}$  and  $C_{r2}$  are equal to 50 pF. The waveform of output voltage  $V_{out}$  of the two-phase switching-mode converter with ZVS is shown in Fig. 6.

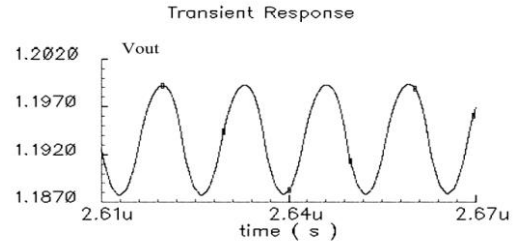


Fig. 6. The waveform of output voltage  $V_{out}$  of the two-phase switching-mode converter with ZVS

The waveforms respectively of  $I_{L1}$  and  $I_{L2}$  of two-phase switching-mode converter with ZVS are presented in Fig. 7.

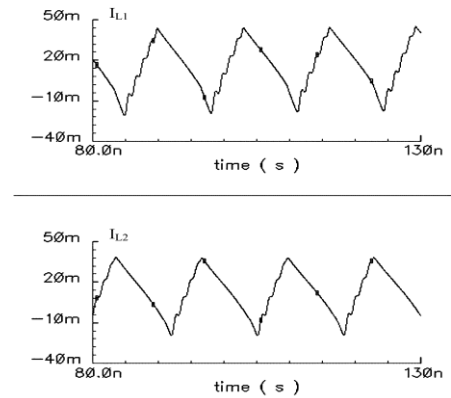


Fig. 7. Waveforms of the inductor currents  $I_{L1}$  and  $I_{L2}$  of two-phase switching-mode converter with ZVS

The waveforms of control signal  $VCP$ , which regulate the mode of operation of main PMOS transistors of the first buck stage, and the capacitor’s voltage  $V_{Cr}$  are shown in Fig. 8.

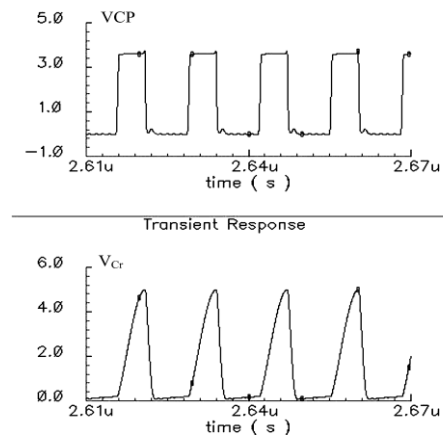


Fig. 8. The waveforms of control signal  $VCP$  and the capacitor’s voltage  $V_{Cr}$

The total power losses in main PMOS transistors ( $P_{PMOS}$ ) of two-phase dc-dc converter, designed on CMOS 0.35  $\mu\text{m}$  technology, are investigated as a function of the load  $R_L$ . The investigations are performed when the circuit works with and without ZVS. The efficiencies of two-phase switching-mode converter, respectively with and without ZVS are evaluated. The received results are presented in Table I.

TABLE I  
POWER LOSSES IN MAIN PMOS TRANSISTORS AND EFFICIENCY OF TWO-PHASE DC-DC CONVERTER AS A FUNCTION OF LOAD  $R_L$ , RESPECTIVELY WITH AND WITHOUT ZVS

	$R_L=10$ [ $\Omega$ ]	$R_L=20$ [ $\Omega$ ]	$R_L=25$ [ $\Omega$ ]	$R_L=30$ [ $\Omega$ ]
$P_{PMOS}$ [mW]	21.4	13.8	11.8	8.7
$P_{PMOS-ZVS}$ [mW]	19.4	11.4	9.6	6.3
Efficiency [%]	67.89	71.25	74.2	76.84
Efficiency-ZVS [%]	69.14	73.43	76.14	78.26

All the results shown in Table I are received when average value of the output voltage  $V_{out(av)}$  of converter is equal to 1.2 V.

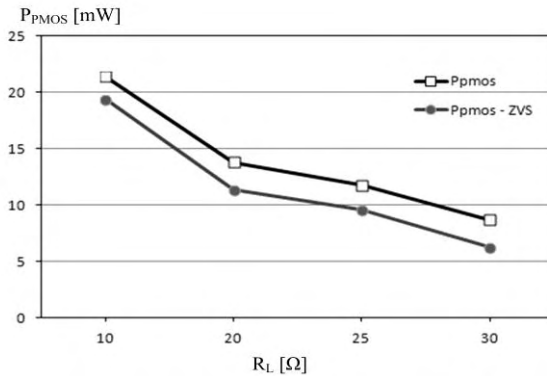


Fig. 9. Power losses  $P_{PMOS}$  of two-phase dc-dc converter as a function of load  $R_L$  with and without ZVS

The power losses of PMOS transistors of two-phase dc-dc converter are presented graphically in Fig. 9 as a function of the load  $R_L$  with and without ZVS.

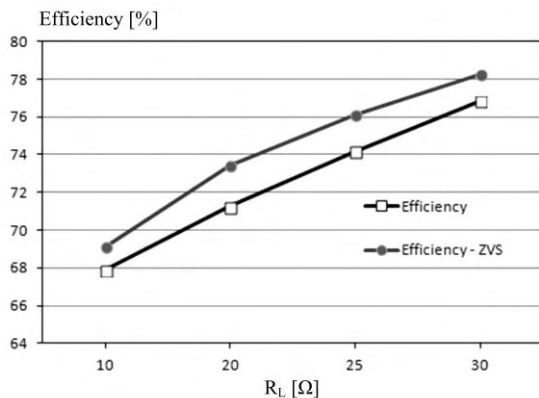


Fig. 10. Efficiency of two-phase dc-dc converter as a function of load  $R_L$ , respectively with and without ZVS

As it can be seen from the picture shown in Fig. 9, the total power losses in the main transistors of two-phase switching-mode ( $M1$  and  $M3$ ) are decreased by about 11% when ZVS technique is used. The reason is that switching power losses are minimized. This effect can be seen in Fig. 10, where the efficiency results of two-phase dc-dc converter as a function of the load  $R_L$  are graphically presented. As it can be seen from the picture efficiency of the whole converter system is increased by about 3% when ZVS is used. The values of the load resistance  $R_L$  are changed between 10  $\Omega$  and 30  $\Omega$ , because this range represents the practical equivalent value of PA used as a load [8].

## IV. CONCLUSION

In this paper two-phase buck dc-dc converter with ZVS designed on CMOS 0.35  $\mu\text{m}$  technology has been proposed. This circuit could be used as switching-mode regulator in parallel hybrid envelope amplifier for LTE applications. The investigation result shows that efficiency of two-phase dc-dc converter can be increased by about 3% if ZVS technique is used. The reason is that the total losses in the main switch of the converter's power stage are minimized.

## ACKNOWLEDGEMENT

The research described in this paper was carried out within the framework of UNIK – dog. DUNK – 01/03 – 12.2009.

## REFERENCES

- [1] M. Hassan, "Wideband high efficiency CMOS envelope amplifiers for 4G LTE handset envelope tracking RF power amplifiers", University of California, San Diego, 2012.
- [2] F. Wang, D. Kimball, D. Lie, P. Asbeck and L. Larson, "A Monolithic High-Efficiency 2.4-GHz 20-dBm SiGe BiCMOS Envelope-Tracking OFDM Power Amplifier", IEEE Journal of Solid-State Circuits, vol.42, no.6, pp.1271-1281, June 2007.
- [3] Y. Li, J. Lopez, D.Y.C. Lie, K. Chen, S. Wu, Tzu-Yi Yang and Gin-Kou Ma, "Circuits and System Design of RF Polar Transmitters Using Envelope-Tracking and SiGe Power Amplifiers for Mobile WiMAX", IEEE Transactions on Circuits and Systems I, vol.58, no.5, pp.893-901, May 2011.
- [4] G. Zhu, B. McDonald and K. Wang, "Modeling and analysis of coupled inductors in power converters", IEEE Transactions Power Electronics, vol. 26, no. 5, pp. 1355-1363, May 2011.
- [5] J. P. Lee, H. Cha, D. Shin, K. J. Lee, D. W. Yoo and J. Y. Yoo, "Analysis and Design of Coupled Inductors for Two-Phase Interleaved DC-DC Converters", Journal of Power Electronics, vol. 13, no. 3, May 2013, pp. 339-348, 2013.
- [6] N. Mohan, T. Undeland and W. Robbins, *Power Electronics, JWES, NY*, 2003.
- [7] V. Kurson, "Supply and Threshold Voltage Scaling Techniques in CMOS Circuits", University of Rochester, NY, 2004.
- [8] J. Ham, H. Jung, H. Kim, W. Lim, D. Heo and Y. Yang, "A CMOS Envelope Tracking Power Amplifier for LTE Mobile Applications", Journal of Semiconductor Technology and Science, vol.14, no.2, pp. 235-245, April 2014.



# Converters with Energy Dosing for Charging of EV's Li-ion Batteries

Nikolay Madzharov <sup>1</sup>, Goran Goranov <sup>2</sup>

**Abstract** - The scientific and the applied problems treated in the present paper are related to the development of a DC/DC power supplies with energy dosing for charging of electrical vehicles (EVs) batteries. They are a hybrid between the achievements in modern microelectronic components - frequency capabilities and low commutational losses, and the trends in the development of power conversion circuits which maintain the power or/and current constant and independent from the battery state operating of charge (SOC).

Some converters are presented together with their modifications. Their operation principle has been pointed out, as well as the investigations that have been carried out. Conclusions have been drawn about the possibility of obtaining good charging characteristics when the battery parameters change during the different charging scenarios.

**Keywords** – Converter, Energy dosing, Battery, Electrical vehicles, Battery state operating of charge.

## I. INTRODUCTION

The power electronic topologies for high power battery chargers can be largely classified into two categories: single phase types and two phase types [1,3,9,12]. Single phase battery chargers generally combine the power factor correction stage and the DC-DC conversion stage into one. They can be more efficient than the two phase types. However, single phase type battery chargers have a low frequency ripple in the output voltage and as a result, the switch and transformer ratings become larger [1,7,12]. The two phase types generally use a boost type converter to improve the power factor at the first stage and a DC/DC converter for the control of the voltage, current and power at the second stage. Two phase power converters use their own controllers to control the input current and the output voltage, respectively power, at the same time. They have a higher power factor and a lower harmonic distortion i.e., they have an advantage that there is almost no low frequency ripple in the output.

The other main characteristic of the charging power supplies is their universality regarding the battery parameters and good regulating possibilities. Not with standing the progress made, the methods used to regulate the output voltage of power sources, are not sufficiently smooth and envisage the use of relatively complex matching transformers, capacitors and ect.

With a view of solving the problems in this area, DC, converters with energy dosing (ED) have been synthesized analysed and tested [1,2,9,12]. According to ED method of operation, they generate, with a specified power, an output voltage corresponding to the particular parameters of the battery [1-3].

The main purpose of the present paper is also in this direction - presentation of circuits of power supply with ED for battery charging and the preliminary investigation performed on them.

A 32 kW battery charger is implemented to demonstrate the stability and performance of the system with ED. The validity of the concept is then verified through the constant current (CC) mode and the constant power (CP) mode charge of an actual Li-ion battery.

## II. DC CONVERTERS WITH ENERGY DOSING DEVELOPED FOR EV CHARGERS

The EV battery, such as electrical load, is characterized with strongly variable parameters during changing process - from idle running to short circuit. In table 1 are presented parameters of tested EV battery (Fig.1a) that contains more than a hundred pieces of AMP20m1HD-A single battery (Fig.1b), production of A123 Systems [9]. In connection with this, the charging convertor should have specific characteristics – on one hand to be able to limit and support output current and on the other hand to provide the dynamic feeding of the necessary power to the battery.



Fig. 1. Tested EV battery

A great number of circuits of DC/DC converters with ED are known [1,2,4-12]. Their distinctive feature is the presence of a dosing capacitor included in series in the battery loop through the interval of energy consuming by the mains. All of them provide dosing of the energy supplied to the battery, reliable work of loads changing and high commutation stability in the dynamic operating mode.

In table 2 are showed the basic circuit and respective time charts, illustrating the principle of converters with ED operating mode. It can be seen that the dosing capacitor voltage is fixed always to the value of the supplying DC

<sup>1</sup>Nikolay Madzharov is with the Faculty of EEE at Technical University of Gabrovo H.Dimitar str. 4, 5300 Gabrovo, Bulgaria  
e-mail:madjarov@tugab.bg, phone: +359 66 827557

<sup>2</sup> Goran Goranov is with the Faculty of EEE at Technical University of Gabrovo H.Dimitar str. 4, 5300 Gabrovo, Bulgaria  
e-mail:g\_goranov@tugab.bg, phone: +359 66 827531

voltage. Consequently, at constant work frequency the power given in the battery will always be one and the same. For the circuits with combined recharge of the dosing capacitor, in the expression for the power takes part the coefficient  $K$  which is less than 1 and depend on the load parameters.

TABLE I  
PARAMETERS OF EV BATTERY

Energy	18.7kWh (67.32MJ)				
Battery capacity	56.7 Ah				
Output voltage	310 ÷ 374V DC				
Charging current	60 ÷ 90A DC				
Internal resistance	0.10 ÷ 0.14Ω				
Battery voltage $U_{DCbat}, V$	310	330	350	374	
$I_{DCbat} = 60A$	$R_{BAT} [\Omega]$	5.17	5.50	5.83	6.23
$I_{DCbat} = 75A$		4.13	4.40	4.67	4.99
$I_{DCbat} = 90A$		3.44	3.67	3.89	NA

The DC converters with ED have the following advantages [1-4, 6, 7, 11, 12]:

- wide range concerning the load parameters and power;
- good regulation characteristics at satisfactory for the practice range and sensitivity;
- with the enlarging of the regulation depth the conditions for steady and optimum commutation of the transistors are kept, i.e. the switching is ZCS and/or ZVS;
- good electromagnetic compatibility with the mains.

TABLE II  
DC/DC CONVERTERS WITH ENERGY DOSING

Circuit	Method of ED	Intervals
1) DC/DC converters with ED and capacitor recharging by load current		
		$0 \div t_1$ $T_1$ $t_1 \div t_2$ $T_2$ $t_2 \div \pi$ $D$ $P = E^2 C_d f$
		$0 \div t_1$ $T_1, T_3$ $t_1 \div t_2$ $D$ $t_2 \div t_3$ $T_2, T_4$ $t_3 \div \pi$ $D$ $P = E^2 C_d f$
2) DC/DC converters with ED and combined recharge of the dosing capacitor		
		$0 \div t_1$ $T_1, T_3, D$ $t_1 \div t_2$ $T_1, T_3$ $t_2 \div t_3$ $D_1$ $t_3 \div t_4$ $T_2, T_4, D$ $t_4 \div t_5$ $T_2, T_4$ $t_5 \div \pi$ $D_1$ $P = k E^2 C_d f$
		$0 \div t_1$ $T_1, T_6$ $t_1 \div t_2$ $T_1, T_3, T_6$ $t_2 \div t_3$ $D, D_1$ $t_3 \div t_4$ $T_4, T_5$ $t_4 \div t_5$ $T_2, T_4, T_5$ $t_5 \div \pi$ $D, D_1$ $P = k E^2 C_d f$

Depending on the operating characteristics, the circuits of DC converters with ED can be systematized in the following groups:

- by the way of dosing capacitor recharging - combined or by the battery current;
- by the principle of output voltage regulation;
- by the presence of preparatory capacitor recharging;
- by the amount of the output impulses for one work frequency period – single-cycle and two-cycle circuits.

The comparative analysis of the results from the examinations and the obtained characteristics [2-4] give the possibility that the following recommendations for the usage of the shown in table 2 circuits can be formulated:

-at the lack of requirements to the battery current pulsations it is expedient to be used the circuits from group 1.

-at the necessity of supporting small output current pulsations, at a wide regulation range are used the circuits with combined recharging of the dosing capacitor. Theoretically, they do not have limits in the regulation characteristics;

-a DC/DC converter, combining ED and transforming, take a medial place in comparison to the examined circuits. To keep the small output current pulsations at a range of regulation  $K > 3$  it is necessary a great increasing of the commutation inductance value.

The dosing source in the shown DC/DC converters with ED is a capacitor, because of its great energy reserve and easy realization. It is also possible the use of inductance or inductance-capacitive circuit, tuned to the working frequency.

## II.I. NON SYMMETRICAL DC/DC CONVERTER WITH ENERGY DOSING

Figure 2 presents a non symmetrical DC/DC converter with ED.

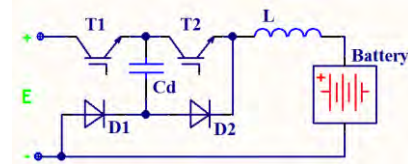


Fig.2. Non symmetrical DC/DC converter with ED

Control pulses are passed from the control system to  $T_1, T_2$ . In case of a battery load the circuit operating principle is as follows. When transistor  $T_1$  is forward - biased, current flows along the circuit (+)E,  $T_1$ ,  $C_d$ ,  $D_1$ ,  $L$ , battery, (-), E. When the dosing capacitor is charged up to voltage equal to E, the diode  $D_2$  is forward - biased and the current is closed along the circuit  $D_2, D_1, L$ , battery.

By the forward-biasing of  $T_2$  the capacitor  $C_d$ , charged up to the supply voltage E, begins discharging through  $T_2, L$ , battery,  $D_1, C_d$ . When the dosing capacitor is discharged the energy stored in the load is closed along circuit  $D_2, D_1, L$ , battery.

The expression

$$W = C_d E^2 / 2, \quad (1)$$

is valid for the energy which is stored and given up by means of the dosing capacitor. Taking into consideration the fact that the amplitude of the capacitor charge  $C_d$  for one half-period is equal to the supply voltage  $U_{cd} = E$ , the power given up in the load can be determined

$$P = C_d \cdot f \cdot E^2, \quad (2)$$

where  $C_d$  is the capacitance of the dosing capacitor,  $f$  - the operating frequency of the dosing device ( $T_1, T_2, C_d$ ),  $E$  - the supply voltage.

## II.II. SYMMETRICAL DC/DC CONVERTER WITH ENERGY DOSING

Figure 3 presents a symmetrical DC/DC converter with ED having a decoupled dosing capacitor.

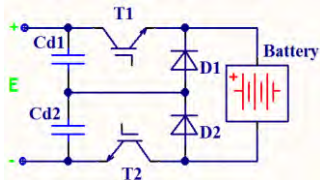


Fig. 3. Symmetrical DC/DC converter with energy dosing

The control pulses are passed from the control system to transistors  $T_1$  and  $T_2$ . In the first half-period  $T_1$  is forward-biased, and current flows along the circuit  $T_1$ , battery,  $D_2$ ,  $Cd_1$ , the capacitor  $Cd_1$  gets discharged and  $Cd_2$  charged to voltages 0 and  $E$ , respectively. The total sum of the dosing capacitors  $Cd_1$  and  $Cd_2$  voltages is always equal to the voltage of the supply source  $E$ . When the dosing diode  $D_1$  gets forward-biased the energy consumption from the supply source stops. The current flows along circuit  $D_1$ , battery,  $D_2$ . After  $T_2$  gets forward-biased the current flows along  $Cd_2$ ,  $D_1$ , battery,  $T_2$  and  $Cd_2$  gets discharged and  $Cd_1$  charged. After that the dosing diode gets forward-biased, and the current flows along circuit  $D_1$  - battery -  $D_2$ .

The output power in the load is determined from the expression

$$P = (C_{d1} + C_{d2}) \cdot f \cdot E^2, \quad (3)$$

where  $C_{d1}$  and  $C_{d2}$  - capacitance of the dosing capacitors,  $f$  - operating frequency of the dosing device,  $E$  - supply voltage.

The first conclusion that can be drawn from the expressions for  $W$  and  $P$  is that when the working frequency, the input voltage and the dosing capacitors value are unchanged, the power transmitted to the battery is constant and does not depend on the battery parameters. Supporting constant power means that the output converter voltage is matched with the battery voltage.

The second special feature of the converter is obtained by replacement the battery current expression  $I_{out} = U_{out} / Z_L$  in (2) and (3). After some transformations it is obtained a correlation giving the connection between the output and the input voltage.

$$U_{out} = 2E \sqrt{\frac{C_d \cdot Z_L}{2 \cdot T}} \quad (4)$$

The conclusion which can be drawn is that by changing the work frequency the output converter voltage, respectively battery voltage can be supported constantly when the battery parameters and/or the input voltage are changed. In figure 4 it is shown the dependence of the output voltage in function from the frequency at different loads (battery parameters). The information from these characteristics is used for converter designing, as it gives an

account of the connection between the load value and the capacitor  $C_d$  value, giving the power on one hand, and the dependence of the output voltage on the frequency and the input voltage, on the other hand.

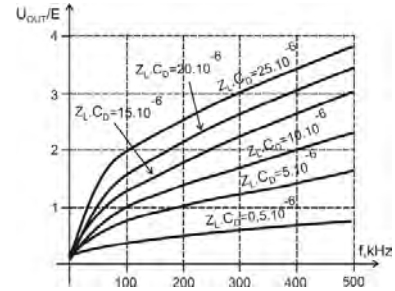


Fig 4. Regulation characteristics

The supporting and the regulation of the output voltage are realized by a feedback which changes the converter working frequency. The analytic dependence of the regulation law can be obtained as the expression for the output power is differentiated in relation to the time  $t$ . After some transformations

$$U_{out} I_{out} = 4C_d E^2 f, \quad (5)$$

$$U_{out} di_{out} + I_{out} du_{out} = 4C_d E^2 df \quad (6)$$

using the expressions

$$du_{out} = di_{out} (Z_L // 1/\omega C_F) \quad (7)$$

and

$$I_{out} = U_{out} / Z_L, \quad (8)$$

$$\text{is obtained } \frac{du_{out}}{df} = 2E \sqrt{\frac{C_d Z_L}{2f}} \frac{1}{1 + \omega C_F Z_L / 2}. \quad (9)$$

This expression is the operating function of the control system and gives the law, by which it should be modified the frequency of changing the load parameters with a purpose of supporting of constant output voltage. Using this function the control system can foresee, define, compensate and obtain the needed converter output characteristics.

## III. EXPERIMENTAL INVESTIGATIONS

The experimental research has been performed using non symmetrical and symmetrical DC converters with ED from figures 2 and 3, having power  $P=32$  kW and frequency  $f=20$  kHz. In fact they belong to the group of converters with ED and capacitor recharging by load current and have all characteristics, defined in paragraph II.

Figures 5 and 6 are presented both studied modes of battery charge at two values of the charging power, respectively current, till the charge level  $SOC = 90\%$ , when the process ends. The initial points in two characteristics starts from  $SOC=10\%$  and  $SOC = 50\%$  level of charge.

At the beginning of the process of charge (0-500 sec) the DC converter with ED maintain constant power and gradually increase in the current value from  $0.8I_{DC}$  to reach  $I_{DC}$ , 500sec after the beginning of the process. The battery voltage could be calculated by following relationship:

$$U_{BAT f} = U_{BAT 90\%} - I_{DC} \cdot R_{BAT} = [(0.9 \cdot (374 - 310)) + 310] - I_{DC} \cdot R_{BAT} \quad (10)$$

where  $U_{BAT f}$  - is battery voltage after end of charging scenario.

The parameter  $U_{BAT 90\%}$  is battery voltage at  $SOC = 90\%$ .



On reaching this value the control system sends a signal to the DC converter with ED for finish the process of charge.

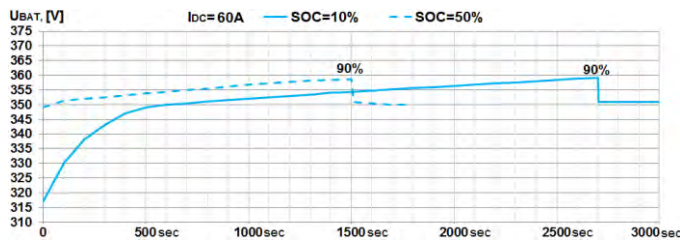


Fig.5. Charging scenario at charging power 21 kW ( $I_{DC} = 60A$ ) and SOC 10 and 50 %.

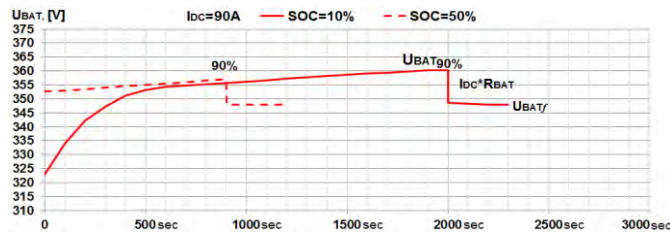


Fig. 6. Charging scenario at charging power 31 kW ( $I_{DC} = 90A$ ) and SOC 10 and 50 %

From the tests implemented can be concluded that the charging source works most efficiently while maintaining maximum charging current or/and output power. Depending on the type of battery used, there is a maximum value of SOC, in which the charging process can be carried out – typically SOC = 80 ÷ 90%.

The obtained from the circuit analysis expressions and tests results allow to be drawn charging characteristics at different batteries. They define the area in which the converter with ED can support constant output current or/and power.

#### IV. CONCLUSION

This paper described the design, control and tests of a DC converter with ED, which can be used to charge the Li-ion battery bank of EVs. A 32 kW converter was designed and implemented to verify the validity of the developed operating mode of ED and control algorithm. The obtained expressions for the control system operating function giving the law for operating frequency changing with a purpose to keep the output power constant when the battery parameters are changed. It has been verified that in ED mode algorithm the charging converter works most efficiently and maintaining maximum charging current or/and output power. It can be concluded that the developed converters with ED may contribute to higher system efficiency and a longer battery life due to its lower ripple current characteristics.

On the basis of the performed analysis and tests the following advantages of the discussed circuits can be pointed out:

- a) possibility for operation with almost constant power and/or current and different loads;
- b) operation in modes close to idle running and short circuit;

c) transistor commutation with zero current and zero voltage;

d) easy algorithm of transistors operation;

e) high power factor in relation to the mains.

The following disadvantages can be pointed out:

a) high maximum values of the currents through the transistors and the diodes;

b) a large number of active and passive elements for some of the circuits.

The obtained results and the drawn conclusions show that the proposed DC converters with ED can be used as charging power supply sources owing to the possibility of working with a wide range of battery parameters.

#### REFERENCES

- [1] Boulatov O.G., A.I. Tsarenko. Thyristor-Capacitor Converters, Energoizdat, 1982 (in Russian).
- [2] Todorov T. S., N. D. Madgarov, D. T. Alexiev, P.T. Ivanov. Autonomous Inverters., Gabrovo, 1996.
- [3] Swagath Chopra, „Contactless Power Transfer for Electric Vehicle" Charging Application“, Delft University of technology, July 2011
- [4] Danila E., Livint G., Lucache D.D., "Dynamic modelling of supercapacitor using artificial neural network technique", EPE2014-Proceedings of the 2014 International Conference and Exposition on Electrical and Power Engineering, Publisher IEEE, 2014, DOI: 10.1109/ICEPE.2014.6969988, p.642-645
- [5] Danila E., Sticea D., Livint G., Lucache D.D., "Hybrid backup power source behaviour in a microgrid", EPE 2014 - Proceedings of the 2014 International Conference and Exposition on Electrical and Power Engineering, Publisher IEEE, 2014, DOI: 10.1109/ICEPE.2014.6969987, Pages 637-641
- [6] Kraev G., N. Hinov, D. Arnaudov, N. Rangelov and N. Gradinarov, „Multiphase DC-DC Converter with Improved Characteristics for Charging Supercapacitors and Capacitors with Large Capacitance“, Annual Journal of Electronics, V6,B1,TU of Sofia, Faculty of EET, ISSN 1314-0078, pp.128-131, 2012
- [7] Bankov, N., Al. Vuchev, G. Terziyski. Operating modes of a series-parallel resonant DC/DC converter. – Annual Journal of Electronics, Sofia, 2009, Volume 3, Number 2, ISSN 1313-1842, pp.129-132
- [8] Bankov, N., Al. Vuchev, G. Terziyski. Operating modes of a series-parallel resonant DC/DC converter. – Annual Journal of Electronics, Sofia, 2009, Volume 3, Number 2, ISSN 1313-1842, pp.129-132
- [9] “Innovative fast inductive charging solution for electric vehicle” - Smart infrastructures and innovative services for electric vehicles in the urban grid and road environment, part of 7<sup>th</sup> Framework Program of EU, [www.fastincharge.eu](http://www.fastincharge.eu)
- [10] Nikolay D. Madzharov, Raycho T. Ilarionov, Anton T. Tonchev, “System for Dynamic Inductive Power Transfer”, Indian Journal of applied research, Vol. 4, Issue 7, 2014
- [11] Nikolay D. Madzharov – Anton T. Tonchev, “Inductive high power transfer technologies for electric vehicles”, Journal of ELECTRICAL ENGINEERING, Vol. 65, No. 2, pp. 125–128, 2014
- [12] Madzharov N.D, Ilarionov R.T, Battery Charging station for electromobiles with inverters with energy dosing, PCIM'11, Power Conversion, Nurnberg, Germany, 2011.



# Mathematical model of thermoelectric Peltier module

Liliya Staneva<sup>1</sup>, Ivaylo Belovski<sup>1</sup>, Anatoliy Alexandrov<sup>2</sup> and Pavlik Rahnev<sup>1</sup>

**Abstract** – Thermoelectric Peltier modules (TEM) are devices that convert electrical power in a temperature gradient. Usually in their catalogue data presents some transducer characteristics and maximum parameters, but they are insufficient to create highly efficient thermoelectric systems. The purpose of this article is to offer a relatively easy method for modelling of TEM, and the results are presented in tabular and graphic form by Matlab.

**Keywords** – Thermoelectric elements, Peltier cooler, Modelling

## I. INTRODUCTION

Thermoelectric Peltier modules (TEM) are devices which convert electrical power in temperature gradient. In their work they use the effect of Peltier, consisting of simultaneous heating and cooling of the two opposite sides of TEM [1, 2].

TEM have increasing interest due to simultaneous improving of their economic and technical parameters as well as and widely application which they get.

As a result of this the producer of TEM supplied to the market the wide assortment of modules with different thermoelectrically parameters, shape and sites [3].

Usually in the data sheets for given TEM some converting characteristics and maximum permissible parameters are shown: maximum temperature difference between the sides of the TEM –  $\Delta T_{max}$ , maximum current –  $I_{max}$ , maximum supplying voltage –  $U_{max}$ , and maximum absorbed from the cool side of TEM power –  $Q_{c_{max}}$  [4,5]

For creating of one high effective thermoelectrically system (TES), except these data it is necessary the optimal parameters of the real module to be known, as well as the base thermoelectrically parameters of the used for modules materials – coefficient of Zeebek  $\alpha$  [V/K], specific resistance of the materials  $-\rho$  [ $\Omega.cm$ ] and the coefficient of thermal conductivity  $k$  [W/cm.K].

Unfortunately the producer does not show that information in the data sheets and that is why it is necessary to have a method for calculation of these parameters.

<sup>1</sup>Liliya Staneva is with the Faculty of Technical science at University “Prof. D-r As. Zlatarov” of Burgas, 1 Prof. Yakimov. Blvd, Burgas 8000, Bulgaria, E-mail: anestieva@mail.bg.

<sup>1</sup>Ivaylo Belovski is with the Faculty of Technical science at University “Prof. D-r As. Zlatarov” of Burgas, 1 Prof. Yakimov. Blvd, Burgas 8000, Bulgaria, E-mail: ivbel@abv.bg

<sup>2</sup>Anatoliy Aleksandrov is with Faculty of Electrical Engineering and Electronics at Technical University of Gabrovo, 4 H. Dimitar Str, Gabrovo 5300, Bulgaria, E-mail: alex@tugab.bg

<sup>1</sup>Pavlik Rahnev is with the Faculty of Technical science at University “Prof. D-r As. Zlatarov” of Burgas, 1 Prof. Yakimov. Blvd, Burgas 8000, Bulgaria, E-mail: pavlikrahnev@abv.bg.

The goal of this paper is the easy and useful method for calculation of the thermoelectrically parameters of TEM.

These are  $\alpha_m$ ,  $\rho_m$ ,  $k_m$ , coefficient of conversion  $\eta$ , quality factor  $Z_0$  and the parameters of the materials  $\alpha$ ,  $\rho$  and  $k$ , based on the information from the producer for the limited parameters of TEM.

The results observed are presented in table and graphical mode with the help of graph editor MATLAB.

## II. MATHEMATICAL ANALYSIS

### A. Expression of base dependences for cooling TEM.

Next equation (1÷4) are fundamental and they are described in books and papers [6,7,8,9]:

$$Q_c = 2N \left[ \alpha I T_c - \frac{1}{2} I^2 \frac{\rho}{G} - k G \Delta T \right], \quad (1)$$

Where:

- $N$  – number of thermocouples in TEM;
- $G$  – factor of geometry expressing the relation between the surface and height of the semiconductor element;
- $I$  – electrical current.

The voltage  $U$  is given with:

$$U = 2N \left[ I \frac{\rho}{G} + \alpha \Delta T \right] \quad (2)$$

And the consummated power from TEM  $W$  is:

$$W = U \cdot I \quad (3)$$

Quality factor  $Z_0$  is the parameter, directly connected with the possibility of TEM to pump thermal power:

$$Z_0 = \frac{\alpha^2}{\rho \cdot k} \quad (4)$$

Definition of the parameters  $\alpha_m$ ,  $\rho_m$  and  $k_m$ :

$$\alpha_m = 2 \cdot \alpha \cdot N \quad (5)$$

$$\rho_m = \frac{2 \cdot \rho \cdot N}{G} \quad (6)$$

$$k_m = 2 \cdot N \cdot k \cdot G \quad (7)$$

Using equation (5-7), the equation (1, 2 and 4) can be presented as:

$$Q_c = \alpha_m IT_c - 0,5I^2 \rho_m - k_m \Delta T \quad (8)$$

$$U = \alpha_m \Delta T + I \rho_m \quad (9)$$

$$Z_0 = \frac{\alpha_m^2}{\rho_m \cdot k_m}, \quad (10)$$

#### B. Calculating thermal electrical parameters of TEM.

After reading the parameters from the producer data sheet:  $\Delta T_{max}$ ,  $I_{max}$ ,  $U_{max}$  u  $Q_{c_{max}}$ , the thermoelectrically parameters of TEM –  $Z_0$ ,  $\alpha_m$ ,  $\rho_m$  and  $k_m$  could be calculated.

This method uses three of the limit parameters -  $\Delta T_{max}$ ,  $I_{max}$  u  $U_{max}$ .

$$Z_0 = \frac{2\Delta T_{max}}{(T_h - \Delta T_{max})^2} \quad (11)$$

$$\alpha_m = \frac{U_{max}}{T_h} \quad (12)$$

$$k_m = \frac{(T_h - \Delta T_{max})U_{max}I_{max}}{2T_h \Delta T_{max}} \quad (13)$$

$$\rho_m = \frac{(T_h - \Delta T_{max})U_{max}}{T_h I_{max}}, \quad (14)$$

Where:

- $T_h$  is the temperature of the hot side of TEM.

After the calculation of thermoelectrically parameters of the module, this easy the thermos physic parameters of the semiconductors to be calculated, from which the thermocouples are created – coefficient of Zeebec  $\alpha$ , specific resistivity of the materials  $\rho$  and coefficient the thermal conductivity  $k$ .

It is done with the help of equation 5÷7, but only of the number of thermocouples  $N$  and geometry actor  $G$  which are known.

For calculation of the converting coefficient  $\eta$  and thermal resistivity of the hot radiator  $R_h$  next equation are used:

$$\eta = \frac{Q_c}{W} \quad (15)$$

$$R_h = \frac{T_h - T_a}{Q_c + W} \quad (16)$$

### III. DISCUSSION AND RESULTS

The algorithm which is used for calculation of thermoelectrically parameters is shown on Fig. 1.

After the initially definition of the conditions and checking for correct their import the next calculation are done:

- Physical characteristics of the chosen thermoelectric module – quality factor  $Z$ , ( $K^{-1}$ ); coefficient of Zeebec  $\alpha_m$  (V/K); the resistance of the module  $\rho_m$  ( $\Omega$ ) and coefficient of module resistivity  $k_m$  (W/K);
- The base physical characteristics of the used for TEM thermoelectrically elements: coefficient of Zeebec  $\alpha_m$  (V/K); specific resistance  $\rho$  ( $\Omega.cm$ ) and coefficient of thermal conductivity  $k$  (W/cm.K).

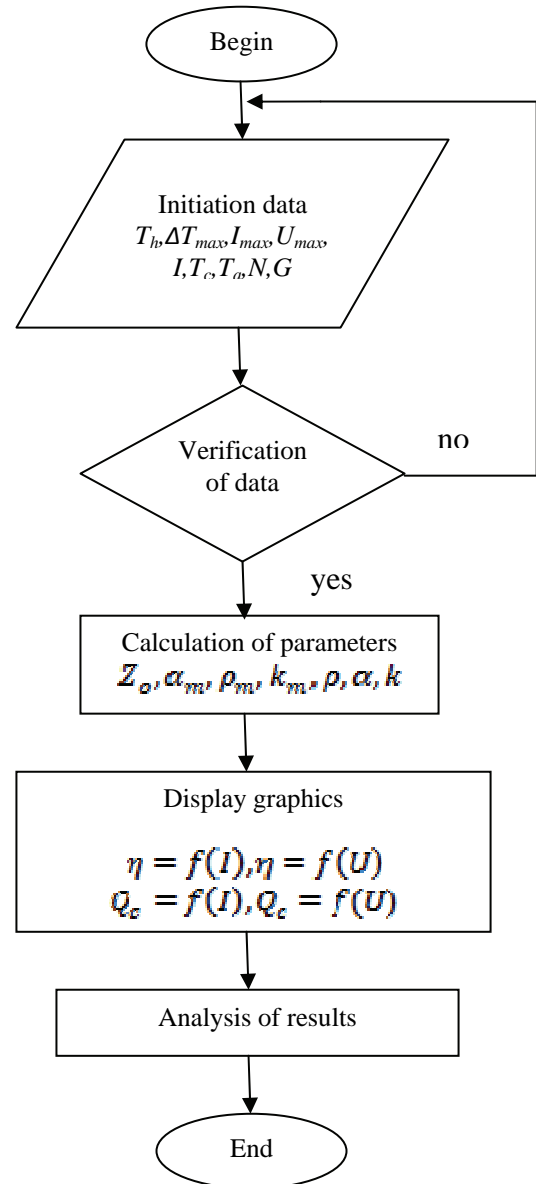


Fig. 1. The algorithm of modeling.

The following is the presentation of the results in the table and graphical mode and their analyses.

**Input the Performance Specifications:**

Th=300K

dTmax=60K

I<sub>max</sub>=10A

U<sub>max</sub>=12V

**Calculate the Physical Characteristics:**

Z=0.00208333 1/K

S<sub>m</sub>=0.04 V/K

R<sub>m</sub>=0.96 om

K<sub>m</sub>=0.8 W/K

**Input the Operating Conditions:**

I=0:10 A

T<sub>c</sub>=300:-20:240 K

T<sub>a</sub>=298 K

N=127

G=0.072 cm

**Calculate the Basic Physical Properties:**

p=0.000272126 om cm

s=0.00015748 V/K

k=0.0437445 W/(cm K)

Calc

Clear

Fig. 2. Input and calculated part of the program interface.

The visualization of the results from the modeling of the thermal electrical module is performed with the help of the program product realized on the base the graphical editor MATLAB.

The program gives wide possibilities for the user, who can import high number parameters: limited parameters of TEM, shown in the producer data catalogue -  $\Delta T_{max}$ ,  $I_{max}$ ,  $U_{max}$ , and well the working condition in which thermoelectrically module will be used.

They are:

- Input current  $I$  (in determined limits);
- Temperatures of the hot and cool side of the module –  $T_h$  и  $T_c$ . The hot  $T_h$  is firmly determined, but  $T_c$  is determined limits;
- The ambient temperature  $T_a$ ;
- The number of the semiconductors thermocouples  $N$ ;
- Geometry factor  $G$ .

Practically it can be simulated the work of every one arbitrary chosen TEM, if for all input data.

On Figure 2 the table the input/exit part of the working interface is shown.

The results, except in the table, are presented in graphical type.

The program proposes possibility four type of dependence to look at on and analyzed.

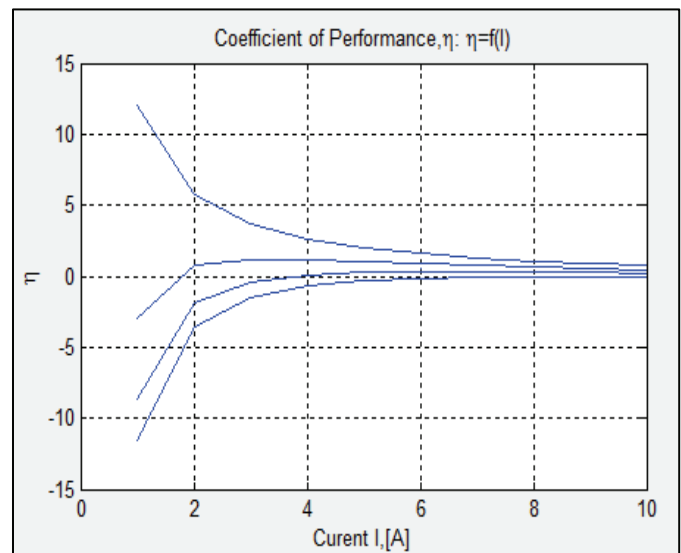


Fig. 3. Dependence of the performance coefficient  $\eta$  on the input current  $I$ :  $\eta = f(I)$ .

- Coefficient of performance  $\eta$  as a function of the input current  $I$ :  $\eta = f(I)$  – Fig.3;
- Coefficient of the performance  $\eta$  as a function of the voltage supply  $U$ :  $\eta = f(U)$  – Fig. 4;

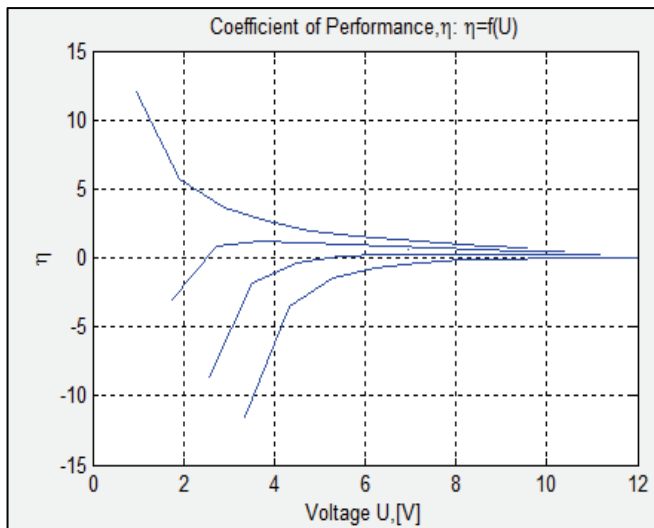


Fig. 4. Dependence of the performance coefficient  $\eta$  on the voltage supply  $U$ :  $\eta = f(U)$ .

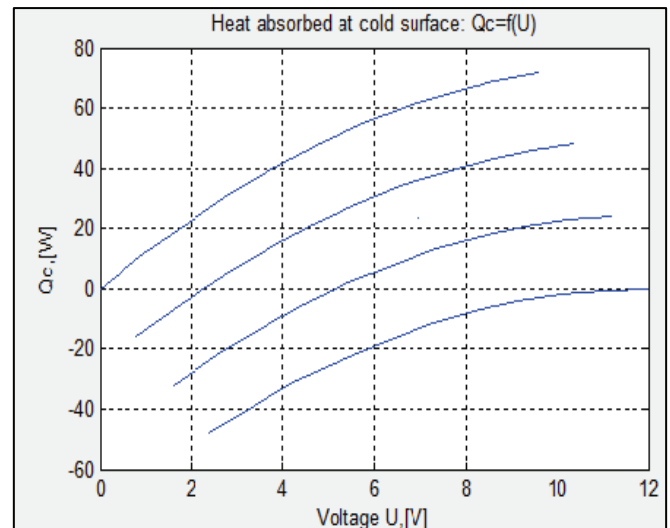


Fig. 6. Dependence of the absorbed thermal power  $Q_c$  on the voltage supply  $U$ :  $Q_c = f(U)$ .

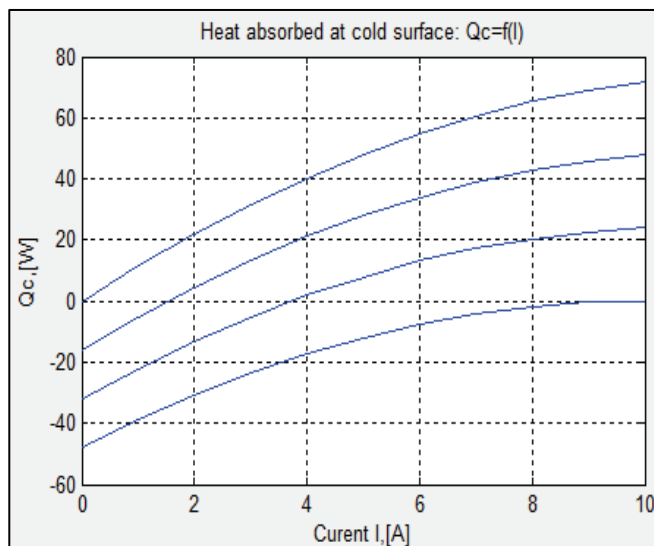


Fig. 5. Dependence of the absorbed thermal power  $Q_c$  on the input current.

- The absorbed thermal power  $Q_c$  from the cool side of TEM as a function of input current  $I$ :  $Q_c = f(I)$  – Fig. 5;
- The absorbed thermal power  $Q_c$  from the cool side of TEM as a function of voltage supply  $U$ :  $Q_c = f(U)$  – Fig. 6;

From Fig. 3 and Fig. 4 the efficiency of the given TEM can be estimated in dependence on DC mode of work.

The graphics from Fig. 5 and Fig. 6 show in which optimal values of input current and voltage the maximum values of the absorbed thermal power from the cool side of TEM could be reached.

#### IV. CONCLUSION

The realized mathematical model of thermoelectrically cooling module is a method with the user easily can calculate the base thermal physical parameters of TEM and for the semiconductors thermocouples, and in graphical way to report the absorbed thermal power  $Q_c$  in dependence on input current and voltage using a catalog information.

On the base of the received results it can select suitable cooling TEM in the design of thermoelectrically cooling – heating system.

#### REFERENCES

- [1] Smith, A.N, Norris, P.M.; “Heat Transfer Handbook”, John Wiley and Sons, 2002
- [2] Иоффе, А.Ф., “Энергетические основы термоэлектрических устройств из полупроводников”, АН СССР, 1950.
- [3] Riffat, S., Ma, X., “Thermoelectrics: A Review of Present and Potential Application,” Applied Thermal Engineering, Vol. 23, pp. 913-935, 2003.
- [4] <http://www.melcor.com/> - to MELCOR thermoelectric product manufacturer
- [5] [www.tetech.com](http://www.tetech.com) – Weblink to TETECH thermoelectric product manufacturer
- [6] Gierscheck, J. and Johnson, D., “Latest Developments in Thermoelectrically Enhanced Heat Sinks,” ElectronicsCooling, Vol. 11, No. 3, August 2005.
- [7] Rowe, D., “CRC Handbook of Thermoelectrics,” CRC press, Inc., 1995.
- [8] Gierscheck, J. and Johnson, D., “Latest Developments in Thermoelectrically Enhanced Heat Sinks,” ElectronicsCooling, Vol. 11, No. 3, August 2005.
- [9] Luo Z, A simple method to estimate the physical characteristics of a thermoelectric cooler from vendor datasheets, Electronics Cooling Magazine, 2008; 14:3: 22-27.



# Mathematical model of control system based on programmable environment

Anatolii Aleksandrov<sup>1</sup>, Goran Goranov<sup>2</sup>, Pavel Hubenov<sup>3</sup>

**Abstract** – The purpose of this article is to show consistency in implementation of the control system of a machine based on a programmable logic. It starts with a clarification of the requirements which determining its functionality and move on with calculation of a mathematical model of each block. The algorithm and the choice of the programmable environment for implementation are a topic of another article. Analysis of researched results.

**Keywords** – Logic gates, Control system, PWM, CPLD.

## I. INTRODUCTION

For reliable engineering sizing and implementation of a control system is extremely important to clarify the requirements which is based on. They are determined by industry, functionality, complexity, till the level of operators and lots of other factors but in conclusion – by the customer. In turn, requirements set the rules by which it will be designed and implemented - electric power modes, speed, management modes, to determine the nature and essence of each piece separately. The purpose of the designed machine is to drill holes and winding the thread in aluminum details.

Its principle of operation is as follows: the aluminum detail is placed in a magazine, loaded into the conveyor, transport towards a position, where the detail is preciously positioned, drilled by "Station1" for two holes, transported to the next position to the "Station2", wherein the cut thread, and exported by the conveyer. Then the cycle is repeated.

The operating accuracy of the machine depends on mostly of the accuracy of the angular displacements. There are few approaches for their realizations. The simplest block diagram is shown on Fig.1, a). An advantage of the method is easier to maintenance the system in time and at a lower cost. The disadvantage is the lack of feedback to control the speed by direct data delivery in case of faults situation. Tracking the speed of the shaft directly is impossible, and everything depends on the accuracy of microprocessor driver and the constant load. The only way to observe the statement of the motor is by consumed current. It is widely used in packaging machines with stretch film. [1, 2]

Second method includes feedback realized by three Hall

sensors, installed at 120 °, which allows the monitoring of the speed and direction of the motor shaft. The advantage of this configuration is the timely reaction of the driver in the event of faults situation. Also, the motor speed can be controlled by managing more precise and reaction on time, even if the load is dynamic, Fig. 1, b). It is used in the management of conveyors or drives with frequent changes of direction or speed. [3, 4]

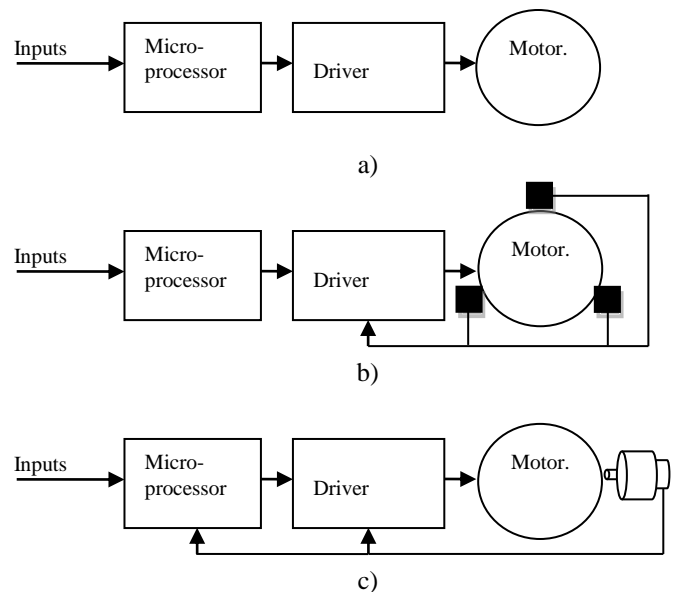


Fig. 1. Basic motion solutions

The most precise of the three configurations is shown in Fig. 1, c). The shaft is coupled with an encoder (photoelectric converter). Its function is to convert the angular displacement in a pulse sequence, i.e. pulses are corresponding to just one turn of 360 °. Basically pulses per round are multiple of 2<sup>n</sup>. There are exceptions, such as those with resolution 200, 400 or more pulses per turnover. For precious tracking the location of the shaft is necessary installation of a sensor for a specific referent point. In this case, the microprocessor looks for this signal and launches counting pulses from the converter. For achieving this goal can be used the embedded in the encoder reset pulse, which is only one for a turnover. The disadvantage of the approaches with sensor or reset pulse, the error that occurs due to the timing of engine braking and the subsequent vibration. When positioning is critical, processor permanently monitor this pulse sequence, ready to stop the cycle of operation of the machine and put the alarm. It can be avoided by using a delay circuit ensuring a delay. The time delay method allows the signal of the sensor to be delayed within a few pulses or time respectively few microseconds. [5]

<sup>1</sup>Anatolii Aleksandrov is Head of Department "Electronics" at Technical University of Gabrovo, 5300 Gabrovo, 4 Hadji Dimitar street, Bulgaria, E-mail: alex@tugab.bg.

<sup>2</sup>Goran Goranov is Assistant Professor – Doctor at Technical University of Gabrovo, 5300 Gabrovo, 4 Hadji Dimitar street, Bulgaria, E-mail: g\_goranov@bitex.bg.

<sup>3</sup>Pavel Hubenov is PhD student at Technical University of Gabrovo, 5300 Gabrovo, 4 Hadji Dimitar street, Bulgaria, E-mail: pavel\_hubenov@yahoo.com.

Even better positioning is achieved using absolute encoder. Differentiations from the previous one is that each position corresponds to a set of  $n$ -bit digital code. In this way positioning is sufficiently precise because every single position defines data which is unique. Although power supply interruption could occur, the position will be remembered and the cycle resumes without failures.

The provided data is usually in binary code 8-4-2-1 but higher noise protection can be achieved easily when output data is in Gray code. This method of control allows the realization of dynamic brake at which the processor supports the entire drive unit in a stable stationary. These are so-called servo systems. Widely applicable at positioning of robots, CNC class machines, confection cutting machines and extruders in polyethylene industry, and more. [6]

To ensure correct positioning of components, set system requirements determine the type of approach used and the specific drivers and their associated control units.

## II. MATHEMATICAL MODEL OF THE BASIC MODULES

When the requirements are completely fixed, has to design the plan for algorithm design. The more detailed it is, the more easily design of as small as possible modules regarding the functionality.

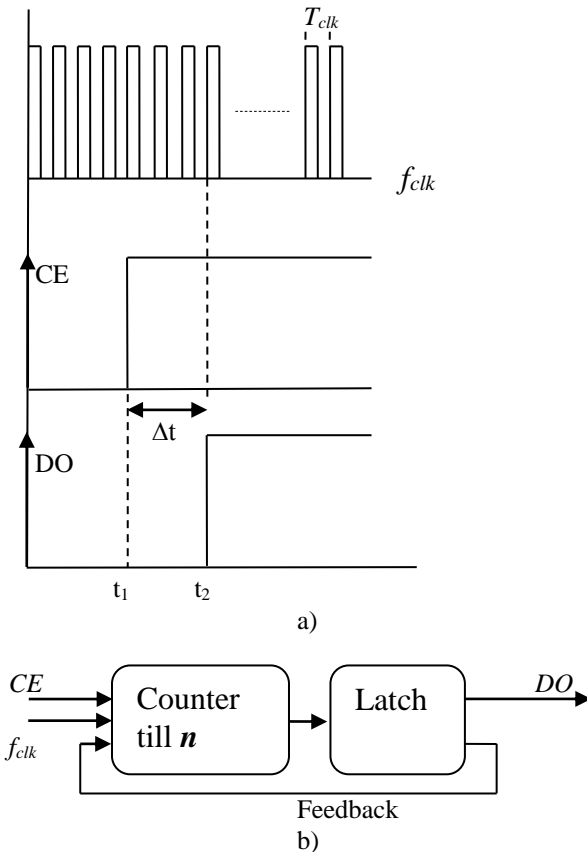


Fig. 2. Transition diagram and schematic of a delay line

Thus, by making easier for synthesis modules is achieved faster integration in a complex system, quick add or remove functionality and easier debugging. Using this approach, the

main types of circuit solutions are simplified to next four modules - Time Delay Line, combinational scheme with X-input and Y-output (Code Converter), Digitally Controlled Frequency Generator, Pulse Width Modulation Generator.

### 1. Time Delay Line

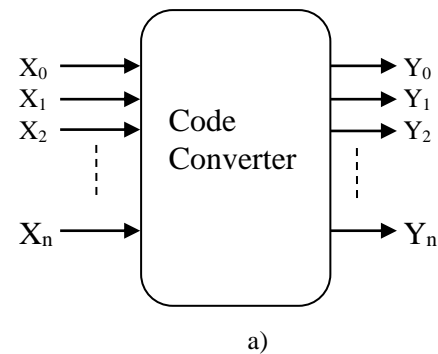
The block diagram of the delay line circuit is given in Fig.2, b). It provides accurate time delay. The difference between the received input signal and delayed output one is

$$\Delta t = t_2 - t_1 = n \cdot T_{clk} [\text{sec}], \quad (1)$$

where  $n$  is the number of periods which the counter has to count and then triggers the latch, fig.2, a). The latch itself ensures a stable and resistant statement and as per the feedback blocks the clock frequency pulses.

### 2. Code Converter

The second module Code Converter is shown in Fig. 3. Basically it is a combinational circuit which provides output binary data according to the input binary data. Applicable when switching relays, e.g. controlling a pneumatic distributor, a pump, a fan, etc., based on the input performance of various conditions. They could be driven by signal of active sensor, limit switch, or signal for completed cycle of a previous module within the algorithm. This kind of block has from one to lots of number of outputs depending on its application - provide simultaneously enable or disable output signal towards other modules.



$X_n$	...	$X_1$	$X_0$	$Y_n$	...	$Y_1$	$Y_0$
$[CS_n]$		$[SS_l]$	$[CE]$	$[t+1]$	$[t+1]$	$[t+1]$	$[t+1]$
0		0	0	0		0	1
		...				...	
1	1	1	1				

b)

Fig. 3. Schematic and diagram of Code Converter

The state transition diagram is shown in Fig. 3, b) where in the left side are filled input magnitudes and in the right side are output ones corresponding to the adjacent input sets. The right side is the statement in the next moment  $t + 1$ . Left located sets could be different conditions (for example:  $CS_n$  – Control

Signal  $n$ ,  $SS_1$  – Signal Sensor 1,  $CE$  – Control Enable), which will provide signal for execution. Mostly converters owned a large number of input signals - 1 till 10 and above, but could be used only strongly defined sets. In this case, the synthesis must take into consideration that such a function is incomplete and should be threat using an extra approach to avoid unacceptable switching. They are in relationship to the statement so called “not defined” which has to be include in the synthesis. When the outputs are more than one, every single one has to be synthesized by minor members into separate own table using Karnaugh or Veitch mapping output equation, which is converted to the preferred type of basic logic units – AND-OR, NOR, or NAND. Finally, subsequently schemes for different outputs are combined into a common combinatorial scheme. [7]

### 3. Digitally Controlled Frequency Generator

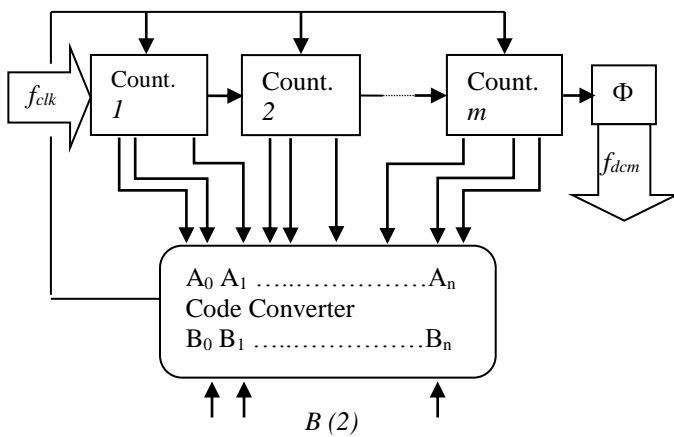


Fig.4. Block schematic of frequency generator

Block schematic of the generator is displayed in Fig.4. Generally, it is useful for frequency control of stepper motor's driver. The driver is allowed to control the motor by micro steps, i.e. a defined number of pulses is needed to be realized one complete revolution of the motor shaft. For example, the selected driver has a maximum number of 51200 steps per revolution, which means that a 360 degrees' rotation will be reproduced by 51,20kHz for just a second. Since this engine drives a conveyor and does not work in synchronization with another one, the requirements for it does not oblige the frequency to be exactly an integer than to be stable in time. This is the reason lead to the choice of this approach for implementing the control scheme. Otherwise could be use other method when a phase noise is observed, but the frequency is an integer and can be changed in small steps, for example 1.0 Hz.

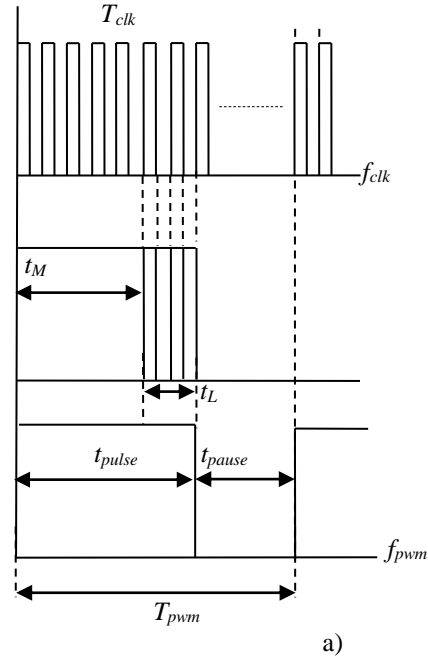
The output frequency is determined by the formula:

$$f_{dcm} = f_{clk} / B(10), \quad (2)$$

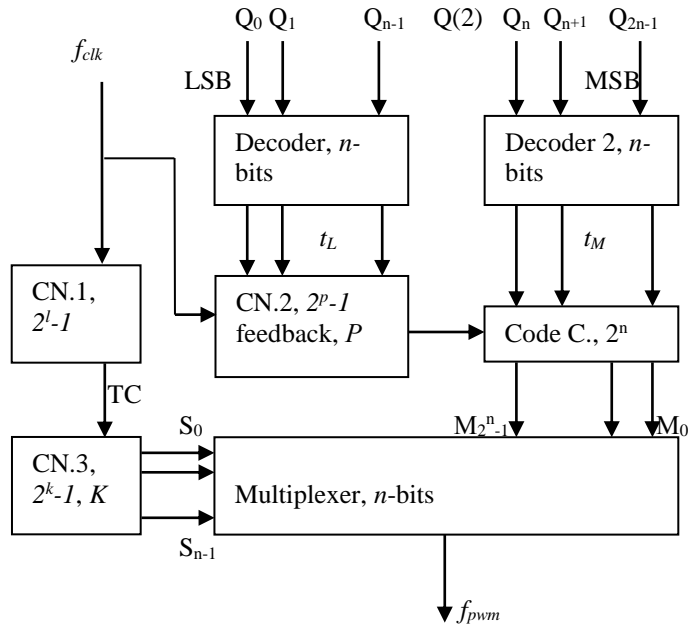
where number  $B$  is in binary-decimal system, but is entered in binary code to the inputs of Code Converter. [8]

### 4. Pulse Width Modulation Generator

The approach for pulse-width modulation is given in Fig. 5, b). As per the block diagram of the generator, the more complexity the more precision of the duty cycle is achieved and the most capacity of programmable platform is available. The principle of operation is as follows: multiplexer provides  $M_2^n$  number of channels, each of them is active for a defined time. It is determined by  $l$ -bit counter "CN.1", which counts the pulses of clock  $f_{clk}$ . This counter provides an output signal when overflow and switches counter "CN.3" which manages the address inputs of the multiplexer.



a)



b)

Fig.5. PWM generator

To realize the logical high level from the period  $T_{pwm}$ , Fig. 5. a) is used code converter "CC" which output code is code of

Johnson, providing a high level to the inputs of the multiplexer. It is managed by "Decoder 2" and "Code C" delivers to all junior multiplexer's channels, logical high level. For the current active channel of the decoder, "Code C" receives a signal from the other decoder. After counting the pulses from  $p$ -bit "CN.2" starts logical low level of the period  $T_{pwm}$ . That period of  $F_{pwm}$  is determined by the equation:

$$T_{pwm} = t_{pulse} + t_{pause} \text{ [sec]}, \quad (3)$$

where  $t_{pulse}$  is the time of the logical high level Fig. 5, a). This time is the sum of the other two:

$$t_{pulse} = t_M + t_L \text{ [sec]}, \quad (4)$$

as they are received from "Decoder 2" and "Decoder 1". Both decoders are managed separately by  $n$ -bit code, but connected as per their ranking by importance – starts with the most insignificant, LSB (*Low Significant Bit*) towards the most important, MSB (*Most Significant Bit*). The time specified by the decoder for the most significant bit (MSB) is:

$$t_M = (K-1) \cdot t(K) \text{ [sec]}, \quad (5)$$

where  $K$  is the number of the active channel of the multiplexer. Active time for each single channel is:

$$t(K) = 2^l \cdot T_{clk} \text{ [sec]}. \quad (6)$$

The time specified by the decoder for the least significant bit (LSB) is:

$$t_L = P \cdot T_{clk} \text{ [sec]}. \quad (7)$$

Here  $P$  is the number to which the external feedback of the counter "CN.2" is adjusted and then resets.

The step, or the quantum, by which the duty cycle of the pulse width modulated signal is determined is by the coefficients  $k$  and  $l$ :

$$Qu = 100 / (2^k \cdot 2^l) \text{ [ \% ]}. \quad (8)$$

Duty cycle  $\lambda$  is determined by:

$$\lambda = (t_{pulse} / T_{pwm}) \cdot 100 \text{ [ \% ]}. \quad (9)$$

The number  $Q$ , managing decoders and sets duty cycle, is the described as:

$$Q(10) = \lambda / Qu \text{ [ \% ]}, \quad (10)$$

then converted into binary and is provided to inputs  $Q_0 - Q_{2n-1}$ . The condition, which must be strongly obligated to be kept in the method of realization of the PWM signal is:

$$p \leq l. \quad (11)$$

Both mentioned generators in the above are designed so as to be controlled by binary parallel code. This approach allows changing the speed of the controlled motor to be performed from minimum to maximum within a single period of the clock frequency. It is useful whether the engine has a number of different speeds.

### III. CONCLUSION

The approach of the control system implemented in the drilling machine is recognized by using easy synthesis of all types of modules. It allows the system to be designed and implemented sufficiently simply, efficiently and reliably. The platform successfully responses to the requirements is CoolRunner-II, with mounted programmable logic matrix XC2C256-7TQG144C. Used matrix delivers flexible enough base for fine adjustments during the revival of the system and subsequently in its implementation as control unit in real facility.

Differenced from other types of architectures of programmable devices, where the number and functionality of the various blocks are factory pre-determined, everything here depends entirely on the designer and made synthesis. The methodology combines mathematical model and synthesis through minimization of logic functions, allows the expansion of the control system with more than one programmable matrix to remain just as easily and reliably. The implementation of functional logic elements makes this approach applicable in almost all the developments of programmable logic controllers series of leading manufacturers. The advantage is much wider audience, unlike using any other language for programming. To be continued with second article which describes algorithm and realizations of logic gates schematics.

### REFERENCES

- [1] Grasblum, Pavel, "Sensorless BLDC Motor Control Using MC9S08AW60", Freescale Semiconductor, Roznov pod Radhostem, 2007;
- [2] Prof. Dr. -Ing. J. M. Pacas, Prof. Dr. -Ing. R. M. Kennel, "Servo Drives State of the Art and Modern Developments", IECON 2005, Raleigh, North Carolina, USA, 6-10 Nov, 2005;
- [3] Bhargavi Nisarga, Daniel Torres, "Sensored 3-Phase BLDC Motor Control Using MSP430", Texas Instruments, Dallas, Texas, 2011;
- [4] Yaskawa America Inc., "AC servo drives, Product Catalog", 2011-2014;
- [5] Sigma II Servo System.;
- [6] Prof.asist., PhD Goranov, G.D., Eng. Kandov, I.O., „Digital schematics – manual for laboratory exercises“, „EXPRESS – Gabrovo, ISBN: 978-954-490-481-4, Gabrovo, 2015;
- [7] Maini, A.K., „Digital Electronics - Principles, Devices and Applications“, John Wiley & Sons Ltd, ISBN: 978-0-470-03214-5, Chichester, England, 2007;
- [8] Agarwal, A., Lang, J.H., "Foundations of Analog and Digital Electronic Circuits", Elsevier, ISBN: 1-55860-735-8, San Francisco, 2005;



# Electronic Module for 2D Positioner Manuel Control

Nikola Draganov<sup>1</sup> Totka Draganova<sup>2</sup> and Dragan Draganov<sup>3</sup>

**Abstract** – Electric drives are wide used in engineering practice. Precise operation and easy control of stepper electric motors had put their application in exact position fixing. The stepper motors in combination with control can provide a flexible operation and awards an opportunity to make reliable electric drives systems. Experimental model of electronic module for 2D positioner control based on two small powerful stepper electric motors and programmable logical matrix is presented in the report.

**Keywords** – magnetic fields control, stepper electric motors, programmable logical systems.

## I. INTRODUCTION

Stepper electric motors are often used in the practice as electric drives of different machines and equipments. Special units are used for this effect which are able to control the motor poles so that can carry out needed operations.

There are electromechanical units doing manipulations as position fixing of details, reading and writing heads which fine control is unimaginable without stepper electric motor. In some cases is not necessary to do this fixing automatic but manuel. So the operator directly assigns a direction, speed and a shifting distance.

Electric motors electronic control can be made by different manners. There are such units with discrete logic, microcontrollers and programmable logical matrixes. The last give a possibility for a flexible change of the discrete logical circuits and also to have a programming of their functions with some much possibilities – operating frequency, time constants etc.[1, 2, 3].

The aim of this elaboration is to propose, fulfill and investigate an experi-mental model of electronic module for control in two directions (2D) positioner by means of two small powerful stepper motors and programmable logical matrix. The developed unit can be applied in auto-mation, electronics and engineering practice for creating of different units for position fixing in 2D plain which are controlled by voltage level. The circuitry and experimental results of the unit which determines the direction and the rotation speed of electric motors in relation to the applied input voltage are presented.

<sup>1</sup>Nikola Draganov is with the Faculty of Electrical Engineering and Electronics at Technical University of Gabrovo, 4 Hadgi Dimitar Str., Gabrovo 5200, Bulgaria, E-mail: [ndrag@abv.bg](mailto:ndrag@abv.bg).

<sup>2</sup>Totka Draganova is with the Faculty of Electrical Engineering and Electronics at Technical University of Gabrovo, 4 Hadgi Dimitar Str., Gabrovo 5200, Bulgaria, E-mail: [totka\\_draganova@abv.bg](mailto:totka_draganova@abv.bg).

<sup>3</sup>Dragan Draganov is with the BDG Company, 20 Yanko Andonov Str., fl. 16, Burgas 8000, Bulgaria

## II. PRESENTATION

A block schematic diagram of the created experimental unit is depicted in Fig.1. It is intended for the position in 2D plain manuel control. Two small powerful stepper electric motors (SM) of the type KP39HM4-016 produced by Matsushita [4] are used. They control the motion of the metal frame in two directions. The block schematic diagram of the electronic unit is composed by three modules – set up generator-modulator (SGM), program-able logical device (PLD) and power device (PD). The control module executes double function. It assigns a motion direction of a corresponding motor (to X or to Y) and its rotation speed. SGM is created by two identical set up circuits (SC<sub>X</sub> or SC<sub>Y</sub>). They produce a signal to PLD which shows the rotation direction and a signal to controlled by voltage generator UFG. The rotation direction of the corresponding motor is determined by the sliding contact of the potentiometer position (control X or control Y). The resistance change of the potentiometer makes a control signal change to controlled by voltage block generator (UFG). It is a system of two channels generator for rectangular pulses. Its frequency is controlled by voltage given from corresponding set up circuit. So by SGM are made two independent pair pulses XCLK, XDIR and YCLK, YDIR with different (independent) frequencies. They give a clock signal and the rotation direction of both motors.

The LPD is the basic module. It is made on the base of a developmental system CoolRunner II, produced by Xilinx [5]. The platform consists of CPLD programmable logical matrix XC2C256 in case TQG144 and peri-pheral modules – JTAG programmer, clock generator, display, buttons, etc [5]. The programme is written down in a this block memory. It generates a defined sequence of pulses for motors correct management. After their for-ming the ruling signals come in power module (PM). Two independent each other signal pairs X<sub>12</sub> and Y<sub>12</sub> enter to a driver block (PD) input. They control the motion to X and Y. PD supplies with needed current and voltage motors coils SM\_X and SM\_Y.

The set up generator-modulator (SGM) schematic circuit diagram is depicted in Fig. 2. It consists of set up circuit SC<sub>X</sub> and controlled by voltage generator UFG. SC<sub>X</sub> defines a motor rotation direction. It is made by means of opera-tional amplifiers IC<sub>1a</sub>, IC<sub>1b</sub> and IC<sub>2a</sub>, IC<sub>2b</sub> and connected with them elec-tronic components – resistors R<sub>1</sub> ÷ R<sub>12</sub>, capacitors C<sub>1</sub> ÷ C<sub>3</sub>, C<sub>5</sub> and diode D<sub>2</sub>, D<sub>3</sub>. This circuit generates two signals. The first is used for voltage-fre-quency generator UFG control (elements IC<sub>2b</sub>, IC<sub>2c</sub>, IC<sub>2d</sub>, D<sub>1</sub>, R<sub>13</sub> ÷ R<sub>17</sub>). The second signal defines a motor rotation direction. The Y channel control is absolutely the same. The corresponding block generates signals for control of a speed and a direction (YCLK and YDIR).

This block circuit operation is defined by sliding contact voltage of a potentiometer  $R_p$  and its comparing with standard defined. The sliding contact place defines the rotation direction. In its middle position the voltage is equal to a half of supplying voltage. The amplifier  $IC_{2d}$  works as an integrator but  $IC_{2b}$  as a comparator.

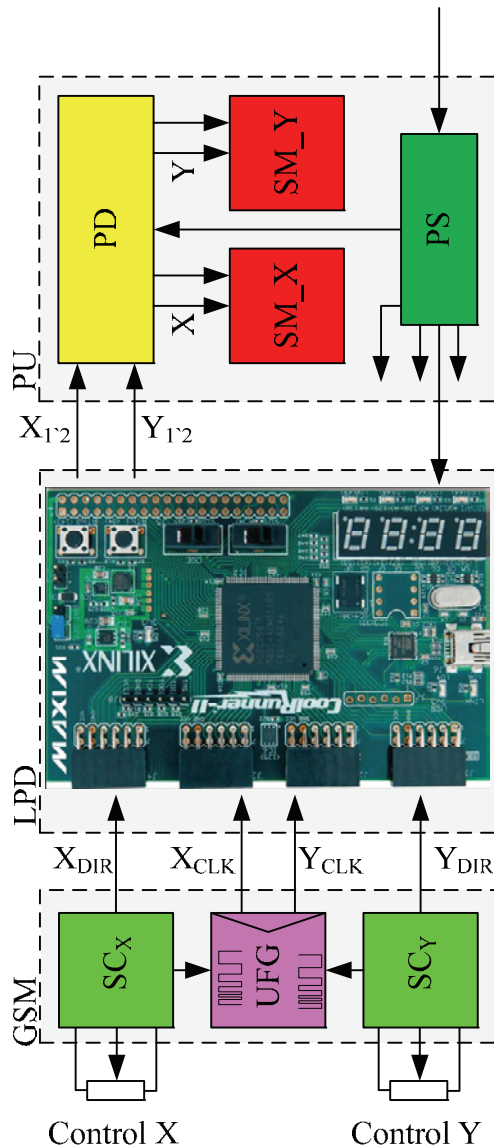


Fig. 1. Block structure of the electronic unit for 2D positioner control

They both build together a generator which is controlled by a voltage limited by IC<sub>2c</sub> and by a connected in its feedback D<sub>1</sub>. The voltage value at which pulse generations appear is dependent on the values of resistors R<sub>1</sub>=10kΩ, R<sub>2</sub>=50kΩ, R<sub>3</sub>=10kΩ. They are so selected and define the control voltage to be equal to a half of supply. When the potentiometer is moved to any direction from the middle position the comparator IC<sub>2a</sub> without delay sets up at its output a corresponding level (0 or 1) and the shifting value is determined by the voltage value at the output of IC<sub>1b</sub> and it

starts to generate pulse XCLK. The describing operation of module SCX timing diagrams are depicted in Fig. 2.

The received from UFG signals determine motor control speed to X and Y.

The signal levels XDIR and YDIR define their rotation direction. They are given to inputs of a programmable logical device PLD.

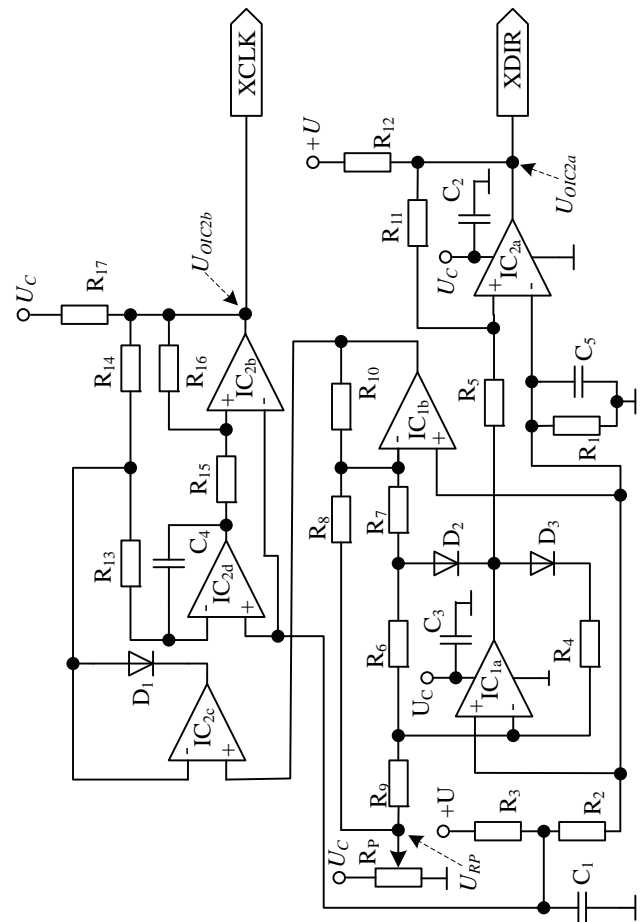


Fig. 2. Simplified schematic circuit diagram of an analog module for X channel control

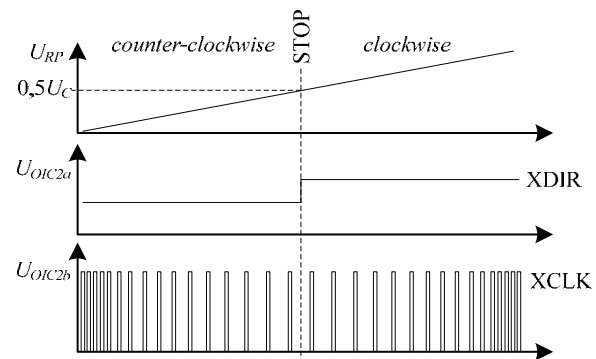


Fig. 3. Timing diagrams describing an analog module operation given a direction and speed rotation

TABLE I  
THE CONFIGURATION OF A PROGRAMMABLE LOGICAL MATRIX PINS

Type of Signal	Control by X direction	Control by Y direction	
	Name of Signal – pin number	Name of Signal – pin number	
INPUTS	XCLK – p124	YCLK – p39	CLOCK
	XDIR – p38	YDIR – p38	DIRECTION
OUTPUTS	CL1X – p10	CL1Y – p143	Control coil L1+
	CL2X – p7	CL2Y – p139	Control coil L1-
	CL3X – p5	CL3Y – p136	Control coil L2+
	CL4X – p3	CL4Y – p134	Control coil L2-

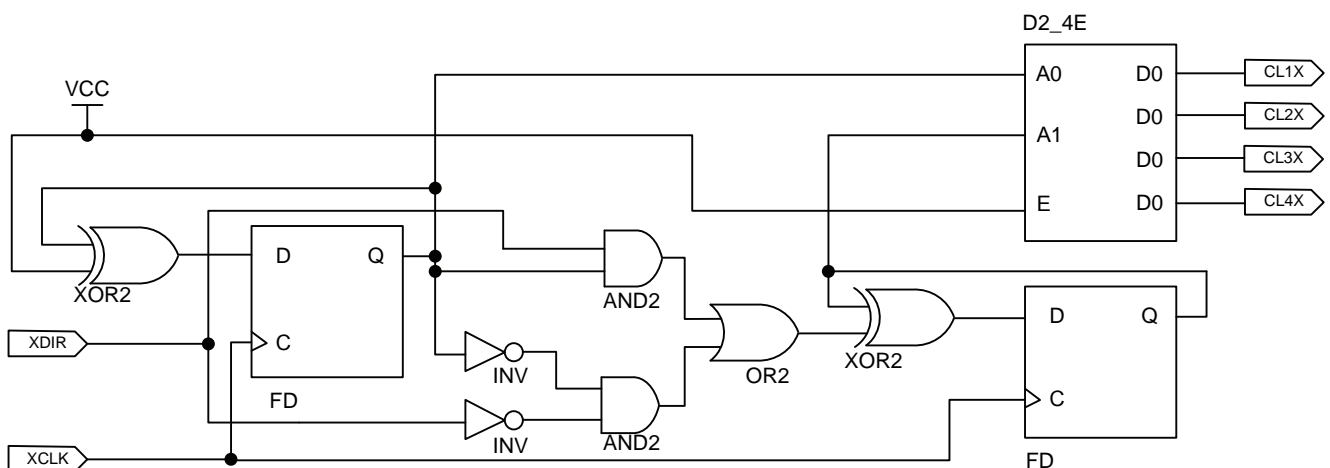


Fig. 4. Programmed in CPLD circuit for control to X only

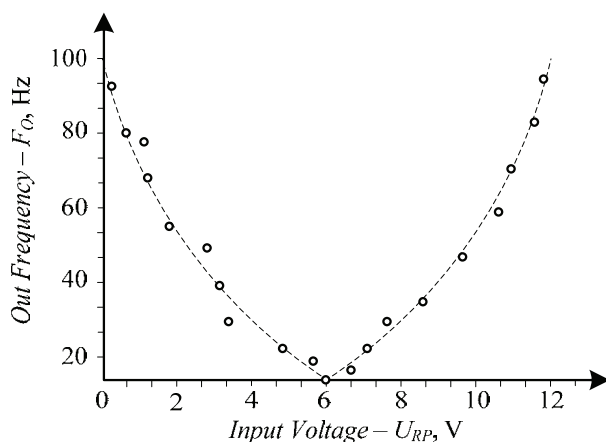


Fig. 5. Conversion characteristic of one channel in device UFG –  $F_o = f(U_{RP})$

The signals from both channels are treated by PLD. For unit realization are necessary both hardware and software. The logical matrix is programmed by a software ISE Design Suit

10.1 produced by Xilinx [5]. The CPLD programming can be made by two methods.

The first is by a programme code drafting. The other used here is by a desired circuit drawing with available in section schematic prepared libraries with function-al elements. A software conceptual electric circuit is depicted in Fig.4.

The formed in device UFG signals go to the programmable logical matrix inputs. They are configured so that the signal XCLK controls D-triggers FD (Fig. 4). A direction signal comes to logical scheme which controls the se-cond trigger so that a decimal counter has to sum up or to subtract. A modu-le PLD generates 4 signals (CL1X, CL2X, CL3X, CL4X) used for driver control of motors. The configuration of a programmable logical matrix pins in relation to input and output signals of both channels is depicted in table1 1.

The output signals for both control channels X and Y are absolutely inde-pendent. They come to driver block PD. The conceptual electric circuit of one channel in driver block is depicted in Fig. 6. The signals from CPLD are inverted by input buffers IC3. Each pair signals CL1X, CL2X and CL3X, CL4X controls MOSFET transistors (Fig. 6) in a H-bridge. One diagonal of both bridges is connected to voltage supply.

Another is connected to electric motor coil. The bridge circuit based on transistors VT1,VT2,VT3,VT4 controls a coil L1 while a bridge with VT5,VT6, VT7, VT8 controls a coil L2 (Fig.6 and table 1).

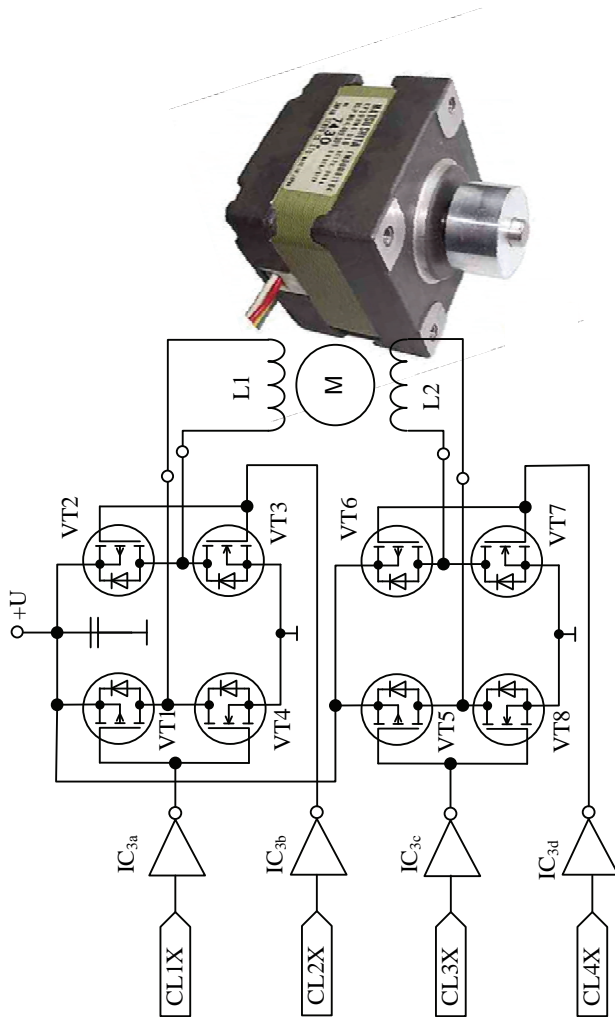


Fig. 6. Conceptual electric circuit of one channel in driver block

### III. CONCLUSION

An experimental unit for manual control of 2D positioner is exploited. It is made up a block circuit describing the operation of electronic unit. A conceptual electric circuit composed of three basic modules – set up generator, programmable logical device and power module is realized. A software for programmable logical device is created in programme medium ISE Design Suit 10.1 produced by Xilinx. The programming is made with build in developmental platform JTAG programmer. The realized unit is experimental investigated. The results are:

1. The timing diagrams in special check points having effect upon unit operation are obtained by means of electronic oscilloscope and are depicted in Fig. 4

2. A conversion characteristics of one channel of block UFG (Fig. 5) is obtained. It shows a dependence between a

generator output frequency  $F_O$  change and a voltage  $U_{RP}$  supplied to its output by means of potentiometer  $R_P$ . The analysis shows that it is symmetrical in relation to the middle value of the stabilized voltage supply ( $1/2 U_C = 6V$ ). The experimental characteristic shows that controlled by voltage generator exhibits nonlinearity.

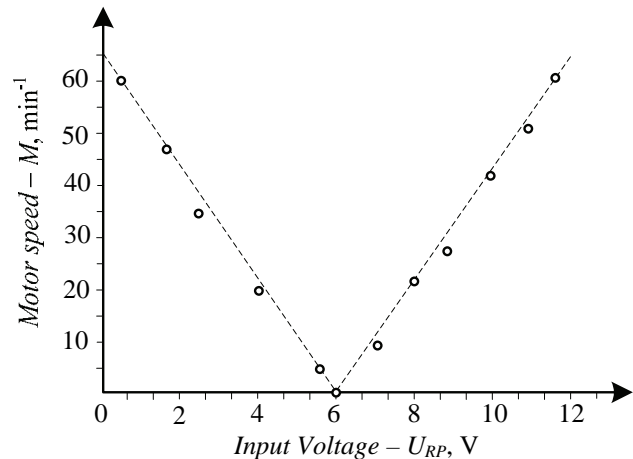


Fig.7 Experimental characteristic  $M = f(U_{RP})$

3. The dependence of electric motor M rotor rotation speed on input voltage coming from potentiometer RP is investigated. The characteristic  $M=f(U_{RP})$  is depicted in Fig. 7. It is symmetrical and linear. When a half of voltage is supplied the stepped motor stops. It is established the same change range of rotation speed from an input voltage  $U_{RP}$  in relation to initial value  $U_{RP} = 6V$ . This range is symmetrical to both rotation directions – from left to right at  $U_{RP} = 6 \div 12V$  and backwards from right to left at  $U_{RP} = 6 \div 0V$  the electric rotor rotation speed rises to  $M = 60 \text{ min}^{-1}$ .

The created unit can find a wide application in research laboratories and in engineering practice by the different positioners, machines and equipments, mechanical cutter, leth, metrology etc.

### REFERENCES

- [1] Draganov, N., T. Draganova. Galvanomagnetic Device for Angular Displacement Measurement. Proceedings of papers ICEST-11, Niš, Serbia, June 29 – July 1, 2011, ISBN 978-86-6125-033-0, pp.910-912
- [2] Draganov, N., T. Draganova. Based on CPLD Programmable Counter for Experimental Digital Electrical Energy Meter Part 2. Proceedings of papers ICEST-14, Niš, Serbia, 25-27 Juny, 2014, ISBN 978-86-6125-109-2, pp. 421-424
- [3] Draganov, N., SENSORS. Principles, Structure, Technology, Characteristics, Parameters and Applications. Part 1. Publishing House X-Press, Gabrovo, Bulgaria, 2014, ISBN 978-954-490-435-7.
- [4] Matsushita electric, Panasonic, [www.panasonic.com](http://www.panasonic.com), last visit April 2016
- [5] Xilinx, [www.xilinx.com](http://www.xilinx.com), CoolRunner II embedded platform, last visit April 2016



# Study of Nonlinear Effects in Parallel Gyrator Resonance Circuits

Atanas Tanev<sup>1</sup>, Ivan Uzunov<sup>2</sup>, Marin Hristov<sup>1</sup>, Sotir Ouzounov<sup>3</sup>

**Abstract** – This paper investigates the behaviour of bandpass filters, realized with parallel gyrator tank, at high input signal levels. The goal is to find in an empirical way correlation between the harmonic distortions of the gyrator amplifiers and the changes of the filter frequency response. When known, this dependency can be used during amplifier design to improve the performance of the filter and decrease its cost.

**Keywords** – nonlinear distortion, OTA-C filters, bandpass filters, gyrators, frequency response.

## I. INTRODUCTION

The optimization of the dynamic range is a significant problem in active filter design [1]. The problem in fact consists of two sub-problems, often solved simultaneously: maximizing the maximum permitted signal and minimizing the circuit noise. The maximum signal at the filter output is limited by the available voltage headroom and the acceptable level of distortions of the output signal. The methods for dynamic range optimization usually aim to equalize the output voltages of the amplifiers in the filter. However, the existing methods typically do not consider how the filter frequency response changes depending on of the amplifier nonlinearities.

The changes of the filter frequency response due to amplifiers' non-linearity are studied since a long time. The focus has been on second-order bandpass filters due to their simplicity, considering basically gyrator equivalent of the LC tank [2,3] and a single operational amplifier biquad [4,5]. Their frequency response inclines at large input amplitudes (Fig. 1) due to nonlinearities in the active elements, which leads to three main effects: 1) The frequency response is no longer symmetric; 2) Its maximum moves usually to lower frequencies; 3) Appearance of the so called "jump phenomenon" at certain nonlinearity levels. The frequency response changes with a jump at a certain frequency and the suggestion is that it is due to appearance of area, in which the response has two values (dotted line in Fig. 1).

Different techniques are applied for analysing the influence of the nonlinearities. In [2,3] analytical expressions are derived that link the level of non-linearity to the filter parameters in gyrator filters. In these studies, the operational trans-

conductance amplifiers (OTA), used in the gyrator, are represented by equivalent large-signal value of the transconductances  $g_m$  where the  $g_m$  is impacted by the amplifier's nonlinearity. More complicated methods have been proposed for achieving better accuracy and analysis the nonlinear effects in higher order filters. For example, the harmonic balance method is applied in [4,5], while in [6,7] Volterra series analysis is used. State-space analysis of the harmonic distortions in OTA-C filters is considered in [8].

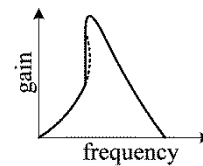


Fig. 1. Inclining the frequency response and "jump phenomenon" in second-order band pass active filters due to amplifier nonlinearities.

The review above shows that most of the existing methods for theoretical analysis of the nonlinear effects in the active filters are approximative and in most of the cases do not rely on powerful tools for simulation of electronic circuits like Cadence, PSpice, etc. Also, there are some questions that remain unclear. For example, the term "weak nonlinearities" is used without precise definition. Further, the relationship between the total harmonic distortion (THD) of the amplifiers and the variation of the filter frequency response it is not well clarified.

The goal of this paper is to study empirically these questions for one of the most popular bandpass filter – the gyrator parallel resonance circuit (gyrator tank). This is done by computer simulation of filter with different types of OTAs with different nonlinearity. The results demonstrate various effects in the frequency response and allow to estimate the limits of OTA's THD in conjunction with the tolerable changes in the filter frequency response.

## II. GYRATOR PARALLEL RESONANCE CIRCUIT AND MODELLING OF OTA V-I CHARACTERISTIC

The circuit of the bandpass filter, realized as parallel gyrator resonance circuit, is shown in Fig. 2(a). The gyrator consists of  $g_{m1}$  and  $g_{m2}$  and forms the resonance circuit together with the capacitors  $C_1$  and  $C_2$ . OTA  $g_{m0}$  at the input converts the input voltage to an equivalent current for proper operation of the tank and  $R$  is the load resistance. The general form of the small-signal filter transfer function is

$$H(s) = \frac{V_o}{V_i} = \frac{hs\omega_0/Q}{s^2 + s\omega_0/Q + \omega_0^2}. \quad (1)$$

Its angular resonance frequency  $\omega_0$ ,  $Q$ -factor and gain  $h$  are:

<sup>1</sup> Atanas Tanev and Marin Hristov are with the Department of Microelectronics, Faculty of Electronic Engineering and Technologies, Techn. University - Sofia, 8 Kl. Ohridski blvd., 1000 Sofia, Bulgaria, e-mails: atanev@ecad.tu-sofia.bg; mhrisov@ecad.tu-sofia.bg

<sup>2</sup> Ivan Uzunov is with Smartcom Bulgaria AD, BIC IZOT, 133, Tzarigradsko Shosse blvd., 1784 Sofia, Bulgaria, e-mail: ivan\_uzunov@smartcom.bg.

<sup>3</sup> Sotir Ouzounov is with Philips Research Europe, High Tech Campus, 5656 AE Eindhoven, Netherlands, e-mail: sotir.ouzounov@philips.com.

$$\omega_0 = \sqrt{\frac{g_{m1}g_{m2}}{C_1C_2}}; \quad Q = R\sqrt{\frac{g_{m1}g_{m2}C_1}{C_2}}; \quad h = g_{m0}R. \quad (2)$$

The corresponding frequency response is symmetrical around  $\omega_0$  when logarithmic scale is used for the frequency axes. Examples for two different  $Q$ -values are shown in Fig. 2(b).

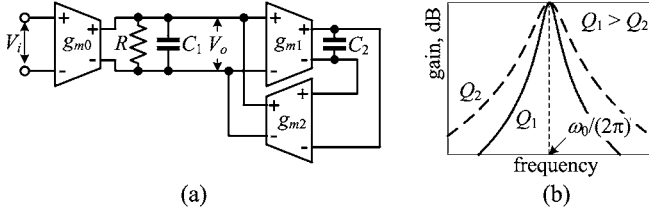


Fig. 2. (a) Circuit of bandpass filter realized as parallel gyrator resonance circuit; (b) filter frequency response at two different pole  $Q$ -factors.

The investigation is focused on the impact of the nonlinearities in  $g_{m1}$  and  $g_{m2}$  on the filter frequency response. The harmonics produced in the input OTA  $g_{m0}$  typically are filtered by the gyrator tank. They may have visible effect only when the input signal is too large and the fundamental frequency of the signal is 2 or 3 times lower than  $f_0 = \omega_0/(2\pi)$ . For this reason this amplifier is considered as an ideal voltage controlled current source (VCCS). Its transconductance is equal to  $1/R$  in all simulations in order to have unity gain at the filter central frequency  $f_0$ .

The other two OTAs  $g_{m1}$  and  $g_{m2}$  are considered as nonlinear VCCS and their transfer characteristics  $i_o(v_i)$  ( $i_o$  – OTA output current,  $v_i$  – OTA input voltage) are approximated by polynomials. Two different circuits are taken as models: the simple differential pair with dynamic load in Fig. 3(a), and a differential amplifier with linearization (Fig. 3(b)). The linearization of the second amplifier is achieved by source degeneration resistors – transistors  $M_3$ – $M_6$  are used for this purpose. This amplifier also uses dynamic load. A negative resistance emulated by the pair  $M_9$ – $M_{10}$  is connected in parallel to the output for increasing the output impedance. The common-mode feedback circuit ( $M_{cf1}$ – $M_{cf4}$ ) is shown also in Fig. 3(b).

The studied circuit has been proposed in [9] and investigated in [10]. Its linearity is controllable by parameter  $a$ , equal to

$$a = \frac{(W/L)_{M1-2}}{(W/L)_{M3-6}}, \quad (3)$$

where  $(W/L)_{M1-2}$  is the aspect ratio of transistors  $M_1$  and  $M_2$  and  $(W/L)_{M3-6}$  of  $M_3$  –  $M_6$ . The value of  $a$  determines the achievable linearity and the best linearity is at an optimal  $a$ .

The  $i_o(v_i)$  characteristics of the amplifiers is obtained by multiple time domain simulations with a sinusoidal voltage source at the input (terminals “+ $V_{in}$ ” and “– $V_{in}$ ”), which amplitude is increased gradually. The output current flows through capacitor of 10nF, connected at the output (terminals “+ $V_o$ ” and “– $V_o$ ”). The frequency is 10kHz and the capacitor impedance is about 1.6k $\Omega$ , which is much less than the amplifier output impedance – i.e. the capacitor realizes an approximate short circuit. The transfer characteristics are simulated with AMS 0.35  $\mu$ m CMOS process used in the education. The sizes of all transistors are  $W = 10 \mu$ m,  $L = 0.35 \mu$ m and only the widths of  $M_1$  and  $M_2$  are varied to achieve the desired param-

eter  $a$ . Tail currents  $I_{ss}$  are 35 $\mu$ A for the circuit in Fig. 3(a) and 100 $\mu$ A in Fig. 3(b). Their values are chosen in order to have approximately equal  $g_m$ 's (about 150 $\mu$ S) of both circuits. The simulated transfer characteristics are shown in Fig. 4(a). Since they almost overlap for the amplifier in Fig.3(b), the dependence of THD from the input voltage, plotted in Fig. 4(b), is used to demonstrate how the parameter  $a$  affects the linearity.

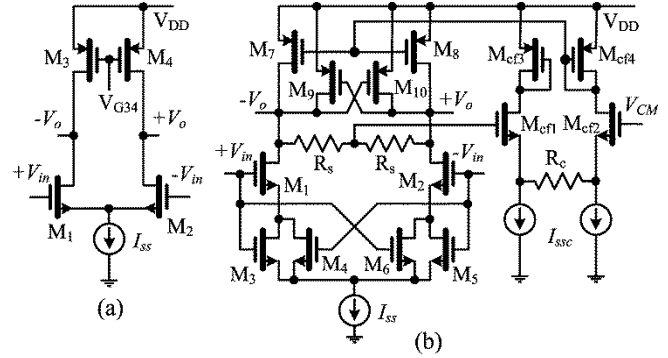


Fig. 3. Circuits used for modelling of OTA nonlinearities: (a) simple differential pair; (b) differential amplifier with linearization.

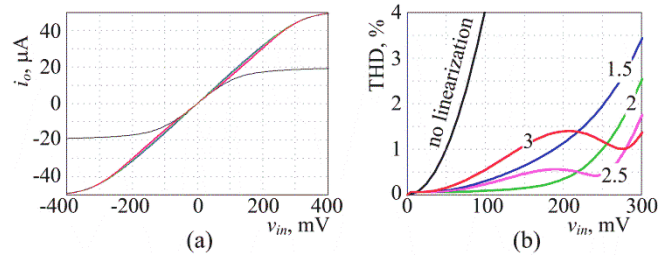


Fig. 4. (a) Transfer characteristics  $i_o(v_{in})$  of the amplifiers; (b) THD of the output current vs. input voltage  $v_{in}$ . Black curves are for the circuit in Fig. 3(a); the others are for Fig. 3(b): blue –  $a=1.5$ , green –  $a=2$ , magenta –  $a=2.5$ , red –  $a=3$ .

The next step is to receive an analytical approximation of the characteristics in Fig. 4(a). Often is used tanh approximation, however the polynomial approximation is preferred here as more flexible, allowing to achieve of better accuracy. The OTAs are fully differential with symmetrical characteristics and the approximating polynomials are odd in the form

$$i_o = \alpha_1 v_i + \alpha_3 v_i^3 + \alpha_5 v_i^5 + \dots \quad (4)$$

The degree of polynomial is increased gradually until achieving relative approximation error of 1% for amplitudes of  $v_i$  up to 0.5V. The polynomial for the OTA in Fig. 3(a) is:  $i_o = 1.454 \times 10^{-4} \times v_i - 0.00242 \times v_i^3 + 0.0325 \times v_i^5 - 0.2694 \times v_i^7 + 1.296 \times v_i^9 - 3.306 \times v_i^{11} + 3.452 \times v_i^{13}$ , where  $v_i$  is in volts and  $i_o$  is in amperes. The polynomials for the OTA with linearization are of 7<sup>th</sup> degree and their coefficients are given in Table 1.

TABLE I. COEFFICIENTS OF APPROXIMATING POLYNOMIALS FOR THE CIRCUIT IN FIG. 3(B).

Parameter $a$	$\alpha_1$	$\alpha_3$	$\alpha_5$	$\alpha_7$
1.5	0.0001648	$-6.587 \times 10^{-5}$	-0.001995	0.004943
2	0.000158	$8.857 \times 10^{-5}$	-0.003008	0.006985
2.5	0.0001503	0.0002396	-0.003932	0.008776
3	0.0001428	0.0003765	-0.004735	0.0103

### III. SIMULATED FREQUENCY RESPONSES OF THE NONLINEAR GYRATOR TANK

Since the amplifiers, forming the gyrator, are nonlinear, time domain analysis should be used for obtaining their frequency characteristics. The small-signal resonance frequency of the gyrator tank is chosen as to be 800kHz in every simulation. A sinusoidal voltage source with amplitude  $V_{im}$  is applied at the input and the maximum of the output voltage  $V_{om}$  is determined by time-domain analysis. The gain in dB for the frequency of the source is calculated as  $20\log_{10}(V_{om}/V_{im})$ . The frequency of the source is varied from 600kHz to 1MHz with step 1kHz and at each frequency is calculated the gain – in this way is obtained the large signal frequency response at the corresponding input amplitude.

The OTAs  $g_{m1}$  and  $g_{m2}$  are considered as nonlinear ideal VCCS, which dependencies  $i_o(v_i)$  are given by the approximating polynomials received in the previous section. Both amplifiers are assumed with identical V-I characteristics. This approach allows to focus on the influence of the nonlinearity of the  $g_m$  only. It also avoids the problems with adjusting the DC voltages at the points of connection of the amplifiers, which could misbalance their V-I characteristics.

The capacitors  $C_1$  and  $C_2$  are equal and their values are calculated from formula (2) in order to have 800kHz pole frequency, taking the transconductances  $g_{m1}$  and  $g_{m2}$  equal to the coefficient  $a_1$  of the approximation polynomial. In order to evaluate the effect of nonlinearities at different  $Q$ -factors the simulations are done for  $Q = 4, 10, 25$ . Its value is adjusted by proper choice of the resistor  $R$  in Fig. 2, changing correspondingly the value of  $g_{m0}$  to keep  $h = 1$ .

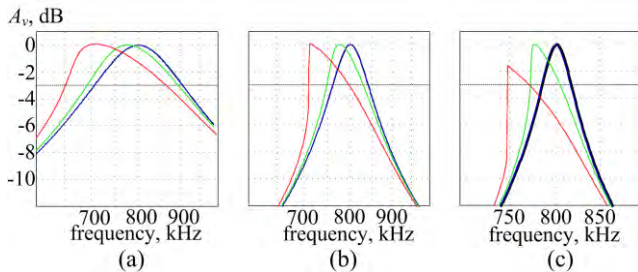


Fig. 5. Frequency responses of the gyrator tank having amplifiers without linearization (Fig. 3(a)). Colours: black – small signal (AC), blue –  $V_{im} = 10\text{mV}$ , green –  $V_{im} = 50\text{mV}$ , red –  $V_{im} = 100\text{mV}$ . (a)  $Q = 4$ ; (b)  $Q = 10$ ; (c)  $Q = 25$ .

The linear frequency responses, received by AC analysis under the assumption of linear amplifiers, are the ideal case and they are used as reference for evaluation the changes of the other frequency responses due to nonlinearities. For this purpose the corresponding AC frequency responses are added in every figure with black colour.

The results for gyrator resonance circuit having amplifiers without linearization are shown in Fig. 5. When the input signal is small – 10mV, THD of the amplifiers is small (~0.3% according Fig. 4(b)) and corresponding curves practically overlap the AC characteristics, i.e. the tank is still linear. Higher input amplitudes change the frequency response and

its deviation is already visible when  $V_{im} = 50\text{mV}$  (1% THD of the amplifiers).

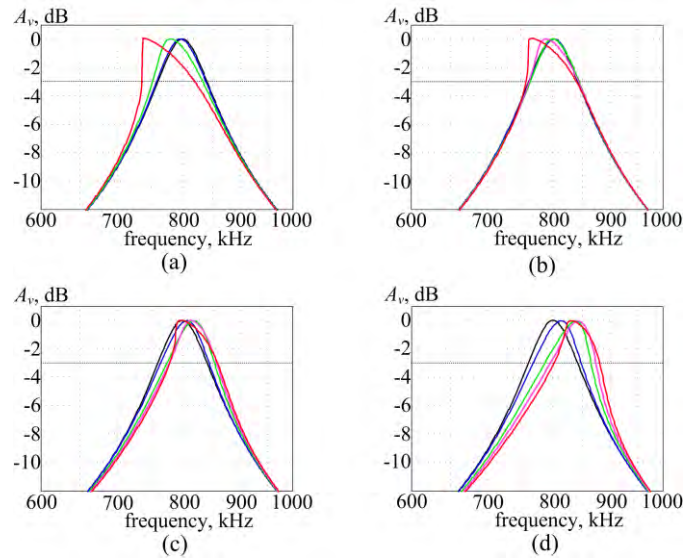


Fig. 6. Frequency responses of the gyrator tank with  $Q = 10$  and amplifiers with linearization. Colours: black – small signal (AC), blue –  $V_{im} = 100\text{mV}$ , green –  $V_{im} = 200\text{mV}$ , magenta –  $V_{im} = 250\text{mV}$ , red –  $V_{im} = 300\text{mV}$ . (a) Amplifiers with  $a = 1.5$ ; (b) amplifiers with  $a = 2$ ; (c) amplifiers with  $a = 2.5$ ; (d) amplifiers with  $a = 3$ .

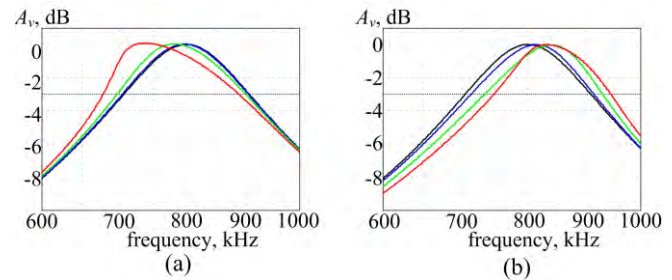


Fig. 7. Frequency responses of the gyrator tank with  $Q = 4$  and amplifiers with linearization. Colours: black – small signal (AC), blue –  $V_{im} = 100\text{mV}$ , green –  $V_{im} = 200\text{mV}$ , red –  $V_{im} = 300\text{mV}$ . (a) Amplifiers with  $a = 1.5$ ; (b) amplifiers with  $a = 3$ .

The use of linearized amplifier in Fig. 3(b) as gyrator OTAs suggests more variants, since the OTA linearity can be controlled by the parameter  $a$ . Fig. 6 shows the family of frequency responses when  $Q = 10$  and  $a = 1.5, 2, 2.5, 3$  (the same values of  $a$  as in Fig. 4). Similar families when  $Q = 4$  and  $Q = 25$  are given in Fig. 7 and Fig. 8 correspondingly.

Several observations can be done from the simulations:

1) The OTA nonlinearities affect the filter frequency response basically around the resonance frequency. The change of the out of band suppression, especially when the attenuation is above 10 dB, is not large.

2) When the amplitude increases enough, the distorting of the frequency response is large and its maximum drops down significantly below the theoretical value of 0dB – e.g. the red curves in Figures 5(c), 8(a) and 8(b).

3) Most of the frequency responses move to lower frequencies when the signal increases. However some of the curves when the parameter  $a$  is 2.5 or 3 move upward. This depends



on the curvature of the function  $i_o(v_{in})$ , which is represented by its derivative  $di_o/dv_{in}$ , shown in Fig. 9(a). When this derivative increases the frequency response moves upward; when it decreases the motion is opposite.

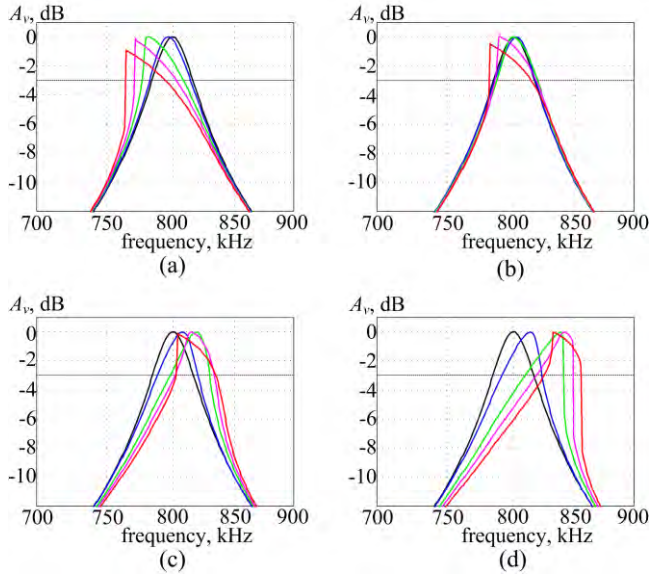


Fig. 8. Frequency responses of the gyrator tank with  $Q = 25$  and amplifiers with linearization. Colours: black – small signal (AC), blue –  $V_{im} = 100\text{mV}$ , green –  $V_{im} = 200\text{mV}$ , red –  $V_{im} = 300\text{mV}$ . (a) Amplifiers with  $a = 1.5$ ; (b) amplifiers with  $a = 2$ ; (c) amplifiers with  $a = 2.5$ ; (d) amplifiers with  $a = 3$ .

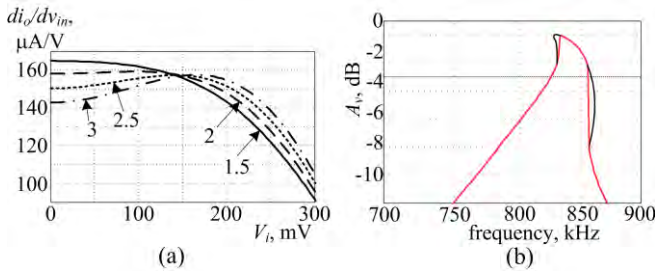


Fig. 9. (a) Dependence of the derivative  $di_o/dv_{in}$  from the input voltage of the amplifier in Fig. 3(b); (b) double inclining of the frequency response for  $Q = 25$ ,  $a = 3$ ,  $300\text{mV}$  input signal.

4) The change of the resonance frequency has very small dependence from the  $Q$  factor if there is no jump in the response. For example when  $a = 2.5$  and the input signal is  $200\text{mV}$ , the maxima are at  $816$ ,  $818$ , and  $818\text{kHz}$  when  $Q = 4$ ,  $10$  and  $25$  correspondingly. However the narrow bandwidth, when  $Q$  is high, makes the circuit more sensitive to the OTA nonlinearity.

5) An interesting case is the red curve in Fig. 8(d), corresponding to  $a = 3$  and  $300\text{mV}$  input signal – it has two jumps. The large input signal covers areas, in which the differential  $g_m$  first increases and after that decreases (Fig. 9(a)). This causes double inclining in the frequency response, shown with black lines in Fig. 9(b).

6) The curves shows that in the gyrator tank THD in the amplifiers above  $0.5\text{-}0.6\%$  cause undesired changes in its frequency response. However this is not strict limit and it can

vary depending on the  $Q$ , area of application and type of OTA nonlinearity.

#### IV. CONCLUSIONS

A gyrator bandpass filters based of fully differential OTAs is investigated by computer simulation concerning the influence of OTA nonlinearity. The collected frequency responses allow to estimate a limit for THD of the amplifiers to  $0.5\%$  approximately when  $Q < 10\text{-}15$ . Then the resonance frequency changes by no more than  $2\text{-}3\%$  and the change of the frequency response is acceptable. It is observed also double inclining of the frequency response in some case – another interesting effects due to nonlinearity of amplifiers

#### ACKNOWLEDGEMENT

The research in this paper was carried out within the framework of Contract No. DUNK 01/03-12. 2009.

#### REFERENCES

- [1] T. Deliyannis, Y. Sun, J. K. Fidler, Continuous-Time Active Filter Design, CRC Press, 1999.
- [2] G. Dinias, T. Deliyannis, Jump phenomenon in a gyrator circuit, Electr. Letters, vol. 11, No. 23, pp. 560-561, Nov. 13, 1975.
- [3] D. L. Hiser, R. L. Geiger, Impact of OTA Nonlinearities on the Performance of Continuous Time OTA-C Filters Bandpass Filters, Proceedings of 1990 Intern. Symp. Circuits and Systems (ISCAS), New Orleans, LA, US, May 1-3, 1990.
- [4] P. Bowron, M. R. J. Motlagh, A. A. Muhieddin, Harmonic characterization of feedback systems incorporating saturation nonlinearities, Electronic letters, vol. 27, No. 20, pp. 1865-1867, Sept. 1991
- [5] X. Zeng, P. Bowron, A. A. Muhieddin, Nonlinear distortion prediction of active filters, Electronic letters, vol. 30, No. 7, pp. 553-554, March 1994.
- [6] S. Szczepanski, R. Schaumann, Nonlinearity-Induced Distortion of the Transfer Function Shape in High-Order OTA-C Filters, Analog Integr. Circ. Signal Processing, 3, 143-151 (1993).
- [7] J. A. Cherry, W. M. Snelgrove, On the Characterization and Reduction of Distortion in Bandpass Filters, IEEE Trans. Circuits and Systems – I: Fundamental Theory and Applications, vol. 45, No. 5, pp. 523-537, May 1998.
- [8] Z. Zhang, A. Celik, P. Sotiriadis, State-Space Distortion Modeling A Weakly Nonlinear, Fully Balanced  $G_m$ -C Filters – A Modular Approach Resulting in Closed-Form Solutions, IEEE Trans. Circuits and Systems – I: Regular Papers, vol. 53, No. 1, pp. 48-59, Jan. 2006.
- [9] Y. Sun, C. J. Jeong, I. Y. Lee, J. S. Lee, S. G. Lee, A 50-300 MHz Low Power and High Linear Active RF Tracking Filter for Digital TV Tuner ICs, Proc. 2010 IEEE Custom Integrated Circuits Conf. (CICC 2010), San Jose, USA, 19-22 Sept. 2010.
- [10] Tanev, I. Uzunov, S. Ouzounov, M. Hristov, Computer Investigation of CMOS Operational Transconductance Amplifier (OTA) with Improved Linearity Implemented on AMS  $0.35\text{ }\mu\text{m}$  Process, Proc. of XXIV Intern. Scient. Conf. "Electronics – ET2015", Sept. 15-17, 2015, Sozopol, Bulgaria.



# Problems and Properties of a Current Amplifier When Realized in Ultra Deep Sub-micron Technology

Simeon Kostadinov<sup>1</sup>, Ivan Uzunov<sup>2</sup>, Dobromir Gaydazhiev<sup>3</sup>

**Abstract** – In this paper are described three versions of a current amplifier and each of them is realized with FETs having different channel length: 90nm, 45nm and 30 nm. Their basic properties are simulated and discussed, demonstrating the major benefit of the shortening of the channel length – extension of the frequency bandwidth. The problems arising with the shorter channels length are also considered briefly.

**Keywords** – amplifiers, current mode operation, short-channel effects, MOSFET.

## I. INTRODUCTION

The scaling of the semiconductor devices in modern CMOS technology improves the performance, lowers the price and reduces the power supply of the fabricated circuits [1,2]. While device downscaling is driven primarily by the benefits it has in digital design, it introduces new effects in transistor behavior, which makes analog design challenging. As channel lengths decrease below 1 $\mu$ m, second order effects, which were not taken into account for long channel FETs must be accounted for. When short channel effects become significant the traditional analytic approach in analog design is not feasible [3]. This necessitates a modification of the existing analytic design methodologies and employment of new techniques.

Shrinking the gate length is accompanied by decrease in power supply voltage, which is a major issue in analog design. However threshold voltage does not scale as much as the supply voltage, and transistors are biased at lower voltages, which results in worse transistor properties [1,4].

Current mode operation overcomes some of the problems with downscaling because it requires lower supply voltage compared to its voltage mode counterparts [4,5]. It also has some other advantages – at low power supply voltage they allow wider signal dynamic range, wider frequency bandwidth, better linearity. For these reasons, current mode designs become very popular in many high performance analog and mixed signal applications – RF front ends, data converters, oscillators, etc.

The legacy design methodology applicable for long channel devices uses theoretical circuit analysis based on expressions taking into account only first order effects. In contrast,

designing an amplifier in sub-micron technology must include the higher order effects, rendering a purely analytical approach no longer viable, since it leads to significant inaccuracies mainly related to the specific short channel effects – drain induced barrier lowering, channel length modulation, velocity saturation [1,3,4]. Even some technology specific parameters, e.g.  $\mu C_{ox}$ , are not constant anymore and depend on transistor geometry.

A method for design of a basic current amplifier, when short channel effects are significant is proposed in [6]. For the cases when short channel effects are moderate the design procedure gives very good matching between specification and simulation results. Further reduction of device sizes causes more and more deviation from the basic theory. The goal of this paper is investigation of these effects when the sizes approach the technology defined limits. A basic current amplifier is designed with different FET devices with different channel lengths: 90nm, 45nm, 30nm.

## II. BRIEF REVIEW OF THE DESIGN METHOD

The circuit of a basic current amplifier is shown in Fig.1. The major problem in its design is the requirement of identical transconductances of  $M_1$  and  $M_3$  at the same drain currents. This is necessary in order to keep the symmetry in upper and lower halves of the circuit. The formulas for long channel device transconductance and drain current are [2,5]:

$$g_m = \mu C_{ox} \frac{W}{L} V_{ov}; I_D = \frac{\mu C_{ox} W}{2L} V_{ov}^2 (1 + \lambda V_{DS}), \quad (1)$$

where  $\mu$  is the carrier mobility,  $C_{ox}$  is gate capacitance per unit area,  $W$  and  $L$  are channel width and length,  $V_{ov}$  is overdrive voltage and  $V_{DS}$  is drain-source voltage of the transistor. The widths of  $M_1$  and  $M_3$  should be connected with the  $n$ - and  $p$ -carriers mobilities by the relationship  $W_1/W_3 = \mu_p/\mu_n$  in order to have equal  $g_m$ 's of these transistors [6]. It follows from (1) and is valid for long-channel transistors only.

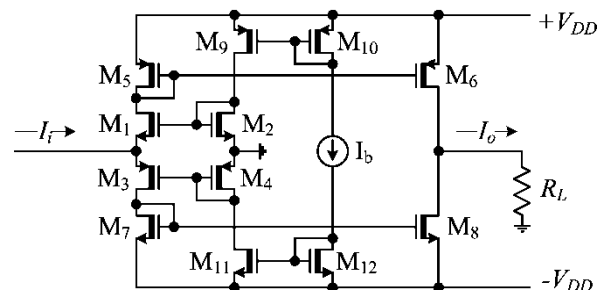


Fig.1. The circuit of the current amplifier

The above mentioned reasons do not allow to implement this relationship for short channel transistors. The design

<sup>1</sup> S. Kostadinov is with Smartcom Bulgaria AD, Department of Microelectronics, 36 Dragan Tzankov blvd., 1040 Sofia, Bulgaria, e-mail: simeon\_kostadinov@smartcom.bg.

<sup>2</sup> I. Uzunov is with SmartLab AD, BIC IZOT, Build. 1, 133 Tzarigr. Shosse blvd., 1784 Sofia, Bulgaria, e-mail: ivan\_uzunov@smartcom.bg.

<sup>3</sup> D. Gaydazhiev is with Smartcom Bulgaria AD, Department of Microelectronics, 36 Dragan Tzankov blvd., 1040 Sofia, Bulgaria, e-mail: dobromir\_gaydazhiev@smartcom.bg

procedure proposed in [6] uses a graphical and analytical approach for initial design centering and then refining the device dimensions by parametric simulation. Firstly the dependences of  $g_m/W$  and  $I_D/W$  from gate-source voltage for PFET and NFET are plotted [2]. Short channel effects make these normalized curves slightly different. However the differences are small and will be compensated in the following optimization. The next steps are:

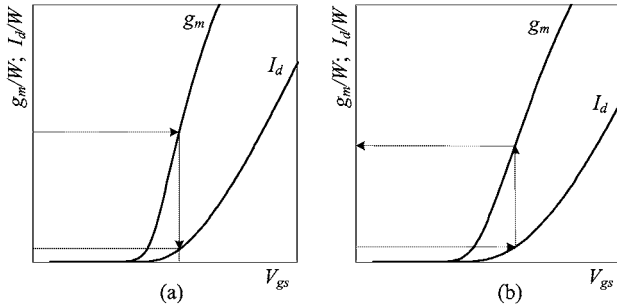


Fig.2. Plots of normalized  $g_m$  and  $I_D$  vs.  $V_{GS}$ : (a) NMOS; (b) PMOS.

1) The transconductance of the input transistors are defined by the input impedance:

$$g_m = g_{m1} = g_{m3} = \frac{1}{2.2R_i} \quad (2)$$

A normalized value of  $g_m$  for NMOS is chosen in the linear region of the curve (Fig. 2(a)). The width of the transistor can be determined by dividing the desired transconductance by the chosen value.

2) The normalized  $g_{m1}$  defines  $V_{GS}$  and the corresponding normalized value of  $I_{D1}$  of  $M_1$ . Since the width of  $M_1$  is already known the required drain current of  $M_1$  can be calculated.

3) The drain currents for  $M_1$  and  $M_3$  must be the same and equal to the total input branch current –  $I_{D1} = I_{D3} = I_D$ . Looking at Fig. 2(b) we pick a value  $I_{D3}/W$  in the region where drain current is a quadratic function of  $V_{GS}$ . This value also defines the  $V_{GS}$  and the width of  $M_3$ .

4) For the selected value of  $I_{D3}/W$  it should be verified that the corresponding  $g_m/W$  is in the linear region in Fig. 2(b). If it is not, another value of  $I_D/W$  should be chosen.

5) These steps do not guarantee the exact matching of  $g_{m1}$  and  $g_{m3}$  since the transistors are of different types. The received values will be used as initial for the following optimization based on parametric analysis, in which the values of  $g_{m1}$  and  $g_{m3}$  are equalized at the same drain current.

### III. PROBLEMS ARISING WITH DEVICE DOWNSCALING

To study the changes in the amplifier parameters and behavior, it is designed in 32nm bulk CMOS technology, developed for SRAM, logic and mixed-signal applications. Three versions of the amplifier are considered: 1) realized with hpar FETs with gate length of 90nm; 2) realized with slvt FETs with gate length of 45nm; and 3) realized with slvt FETs with gate length of 30nm (the minimum allowed drawn gate length for thin oxide in the used 32nm technology). The abbreviations hpar and slvt are for different types of MOS

transistors in the Process Design Kit [7]. All three version are supposed to meet the following specifications:

- input impedance  $R_i < 500\Omega$ ;
- power supply  $V_{DD} = \pm 1V$ ;
- current gain  $A_i$  equal to 1, i.e. current buffer.

The design is based on the procedure outlined in the previous section. All transistors are working in strong inversion in saturation and the corresponding DC branch currents are listed in Table 1.

TABLE I  
DC CURRENTS IN THE AMPLIFIER BRANCHES

Current in $\mu A$ through	$I_b$	$M_5, M_1, M_3, M_7$	$M_9, M_2, M_4, M_{11}$	$M_6, M_8$
Version 1	74	72.6	73.2	98.4
Version 2	93	79.3	85.7	121.5
Version 3	115	97.6	105.4	186

Several problems were faced during the design:

1) Due to channel length modulation the current mirrors ratios depend strongly on the drain-source voltage. Then the current gain deviates from the ratio of the areas  $M_5/M_6$  and  $M_7/M_8$  (the theoretical value of the gain). For example, if these transistors have equal areas, the gain for Version 1 is 1.18, for Version 2 it is 1.25, and for Version 3 it is 1.3. An accurate value of current gain can be achieved by varying the channel widths with parametric analysis.

2) Decreasing the channel length relatively reduces the back-gate transconductance  $g_{mb}$  concerning the main transconductance  $g_m$ . In [6]  $g_{mb}$  is estimated as more than 10% of  $g_m$  and its contribution is reflected in formula (2) by the factor 2.2. For shorter channels  $g_{mb}/g_m$  is less than 5% and this factor should be reduced to 2.1 or even 2.

3) The threshold voltage  $V_{th}$ , which is usually considered to be constant, in fact depends on the drain-source voltage  $V_{ds}$  due to drain induced barrier lowering (DIBL effect). This is a common short channel effect in FETs. The variation of the threshold voltage is investigated by simulation for the used transistors and the corresponding plots are shown in Fig.3.

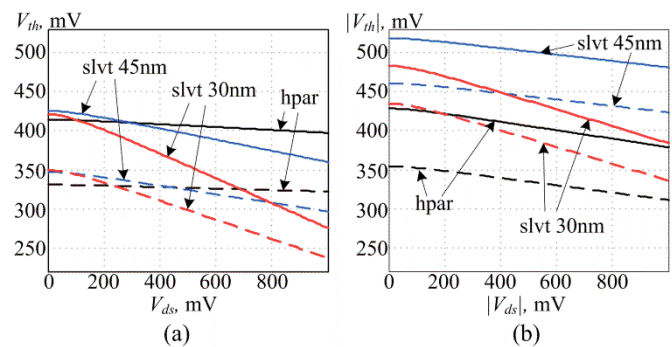


Fig. 3. Dependence of the threshold voltage from the drain-source voltage for the transistors of interest. Solid lines - with bulk effect, dashed lines - without bulk effects: (a) NMOS; (b) PMOS.

Several effects are observed in Fig.3. As expected, shorter channels have greater dependence of  $V_{th}$  on  $V_{ds}$ . Transistor type hpar is optimized for analog applications, which results in negligible difference of threshold voltages for PFET and

NFET and their deviation with  $V_{ds}$ . This is not the case for slvt FETs, which are basically intended for use in digital circuits –  $V_{th}$  has stronger dependence on  $V_{ds}$  and the difference between  $V_{th}$  of PFET and NFET is more than 20%. However overcoming the challenges in amplifier design based on slvt FETs proposes some significant benefits – small die area and operation at higher frequencies.

4) The breakdown voltages for the used transistors are low (1V) for all three types. On the other hand, threshold voltages are relatively high – up to 0.5V, while the circuit has four transistors in series in most of the branches. Both circumstances require operating at a total supply voltage, which is more than the limits of the transistors and here  $\pm 1V$  is used. To guarantee the safe operation it is necessary to check the voltages over transistors at different input current. The limits for the magnitude of the input currents are defined with the following procedure: a DC current source is applied at the input and its current is varied from -3mA to 3mA with DC sweep analysis. The plots of the voltages, which exceed the limits, are given on Fig. 4 for amplifiers with hpar and slvt 45nm. The corresponding plots for slvt 30nm are very similar to slvt 45nm and they are not shown. The maximum input current  $I_{imax}$  determined from the figure are: 1.2mA for Version 1 (hpar), and 1mA for Versions 2 and 3 (slvt).

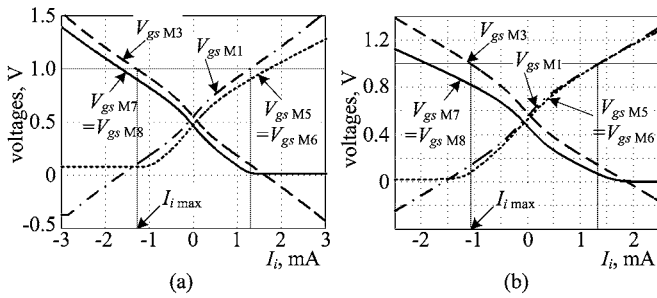


Fig. 4 Gate-source voltages, which exceed the safe operation range: (a) amplifier with hpar FETs, (b) amplifier with slvt 45nm FETs.

#### IV. SIMULATIONS OF THE BASIC AMPLIFIER PARAMETERS

The three versions of the amplifier are compared by several basic parameters: small and large signal current gain, input impedance, and transfer characteristic ( $I_o$  vs.  $I_i$ ) and THD. The small signal parameters are shown in Fig. 5. The current gain is different for the three versions for reasons commented above. More interesting are the frequency bandwidths. The corners frequencies are: 12.03 GHz for Version 1, 25.6 GHz for Version 2 and 45.9 GHz for Version 3. For comparison, the same amplifier designed with another type of MOSFETs from the same technology – zgfets with 270nm minimum gate length [6] has corner frequency of 2.3GHz. The benefit of using short channel transistors is obvious. The input impedance is close to the designed and it keeps its value up to few GHz.

The large signal behavior is illustrated in Fig. 6. They are obtained by parametric time domain analyses using sinusoidal input current source with parameterized amplitude. Its frequency is fixed at 10 KHz - low enough to avoid slew rate

effects and suppression of the harmonics, generated in the amplifier, from its frequency response.

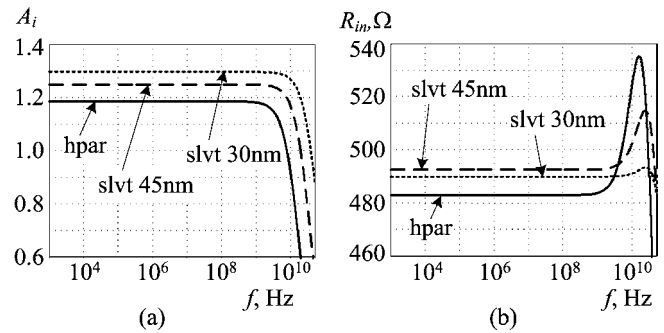


Fig. 5 Small signal characteristics of the designed amplifiers: (a) current gain; (b) input impedance.

The dependence of the amplitude of the output current from the amplitude of the input current is shown in Fig. 6(a). The curves are very close to each other due to the approximately equal current gains of all three versions and it is difficult to distinguish them. For this reason two other characteristics are added: large signal current gain  $A_i = I_o/I_i$  and the THD, both as functions of the input amplitude. These characteristics show that the decreasing of the channel length increases the nonlinearity. Evidently this is the price for the extended frequency bandwidth. However, the amplifier with hpar transistors has similar THD as the amplifier with three times longer (270nm) zgfet transistors, which has 1.7% THD at  $I_i = 1mA$  [6].

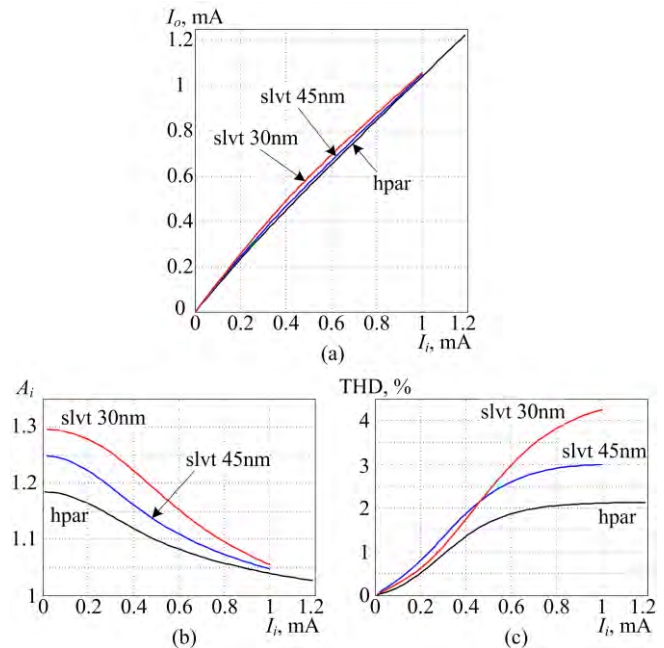


Fig. 6 Dependences of large signal parameters of the designed amplifiers from the input current: (a) output current; (b) large signal current gain; (c) THD.

## V. CONCLUSION

One of the most common current amplifiers is designed with three different types of deep submicron MOSFETs from a 32nm CMOS technology: 90nm hparfet, 45nm and 30nm slvtfet. The major benefit of short channel transistors is significant extension of the frequency bandwidth – up to 45 GHz for the circuit with shortest channel transistors. The large signal behavior is also investigated and the hpar version of the circuit shows similar non-linearity compared to the same circuit with much longer transistors. Certain increase of the non-linearity is observed for the versions with shorter channel transistors (slvtfets). The maximum amplitude of the input current is lower due to the reduced operating voltages of the used transistors, but this limitation can be relaxed by using wider FETs and higher biasing current.

## REFERENCES

- [1] A.-J. Annema, B. Nauta, R. van Langevelde, H. Tuinhout, "Analog Circuits in Ultra-Deep-Submicron CMOS", IEEE Journal Solid State Circuits, Vol.40, No 1, pp. 132–143, Jan. 2005.
- [2] R. Jacob Baker, *CMOS Circuit Design, Layout, and Simulation*, 3<sup>rd</sup> ed., IEEE press, 2010.
- [3] E. Manolov, "Graphical Approach to Dimensioning of Amplifying Stages Using CMOS Submicron Transistors", Transactions of National Forum "Electronics 2015", Sofia, Bulgaria, May 14-15, 2015, pp. 79-83, (in Bulgarian).
- [4] L. Safari, S. Azhari, "A novel low voltage very low power CMOS class AB current output stage with ultra high output current drive capability", Microelectronics Journal 43 (2012), pp. 34-42.
- [5] W. Sansen, *Analog Design Essentials*, Springer, 2006.
- [6] S. Kostadinov, I. Uzunov, D. Gaydazhiev, E. Manolov, "Design Considerations for Current Mode Amplifier in Deep Sub-micron CMOS Technology", XXIV International Scientific Conference ELECTRONICS ET 2015, Sept. 15-17, Sozopol, Bulgaria.
- [7] IBM 32LP Technology Design Manual, IBM Corp., 2010



# Optimization and analysis ALL Digital PLL For resonant inverter

Goran Goranov <sup>1</sup>, Nikolay Madzharov <sup>2</sup>

**Abstract** – The purpose of this article is to determine the optimal ALL digital PLL for control resonant inverters. Various methods of synthesis of the digital frequency are examined and compared. On the base of the results we determine which control approach is applicable. Phase detector, loop filter and digital controlled oscillator realized in programmable logic are used.

**Keywords** – Logic gates, Control system, PLL, CPLD.

## I. INTRODUCTION

The widespread use of transistor converters in various areas of practice - industry, medicine, service sector etc., imposes the new requirements to the control systems related with regulation and maintenance of certain operating modes. The implementation of the PLL functions into control systems for power converters and their investigation are presented in details in [1], [2], [3]. The use of programmable logic devices allows for the development of such systems by expanding their functionality and adaptability.

The optimization and analysis on this digital PLL gives us opportunities to work in a relatively wide frequency range and small steps for adjustment. This is also and an easy connection to other digital systems. The digitally frequency synthesis is based on programmable divide a quartz stabilized frequency to obtain a change of the output frequency in a certain range and a predetermined step. There are various methods for digital frequency synthesis - a divisor of N, direct digital synthesis (DDS), I/D - counter, fractional N, division by comparison schemes, etc. They are hardly realizable in a discrete implementation [4], [5], [6].

The analysis of digital PLL is limited to the examination of each module separately - Fig.1.

In the previous studies and researches of the author feasibility of a digital PLL, using I/D - counter and FRACTIONAL N method are affected. The conclusion is that various digital controlled generators require a substantial change in the functions of the previous blocks in the digital PLL. Often this refers to the digital filter, which is connected to the generator. Phase detector in most of the cases does not change unless you change monitoring requirements for the phase difference. Figure 1 shows a digital PLL, comprising a

phase detector DPD, converting the phase difference into a signal of a specified duration, the digital filter DLF, converting the signal from the phase detector and the digital controlled oscillator DCO. In operation of the system, frequency  $F_{out}$  of the output signal of the DCO seeks to achieve input frequency  $F_{in}$  with certain accuracy. The change of the input frequency  $F_{in}$  generates an increase or decrease of the phase difference between it and the output frequency  $F_{out}$ , which leads to their alignment ( $F_{in} = F_{out}$ ).

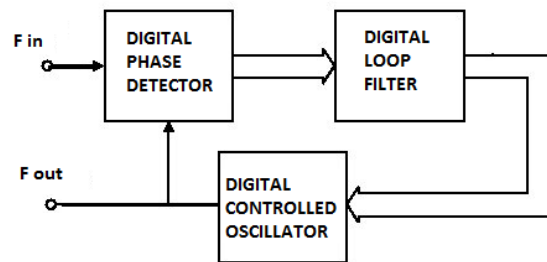


Fig. 1. ALL digital PLL

## II. SYNTHESIS OF THE BASIC MODULES

### Phase detector.

The most felicitous for use from the three main types of phase detectors (XOR, JK-Flip Flop and frequency phase detector) in the management scheme of the resonance inverters is the frequency phase detector. This is a detector with three states, managed by the front of the input signal. The range is  $\pm 360^\circ$ . From the timing diagram shown in Figure 2 we can see the advantage of the phase frequency detector. In phase difference  $\varphi = 0$  signals UP and DN are also zero. Depends on the direction of the phase shift, in one of the two outputs, is obtained the pulse K3, proportional to the phase difference.

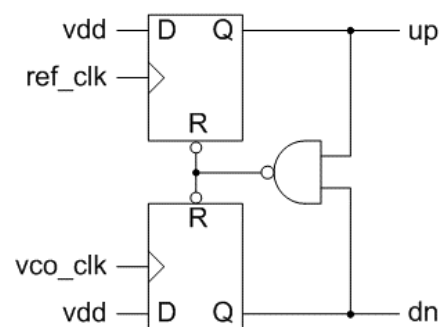


Fig. 2 a)

<sup>1</sup> Assist. Prof. Goran Goranov, Ph. D., at Department "Electronics" of Technical University of Gabrovo, 5300 Gabrovo, 4 Hadji Dimitar st., Bulgaria, e-mail: [g\\_goranov@bitex.bg](mailto:g_goranov@bitex.bg).

<sup>2</sup> Assoc. Prof. Nikolay Madzharov, Ph.D., at Department "Electronics" of Technical University of Gabrovo, 5300 Gabrovo, 4 Hadji Dimitar st., Bulgaria, e-mail: [madjarov@tugab.bg](mailto:madjarov@tugab.bg).

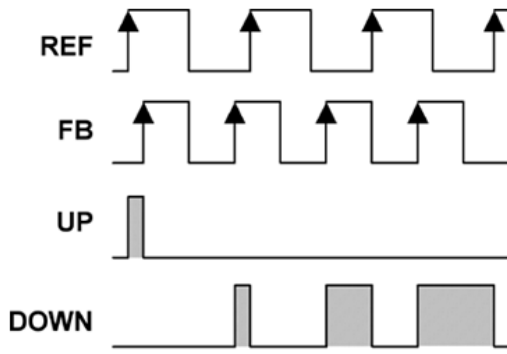


Fig. 2 b)

### Digital controlled oscillator

#### FRACTIONAL - N

The block diagram representing the method consists of two main blocks for division of an integer and a fractional part - Fig. 3

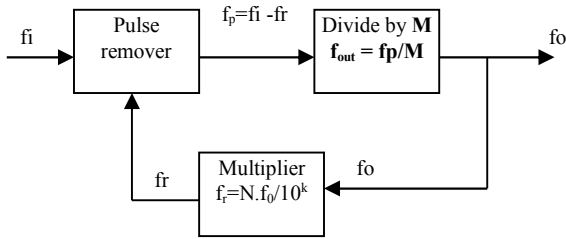


Fig.3. Basic Fractional-N.

The input clock frequency  $f_i$  is used. Part of the pulses is removed in reducing impulses block. The number of the removed pulses is proportional to the ratio  $N$ , issued by the block of the multiplier. The removal of the phase noise (asymmetry between successive pulses) is achieved most easily as "blur" in time. The division factor  $M$  is used, which is an integer - block divider of  $M$ . The output frequency is calculated by the following formula:

$$f_{out} = \frac{f_p}{M} = \frac{f_i - f_r}{M} = \frac{f_i - \frac{N \cdot f_{out}}{10^k}}{M}, \quad (1)$$

from where

$$f_{out} = \frac{f_i}{M + \frac{N}{10^k}}, \quad (2)$$

$$f_{out} = \frac{f_i}{M + \frac{N}{10^k}} - \frac{f_i}{M + \frac{N-1}{10^k}}, \quad (3)$$

In the presented formulas  $M$  is an integer,  $K$  is number of decades, and  $N$  is an integer in the range of  $0$  to  $10^k - 1$ . On this basis, for the realization of FRACTIONAL -  $N$  method are needed binary decimal counters for  $M$  divider and multiplier for divisor  $N$ . If the coefficient of division  $M$  varied from  $0$  to  $9$ , you will need a counter with two digits of  $M$  i.e. an amendment from  $0$  to  $99$ , respectively, two counters and etc. The same goes and for the divisor  $N$ , which determines the fractional part. For example in the number of decade  $K=2$  is obtained by dividing the frequency  $1,00$  to  $99,99$ .

The implementation of the digital Fractional- $N$  divider for management of the resonant inverter has the following advantages and disadvantages:

- Works in a wide frequency range.
- Receives relatively low phase noise and hence little impact on the optimal operation of the power transistor converters.
- Easy to implement coordinating digital filter

#### Disadvantages:

- Nonlinear step of modifying the frequency - depending on the divisor.
- Receives relatively low phase noise and hence little impact on the optimal operation of the power transistor converters.

### DDS – generator

DDS – generator is a type of generators, which is driven by a digital code (word), called the DCO. It works with digital converter filter time-digit code. On Fig. 4 is shown the block structure of this type of frequency synthesis.

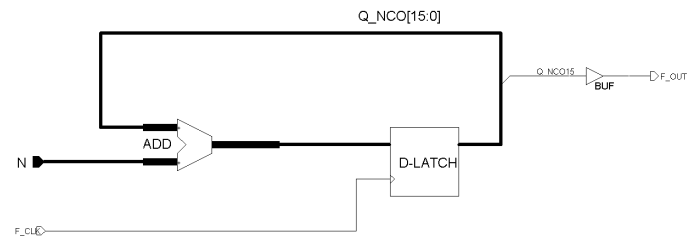


Fig. 4. DCO.

The output frequency is given by:

$$f_{out} = f_{clk} \cdot N / 2^k \quad (4)$$

where  $2^k$  is the sparsity of the ADD.

The amendment of  $f_{out}$  varies from  $0$  [Hz] to  $f_{clk}/2$  [Hz] with step:

$$\Delta f_{out} = f_{clk} / 2^k \quad (5)$$

Advantages and disadvantages of using DDS method for managing resonant inverters are:

- Working in a wide frequency range  $\max \frac{1}{2} f_{in}$ .
- A very small step of frequency variation  $\Delta f$ .
- Easy to implement digital matched filter.

**Disadvantages:**

- The main disadvantage is the phase noise, its value reaches  $1/f$ , at  $f_{out}=1/2f_{in}$ .
- Requires large program capacity.

**Increment/decrement counter**

This is a bias scaler, whose amendment is managed by two entrances carry and borrow - fig.5.

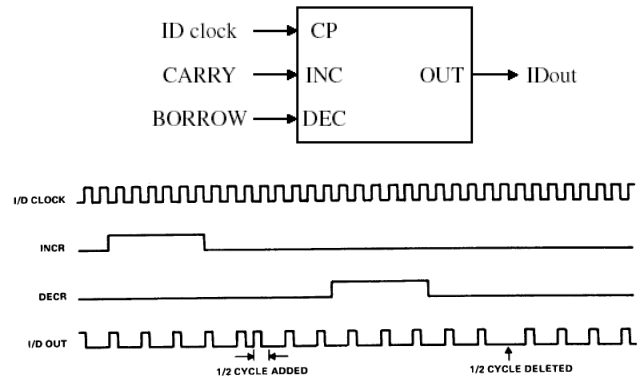


Fig. 5. I/D - counter and timing diagram

The resulting output frequency (I/Dout) of I/D counter – is  $\frac{1}{2} I/Dclk$ . It may be amended "up" or "down" depending on which input is received 1-0 transition. The amendment is obtained by adding or deleting  $\frac{1}{2}$  cycle. To reduce the phase noise can be used an additional counter (divisor N), then phase noise will be reduced by  $1/N$ .

Advantages and disadvantages in using this method are:

- Relatively small step of frequency variation  $\Delta f$  (tens of Hz).
- Small phase.

**Disadvantages:**

- Complex to implement digital filter
- Limited range of frequency variation
- Dependence on many parameters.

**Loop filters.**

The task of the filters is to convert the signal so that it is convenient to control the synthesized oscillator. Depending on the digital synthesized adjustable generator are used various filters. If the used generator is I/D – counter, then a filter that converts the signal phase difference from the phase detector in the number of pulses is needed. Thus plays the role of an integrator [8]. Such a filter can be realized by a reversible counter with an adjustable division ratio, called K-counter - Fig. 6. The principle is explained with tim diagrams shown in Fig. 7. The number of output pulses (transfers) is proportional to the operating time of the respective counter UP or DOWN, respectively, and the phase difference.

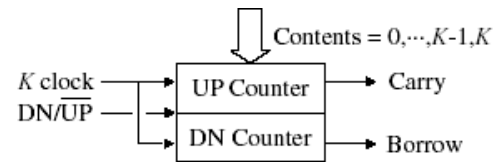


Fig. 6. Programmable frequency divider of K - counter.

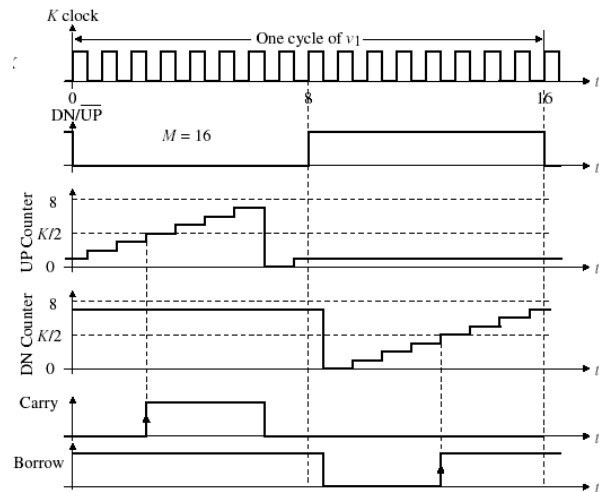


Fig. 7. Time chart of K - counter.

Another type of filter is converter of time into a digital code. The resulting code (word) N directly operates the generator - DDS or Fraction-N, changing its output frequency proportional to the code N. The condition for synthesis of the converter is  $f_{clk} \gg \text{EVENT}$

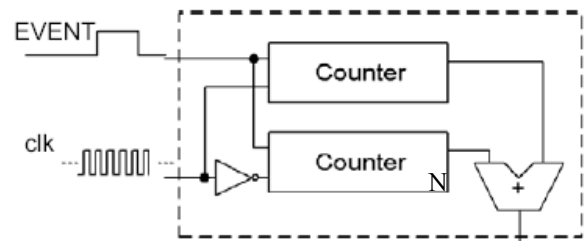


Fig. 8. Converter time numerical code.

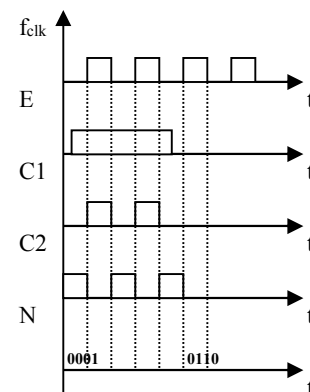


Fig. 9. Time chart of converter - time in numerical code.

Between the considered options are development and synthesized PLL scheme with DDS generator and filter T2C (time to code converter) in CPLD Xilinx CoolRunner2.

The experimental object for management is a transistor half-bridge parallel inverter, which is running with frequency above the resonance one. The input power of the converter is  $P = 1,6 \text{ kW}$ . Load capacitance value is  $C = 1,72 \mu\text{F}$ . It is implemented with non-inductive capacitors. The load inductance - the inductor has a value  $L = 3,1 \mu\text{H}$ , wound copper tube with a diameter of 9 mm with 13 windings. The diameter of the inductor is 73 mm.

A study of system performance under different input power and loads has been made. On Fig. 10 a) is shown the current and voltage in resonance at low load and input voltage  $E = 130\text{V}$  and input current  $I = 12,3\text{A}$ . On the channel one of the oscilloscope is visualized the form of the current and on the channel two the form of the voltage. Both signals were observed after the transformers for feedback. It can be seen that the mode of the inverter is good, when the frequency is maintained by digital PLL. The increase of the input voltage leads to bigger power load. This will lead to its heat and therefore to change the parameters of the circle and therefore the frequency will start to decrease.

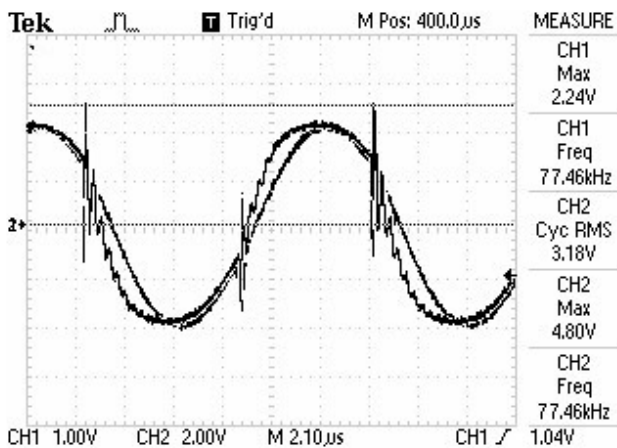


Fig. 10 - A) Voltage on the parallel circuit and the current - hollow material.

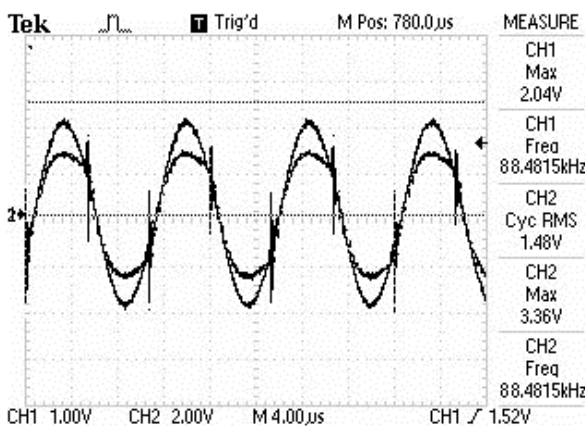


Fig. 10 - B) Voltage on the parallel circuit and the current - dense material.

At higher load (when solid material is used), the operating frequency is changed on 88,48 kHz - Fig. 10 b). Input voltage  $E = 122\text{V}$ ,  $I = 10,7 \text{ A}$ .

### III. CONCLUSION

The analysis shows that the best results are achieved when synthesized management scheme with 16bit DDS generator is used. A 16 bit counter for filter, which converts signal from the phase detector is realized. The response time of the system as a whole depends on the sparsity of the filter and the time of integration. Time integration is proportional to the signal EVENT - phase difference. Under the process of locking frequency (resonance inverter) phase deviation is small and therefore the reaction has several clock cycles. Great importance has the supporting clocked. Although the inverter operates at frequencies from 60 to 100 kHz by using 50MHz reference frequency. Thus the influence of phase noise is minimized.

Good results are achieved also when I/D counter is used, but the digital filter is more complex and requires more programmable logic resources.

Using methods for fully digital control improved adaptability to any electronic system, provide much easier monitoring of parameters and setting of such microcontrollers.

### REFERENCES

- [1] Donchev B., M. Hristov, P. Goranov, *Design of control system for inductive heating converter*, ET 2001, Sozopol, Bulgaria.
- [2] Olsson T., and P. Nilsson. *A Digital PLL made from Standard Cells*", Proceedings of the 15 th European Conference on Circuit Theory and Design In Proceedings of ECCTD' 01, Helsinki University of Technology, Finland 28 th-31 st August 2001, pp. 277 – 281.
- [3] Goranov G., Digital synthesis of PLL for control induction heating inverters, ISC Electronics 2004 - pp. 39 – 44.
- [4] Denny Ambramovitch, Lyapunov Redesign of Classical Digital Phase-Locked Loop, Communication and Optical Research Lab 2003;
- [5] Analog Devices, Fractional-N Frequency Synthesizer – ADF41xx, 2006.
- [6] В.Майская, "Fractional N Синтезаторы" ЭЛЕКТРОНИКА: Наука, Технология, Бизнес, стр. 10 -16, бр.5, 2002.
- [7] Maini, A.K., „Digital Electronics - Principles, Devices and Applications”, John Wiley & Sons Ltd, ISBN: 978-0-470-03214-5, Chichester, England, 2007;
- [8] Thomas Olsson, Peter Nilsson, *A Digital PLL made from Standard Cells*, Lund University, Dept. of Electrosience, Sweden 2001.



# Design and Realization of a 12W Quasi-Resonant Flyback Converter

Zoran Zivanovic<sup>1</sup> and Vladimir Smiljakovic<sup>2</sup>

**Abstract** – In this paper the design of a quasi-resonant flyback converter is presented. The operating principles are briefly explained, including the functional block diagram. The prototype has been built and experimental results are presented to support the theoretical analysis and to demonstrate the converter performance.

**Keywords** – CCM, DCM, Flyback, EMI, Quasi-resonant

## I. INTRODUCTION

The most widely used isolated switch mode power supply topology is the flyback converter. Power levels are in low to mid power range, from 1 W to 50 W or even 100 W for low current outputs. The single magnetic component is actually not a transformer but a coupled inductor that combines functions of energy storage, energy transfer and isolation. The main advantages are wide input voltage range, with multiple outputs, higher or lower than the input voltage and simplicity. Depending on the design of the coupled inductor the flyback converter can operate either in Continuous Conduction Mode (CCM) or Discontinuous Conduction Mode (DCM). However flybacks are not perfect. Major drawbacks are poor magnetics utilization and higher electromagnetic interference as a result of hard switching with high peak currents.

## II. QR FLYBACK CONVERTER BASICS

A quasi-resonant converter is a distant cousin of a resonant converter. However, the waveforms are not sinusoidal like at true resonant converter, in fact they have typical flyback shape. Also QR flyback power stage does not seem any different from a classical DCM flyback. The “quasi-resonance” is effective in the dead time after the core is demagnetized. The switching occurs at the valley of the resonant ringing generated by the circuit parasitics. There is no need to add inductors and capacitors because they are already there. Soft switching results in switching loss reduction because turning the MOSFET on at the valley of the resonant ringing lowers the losses associated with the output parasitic capacitance of the MOSFET. Another advantage is a reduction of conducted and radiated electromagnetic interference. Also the jitter that results from searching for valley of the ringing spreads the frequency spectrum reducing

electromagnetic interference.

The QR flyback converter must operate in DCM because the core must be completely demagnetized. A specialized controller detects core demagnetization and resonant valley and initiates next turn on cycle. In a conventional flyback the constant frequency oscillator initiates the next turn on cycle, so turn on can be at any point of the resonant ringing. QR controllers rely on the switching waveform for valley detection, by detecting the change in slope or a zero-crossing threshold. Some controllers modulate both the switching frequency and the primary current over the most operating range and the others modulate the switching frequency but maintain a constant peak primary current over the most operating range.

The UCC28600 from Texas Instruments (Fig.1) is QR controller operating in different modes, modulating the peak primary current proportional to load and the switching frequency inversely proportional to load.

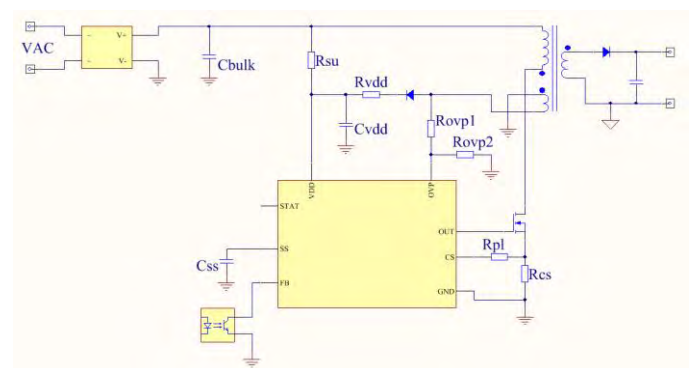


Fig.1. QR flyback converter

For loads below 10 % of rated power the controller will operate in green mode (GM) regulating the output using burst of 40 kHz pulses. The number of pulses increases until the converter is switching consistently at 40 kHz. Frequency fold-back mode (FFM) is reserved for loads between 10% and 30% of rated power. The output is regulated by modulating the frequency from 40 kHz to 130 kHz, holding the peak primary current constant. Finally, for higher loads between 30% and 100% of rated power the controller will operate in either DCM, where the peak primary current is modulated but the switching frequency is clamped to maximum (130 kHz), or quasi-resonant mode (QRM), where the peak primary current and the switching frequency are both modulated.

The above boundaries of steady-state operation are approximate because they are specific design dependent. They are programmed by the flyback transformer and the four resistors  $R_{PL}$ ,  $R_{OVP1}$ ,  $R_{OVP2}$  and  $R_{CS}$  on the Fig.1.

<sup>1</sup>Zoran Zivanovic is with the IMTEL KOMUNIKACIJE AD, Bul. Mihajla Pupina165b, 11070 Belgrade, Serbia, E-mail: zoki@insimtel.com.

<sup>2</sup>Vladimir Smiljakovic is with the IMTEL KOMUNIKACIJE AD, Bul. Mihajla Pupina165b, 11070 Belgrade, Serbia, E-mail: smiljac@insimtel.com.

### III. DESIGN AND ANALYSIS

The task is to design a 12 W QR flyback converter using the UCC28600 controller with careful choice of operating parameters and components. The footprint size must be around 80x40mm. Design specifications are given in Table I.

TABLE I  
DESIGN SPECIFICATIONS

		Min	Typ	Max	
Input voltage	$V_{IN}$	160	220	264	$V_{AC}$
Output voltage I	$V_O$	+12			V
Output power	P	12			W
Full load efficiency	$\eta$	80			%
Output ripple voltage	$V_{RIPPLE}$	$\leq 120$			mV
Switching frequency	f	40		130	kHz

The transformer (inductor) must be sized at the minimum input voltage, maximum power and between frequency clamps (40 - 130 kHz). Knowing that, a good choice of core for the transformer is E20/10/6, N87 material from TDK. Reinforced insulation is a must, but the effectiveness of which must be verified by dielectric strength testing.

Using a switching frequency of 80 kHz we can now calculate the maximum primary inductance using the following equation

$$L_{Pmax} = \left( \frac{V_{DCmin}(V_O + V_F)N \times 0.925T}{V_{DCmin} + N(V_O + V_F)} \right)^2 \left( \frac{f}{2P_{INmax}} \right) \quad (1)$$

We will use 650V MOSFET so the reflected voltage must be around 100V. If we adopt 110V, we can easily calculate the primary to secondary turn ratio using equation

$$N = \frac{N_P}{N_S} = \frac{V_R}{V_O + V_F} \quad (2)$$

In order to prevent the saturation of the core we must obtain inductance factor to be  $A_L = 200nH/T^2$ . For that value we need air-gap in center leg to be  $g = 0.2$  mm. Because the grinding of the center leg is problematic, we will place non-conducting spacers between core halves. The spacer thickness should be half the value used for the center leg gap. In our case we have  $s = 0.1$  mm.

Now we can calculate the number of primary turns using equation

$$N_P = \sqrt{\frac{L_P}{A_L}} \quad (3)$$

Skin depth in mm is given by the following equation

$$\delta = \frac{76}{\sqrt{f}} \quad (4)$$

As a result the wire diameter must be

$$d < 2\delta \quad (5)$$

At the end the number of secondary turns is given by

$$N_S = \frac{N_P}{N} \quad (6)$$

The results are given in Table II.

TABLE II  
BASIC PARAMETERS

		---	Typ	---	
Max. primary inductance	$L_P$	1.2		mH	
Number of prim. turns	$N_P$	78			
Number of sec. turns	$N_S$	9			
Primary peak current	$I_{PPEAK}$	0.5		A	
Primary RMS current	$I_{PRMS}$	0.2		A	
Secondary RMS current	$I_{SRMS}$	1.6		A	
Max. wire diameter	d	0.42		mm	

Now it is time to wind the transformer. We will use a single 0.2 mm enamelled copper wire for the primary and a triple-insulated bundle of 7 twisted wires with 0.3mm enamelled copper wires for the secondary, in order to minimize copper losses taking into account the skin effect and to fulfil demands for reinforced insulation. Using triple-insulated wire instead of layers of tape barrier between the windings we will minimize the leakage inductance which must not exceed 4% of the primary inductance. Also to minimize the leakage inductance the bias and secondary winding are interleaved between the primary halves. The bias winding is designed not only to supply the current for the controller, but also to detect the valley of the resonant ringing

Knowing specific core losses we can now calculate the loss in magnetic component (Table III). Total transformer power loss is 0.354 W. Since the thermal resistance of chosen transformer is 46°C/W the temperature rise will be approximately 17°C above ambient temperature. Satisfied with the results, we will keep the chosen core geometry.

TABLE III  
TRANSFORMER LOSSES

		---	Typ	---	
Core effect. volume	$V_E$	1.49		cm <sup>3</sup>	
Specific core losses	$P_V$	0.15		W/cm <sup>3</sup>	
Core loss	$P_{CORE}$	223		mW	
Primary resistance	$R_{PRI}$	1.73		$\Omega$	
Primary loss	$P_{PRI}$	86		mW	
Secondary resistance	$R_{SEC}$	14		m $\Omega$	
Secondary loss	$P_{SEC}$	46		mW	

For the output diode we will use Schottky diode despite much higher junction capacitance which can cause ringing with parasitic inductance. The rated voltage of used diode is 100V in order to accommodate the sum of the reflected

primary voltage and the output voltage including margin for the leakage inductance spike

#### IV. REALIZATION

AC/DC converter was built on two layer FR-4 substrate with  $35\mu\text{m}$  copper with a footprint of  $80 \times 40\text{mm}$ . The transformer is wound on through-hole coil former according to the calculations. Output voltage is further filtered out by the added LC filter. All electrolytic capacitors are low ESR organic polymer electrolytic capacitors.

Furthermore we have measured efficiency at various loads at various input voltages. For practical reasons we have used input voltages from 250 to  $350\text{V}_{\text{DC}}$ . The results are given in Fig.2. The efficiency is over 84% at full load.

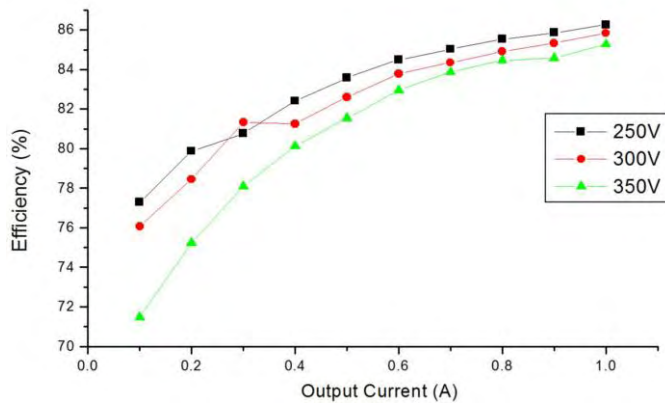


Fig.2. Efficiency as a function of load

The drain waveform of the primary MOSFET in green-mode at 0.1A output load is given in Fig.32. With raising load the converter goes to frequency fold-back mode (Fig.4). At maximum load (1A) the converter operates in quasi-resonant mode (Fig.5). All three waveforms are captured at 300V DC input.

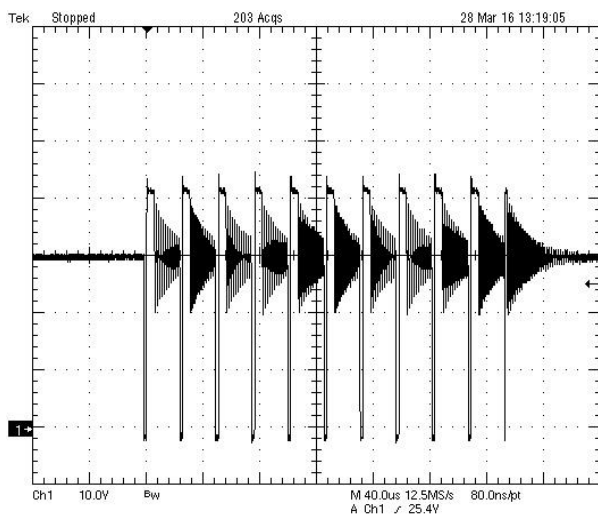


Fig.3. Drain voltage in green-mode (burst)

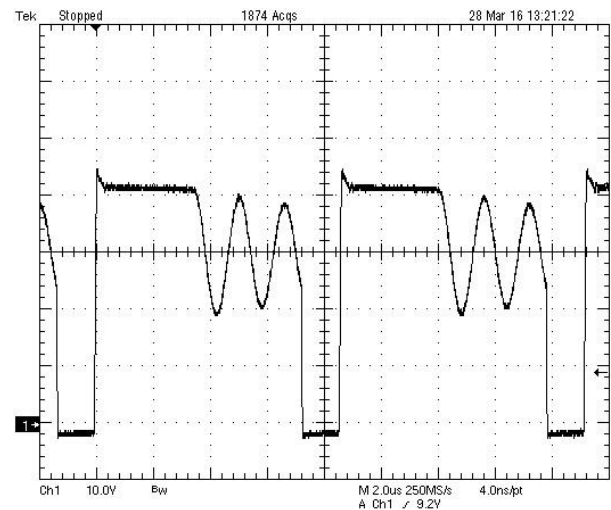


Fig.4. Drain voltage in frequency fold-back mode

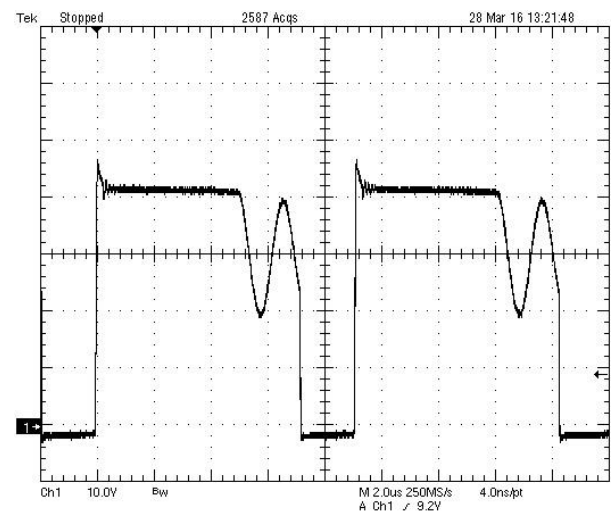


Fig.5. Drain voltage in quasi-resonant mode

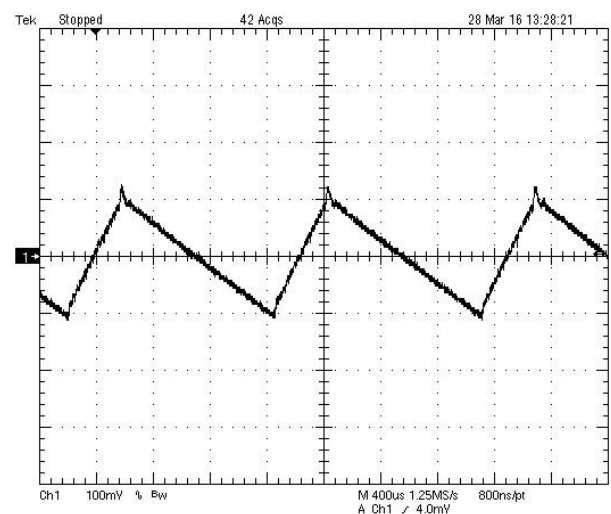


Fig.6. Output ripple in green-mode



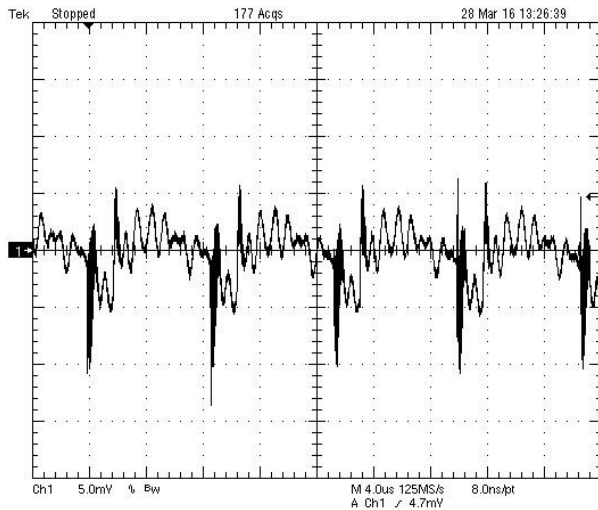


Fig.7. Output ripple in quasi-resonant mode

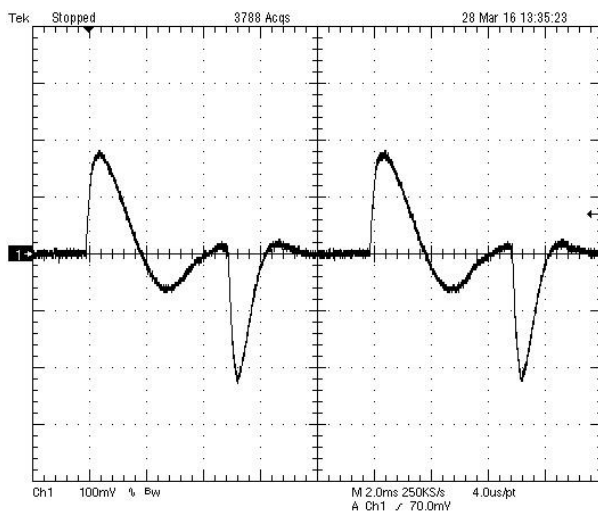


Fig.8. Output voltage load transient response

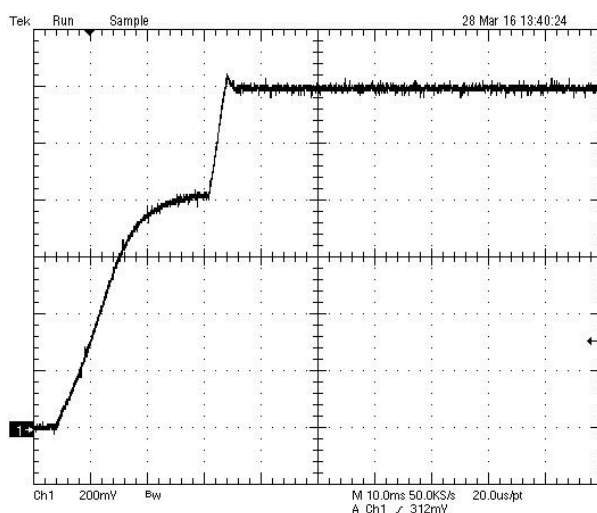


Fig.9. Output voltage rise into 1A load

The output voltage ripple in green-mode is around  $230\text{mV}_{pp}$  (Fig.6) with triangular shape. In quasi-resonant mode output ripple has more complex shape (Fig.7), but much lower level. The basic ripple is about  $10\text{mV}_{pp}$  and the spikes are about  $20\text{mV}_{pp}$ .

Output voltage load transient response was measured with load current changed between  $0.2\text{A}$  and  $0.8\text{A}$  at  $5\text{ms}$  period and at  $0.5\text{A}/\mu\text{s}$ . Fig.8 shows the AC coupled output voltage under this load transient condition. As we can see over/undershoot is under  $200\text{mV}$ .

Output voltage rise (Fig.9) was measured into the electronic load with  $1\text{A}$  load at  $300\text{V}$  input voltage. The rise is monotonic with small  $400\text{mV}$  overshoot.

The picture of the converter prototype is given in Fig.10.



Fig.10. Converter prototype

## V. CONCLUSION

In this paper the design and analysis of  $12\text{W}$  QR flyback converter are presented. The prototype was built and tested. The results verified that the efficiency between  $50\%$  and  $100\%$  of load is over  $80\%$ . With minor changes we can make different versions of this power supply.

## ACKNOWLEDGEMENT

The work is partially supported by the Serbian Ministry of Education and Science (Project III-45016).

## REFERENCES

- [1] Basso, Christophe, "The Dark Side of Flyback Converters" <http://cbasso.pagesperso-orange.fr>
- [2] TDK, "Ferrites and accessories E20/10/6", Data Sheet, [www.tdk.eu](http://www.tdk.eu)
- [3] Power Integrations, "Flyback Transformer Construction Guide", [www.power.com](http://www.power.com)
- [4] Texas Instruments, "UCC28600 8-pin Quasi-Resonant Flyback Green Mode Controller", [www.ti.com](http://www.ti.com)



# Problems in Assessing the Reliability of Electronic Components and Systems by Reliability Testing

Toncho Papanchev<sup>1</sup>, Anton Georgiev<sup>2</sup> and Nikolay Nikolov<sup>3</sup>

**Abstract** – This paper concerns reliability evaluation of electronic devices due reliability tests. Basic groups of tests are described. Some problems in their implementation are observed. A development in this regard is presented and guidelines for further work are outlined.

**Keywords** – Electronic Devices, Reliability, Reliability tests.

## I. INTRODUCTION

Manufacture of electronic devices has two aspects - mining and processing of raw materials on the one hand and on the other - creating a final consumer product. By increasing the reliability as an essential part of quality, the economic efficiency of the devices dramatically increases, thus compensating the lower efficiency of the commodities' conversion process.

Striving to create a competitive production imposes the extension of the classical concept of quality production, which comes down to answer questions related to the processes that follow producing of devices, such as:

- Do devices fail in the early stage of operation?
- Is the *Burn-In* process sufficient?
- Is the failure rate value acceptable in the normal operation lifecycle?
- What adjustments in the design, manufacture, operation and maintenance can lead to higher reliability?

They can be represented three levels of implementation of reliability as a concept [1]:

1. Reliability prediction - at the design stage. Reliability is assessed: - for the design - through parametrical failures; - for the elements - through sudden failure; - for the devices - through failures from defects, caused by production processes.
2. Technical (nominal) reliability – by using the data from testing of first batch samples ("zero" series) under laboratory conditions imitating the real operational once.
3. Operational reliability - according the data from real operation of mass production devices.

For electronic devices such levels are in direct correlation with the main stages of their life-cycle - design, manufacture and operation. Figure 1 shows the sequence of stages and of their constituent phases.

Creating first test samples and test series gives possibility to carry out reliability tests in order to obtain experimental data and more accurately determination of reliability parameters.

<sup>1</sup>Toncho Papanchev is with the Faculty of Automation and Computing at Technical University of Varna, 1 Studentska str., Varna 9010, Bulgaria, E-mail: t.papanchev@tu-varna.bg

<sup>2</sup>Anton Georgiev is with the Faculty of Automation and Computing at Technical University of Varna, 1 Studentska str., Varna 9010, Bulgaria, E-mail: gorgiev\_an@tu-varna.bg

<sup>3</sup>Nikolay Nikolov is with the Faculty of Automation and Computing at Technical University of Varna, 1 Studentska str., Varna 9010, Bulgaria, E-mail: nikolay.nikolov@tu-varna.bg

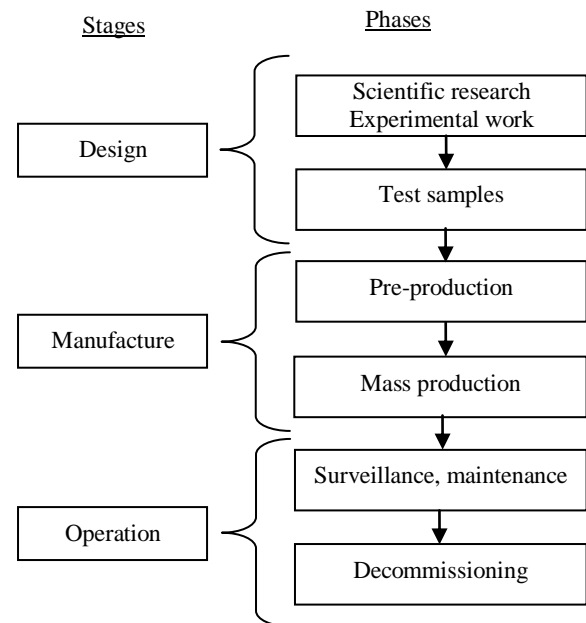


Fig.1. Life-cycle of electronic devices

The manufacturing stage consists of two phases. The first phase, called "pre-production" represents technological preparation and optimization of manufacturing processes, and also the equipment selection. Limited test series of products are produces. During the second phase, "mass production", the devices are produced in their final form, ready for field operation. Large amounts of uniform production are manufactured. It could mean, in present terms, series of several hundred to several million units of production.

Operational stage is the period of real operation of the devices and comprises the largest part of their life-cycle. During this stage the devices reveal their features, including in terms of reliability.

The reliability of electronic products throughout their lifetime is kept at the desired level by technical maintenance, including inspection, failure localization and repair, replacement parts and others. The high cost of maintenance poses questions about the optimality of the implemented procedures. The optimization of maintenance is subject to research by many scientists as Ushakov [2], Benbow and Broom [3], Moubray [4].

After a certain moment a physical and technological obsolescence of products happens. Nowadays, because of accelerated products' development and shortening the duration of the first two stages of the life-cycle the technological obsolescence often overtakes the onset of physical aging.

The assessment of reliability of the products is a procedure of successive stage-followed adjustment of reliability parameters' estimates with regard to construction, technology, manufacturing, operational algorithms, operating rules and conditions, maintenance and repair procedures, and criteria for the

occurrence of failures and boundary states.

In recent years, manufacturers achieve a significant improvement of the reliability of manufactured devices. Furthermore, a considerable part of failures are due to reasons not directly related to components and circuits, such as errors or inaccuracies in the manufacturing process, software problems, unsuitable design choices, external overstresses. This sets new issues related to precise determination of the causes of failures and difficulties associated with assessing the reliability parameters of electronic devices in terms of lack of registered failures or limited number of them.

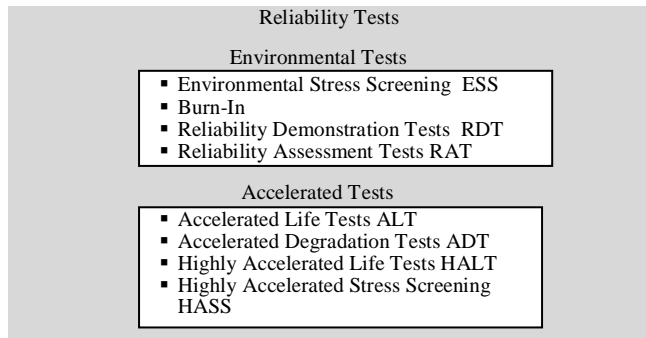


Fig.2. Reliability tests summary

## II. RELIABILITY TESTING

### A. Summary

Figure 2 presents classification of reliability tests, prepared under the criterion "conditions of the tests." Accordingly, tests can be divided into three groups:

- *Environmental tests* are held in conditions similar or identical to the most common in the real working conditions; they could be performed within normal operating conditions, but close to the border values of the main factors determining normal working conditions;
- *Accelerated tests* are held in conditions beyond the specified normal operating conditions.

In economic terms it is crucial the time needed to collect the necessary information that provides the data for assessing the products' reliability. This is the reason further in the article to be examined in details the third group of the above classification - accelerated tests.

The fast development of electronics in the last decade has imposed a shortening of the time from development to the realization of electronic products. Aiming for evaluating the reliability in short time raises the need to develop sufficiently precise and convenient for practical application test methods, which to ensure for a short time correct assessment of the main indicators of reliability – mean time to failure *MTTF*, failure rate  $\lambda$ , availability  $R(t)$ , mean time between failure *MTBF*, etc. Under proper planning and precise execution, the accelerated tests provide reliable information to determine the potential mechanisms of occurrence of failures and modeling of the distribution law of failures of electronic products. The application of reliability testing of electronic systems and components in the initial phases of their lives is crucial to meeting the requirements for reliability in the later phases of production. Elsayed [5] describes the process *TAFT* for

elimination of design defects in the initial samples by testing, analyzing data collected, removal of registered weaknesses and re-testing. This process is generally represented by the term "Reliability growth".

In many cases, the accelerated tests are the only applicable approach to assessing reliability. They can be used for:

- a) reliability assessment electronic products and the quality of materials used;
- b) identifying types of failures, their localization in the devices under test (*DUT*) and taking actions to prevent them;
- c) electro-thermal training to remove items with hidden defects (*Burn-In*);
- d) predicting the performance of the device under real operational conditions and its behavior under extreme load stresses and external impacts.

### B. Environmental tests

When stress values are close to the environmental conditions, such tests are named environmental stress screening *ESS*. In *ESS*, all the products are tested with the aim to speed up and find out the hidden defects as earlier as possible in the production cycle, making their removal cost-effectively. Such tests could be combined with systems for measuring and automation control [6], to extend the benefits beyond the usual objectives of each approach.

A special kind of reliability tests represents technological training of electronic products. In the literature it is known as electrothermal training or "*Burn-In*". *MIL-STD-883* [7], has defined it as "... test performed to monitor or separation of the devices having internal defects or defects caused by deviations in the production process that led to failures dependent on the duration of operation and load." *Burn-In* provides an interesting opportunity in terms of evaluating the reliability of manufactured electronic devices due to the application a stress (thermal or electrical) on all products. This makes it possible to trace the impact on the devices not only as an originated failure or lack of it, but as changes in certain controlled informative indicators. The use of *Burn-in* tests in this type of classification would save costs for additional tests. Difficulties could arise due to lower stress levels and shorter duration of *Burn-In* tests, which limits the range of dispersion of informative indicators for classification, respectively accuracy of the final reliability classification results.

Environmental tests are also used to demonstrate the compliance of the actual assessments of reliability parameters of the devices with the specified once, regarding the requirements - reliability demonstration tests *RDT*. An example of such an approach is represented by Yadav [8]. *RDT* are usually held within normal working conditions of the devices. Similar tests are applied due delivery-and-acceptance procedures – reliability acceptance tests *RAT*. The results of these tests provide information that is used for decision about acceptance or rejection of the series of products of which was formed the tested sample.

### C. Accelerated reliability tests

Highly accelerated life tests *HALT* can be performed with two different goals - to define the limits of the operating conditions of tested products, and to express and analyze all

possible failures in products. *HALT* in electronic products is carried out by applying stress factors temperature and vibrations and is used most frequently for analyzing problems in the functioning of the circuit boards. Subject to certain restrictions, such tests can be used in modeling of processes in power electrical engineering, such as lightning protection [9]. The same stress factors, but with a smaller stress values are applied in highly accelerated stress screening *HASS*, to confirm the ability of the products to function properly under cyclic loading. The main difference between the two tests is the object of study - *HALT* explore usually test samples and series prior mass production, while during *HASS* are tested samples of the mass production of the devices. *HASS* provide opportunities to 'catch' changes in the quality of the production process.

Devices, in which the aging process is manifested at an earlier stage, are tested by accelerated aging tests *ADT* [10].

The application of accelerated tests is studied and systematized by many scientists as Nelson, in his remarkable work "Accelerated Testing. Statistical Models, Test Plans, and Data Analysis" [11], Ushakov [2] and many others. Many publications and methodologies present the application of accelerated tests in different versions aimed at achieving a sufficiently precise estimates of indicators of reliability [12][13][7][14]. Escobar and Meeker, in an extended report, review many types of accelerated tests in order "to outline some of the main ideas behind the accelerated tests" and to formulate "proposals for potential contributions to which statisticians could help to the development the methods and models of accelerated tests" [15]. In another article Meeker draws attention to the problems associated with the selection, planning and implementation of accelerated tests [16]. The great variety of techniques gives engineers multiple reliability analysis tools through accelerated reliability tests, and their task is to choose the most appropriate in accordance with applied tests and available preliminary information.

#### *D. Test planning*

When tests are performed for the first time on certain products, planning is done on the basis of known theories of optimal, best plan and plan with restrictions (compromises) [11]. Prior knowledge, collected due tests on products, similar in functionality or composition, is used to initially determine the range of change in stress factors, the duration of the tests and sample size. However, there is enough uncertainty that makes plans tentative, and may lead to surpluses or shortages of required information. The possibility of further optimization of tests, through dynamic assessment of the results, is not sufficient studied. The results of such optimization can be expressed in additional spending cuts of time and money while keeping the accuracy of the final results within acceptable limits. *IEC 61649* [17] presents set of requirements for conducting accelerated tests. Many developments in the planning and optimization of accelerated tests are presented in the scientific literature [2][17]. Test parameters usually are determined in advance on the basis of preliminary studies of conducted tests on similar devices, normative documents and optimization procedures. An important characteristic of the tests is a manner of recording the moments of occurrence of failures – exactly or roughly. If

there is continuous control, by man or machine, on all tested devices, any failure is recorded at the exact moment of occurrence and there are no errors in assessing the parameters of reliability, caused by uncertainty of timing of failures. The registration of exact times of failure occurrences is difficult to realize, sometimes impossible, especially in data collection during the operation. The information about failures usually does not contain the precise moment and data can be of the following type - censored data (left or right), grouped (interval) data or reported only at the beginning and end of the test duration. It often happens to be handled with different data types in a research. While censored data of different types and exact data are object of a number of studies, the interval data is less common in the science literature.

Recently the accelerated tests are associated with tasks for evaluation of indicators describing various aspects of the reliability of devices. This determines the variety of combinations of types and combinations of stresses applied, ways for stress application and other conditions. The scientific literature describes accelerated tests, applied on electronic products, with different plans and mathematical formalism [18][19][20].

### **III. PROBLEMS OF ASSESSING THE RELIABILITY OF ELECTRONIC DEVICES OVER THE DIFFERENT STAGES OF THEIR LIFE CYCLE**

The products' reliability depends on the decisions taken during the period of design and production. Therefore, it can be concluded that reliability is the property of the product, which is modeled and built into it and requires the investment of considerable resources and time. So considered, the process of „embedding" of reliability outlines many problems. Below are presented only two of them and the ensuing conflicts:

First problem - "materialization" of the term "sufficient reliability" for each particular product.

Second problem - decision of how to achieve the level of "sufficient reliability" in every product manufactured.

First conflict - Time: Growing competition shortens the overall length of all life stages of products: design - testing - production - operation - technological obsolescence - design of completely new product or release of devices with enhanced options. This naturally requires cutting the time spent on reliability testing and analysis of the devices.

Second conflict - Price: The wide variety of products with similar, even identical characteristics requires seeking the solutions that ensure low prices of the output product. This is reflected in the reduction of planned spending on research of reliability.

Wrong decisions lead to losses, resulting not only in unforeseen costs for warranty service, but also in reduced sales and revenue from the negative impact of the customers' dissatisfaction. The effect extends more globally; negatively affect the reputation of the manufacturer, putting question mark over its survival in a highly competitive landscape of electronic products. From the customer's perspective and in a "good" case scenario the insufficient reliability means more periods of inactivity and spending more time and money for maintenance. When it comes to manufacturers, this directly affects their current and prospective revenues. In a "worst"



case scenario, however, the consequences can be expressed with the destruction of material facilities, injuries or loss of human lives. Murthy analyzed some aspects of the above cases in a series of publications [21][22][23]. The complex nature of these problems and conflicts does not allow the formulation of a clear decision, and requires operating in an uncertainty in many ways.

#### IV. CONCLUSIONS AND FUTURE WORK

The problems of dynamically optimization of accelerated tests, aiming to increase the efficiency of tests by using ongoing assessment of their informativeness, are the subject of research by the authors of this publication. Algorithms are examined in Burn-In and accelerated tests [24][25]. The following issues are encountered, which are subject to further work:

1. Insufficient ability to determine the indications of classification of devices after *Burn-In*.
2. Significant complexity and the need arises to establish limits for efficient use – cost/effect.
3. It was reported, during tracing of conducting the accelerated tests, a difficulty in establishing a set of indicators defining the right moment for termination, replacement or continuation of a test. It is obtained a divergence of the chosen indicators goodness-of-fit  $r^2$  and error function  $Ej$ , therefore it is necessary work to identify additional markers for decision making.

With regard to tests on highly reliable devices and registration of a little failures and interval data, an approach “Interval Weibayes” was proposed [26]. Further work is focused on combination of effective strategies for reliability-oriented maintenance and systems for continuous monitoring, in order to develop an approach for detection of approaching failure conditions, with the limitation of “false alarms”.

#### ACKNOWLEDGEMENT

This paper presents results obtained during a scientific research performed within a PD11/2016 project “Study on the reliability of a network structured electronic system”, sponsored by Technical University-Varna scientific research fund.

#### REFERENCES

- [1] Гиндев Е., Изчислителна надеждност в радиоелектрониката, Д. изд. Техника, България, 1979
- [2] Ushakov I. A., Handbook of Reliability Engineering, Wiley-Interscience, March 1994, ISBN 0471571733
- [3] Benbow D., W. Broom, The certified Reliability engineer handbook, ASQ Quality Press, 2008, ISBN: 978-0-87389-721-1
- [4] Moubray J., Reliability-Centered Maintenance, 2nd Ed., industrial Press Inc., New York, 1997, ISBN 0-8311-3078-4
- [5] Elsayed E., Overview of Reliability Testing, IEEE Trans. On Reliability, Vol. 61, No:2, June 2012, ISSN 0018-9529
- [6] Vasilev R., Intelligence System for Measuring and Automation Control of Air and Cable Outlets, Conf. proc. of 6-th International Conference Tehnonav 2008, pp. 534-538, Ovidius University Press, ISBN 978-973-614-447-9
- [7] MIL-STD-883E Test Methods Standards, US Department of Defence, USA 1996
- [8] Yadav O., N. Singh, P. S. Goel, Reliability demonstration test planning: a three-dimensional consideration, Reliability Engineering and System Safety, 91 (2006), pp. 882-893, Elsevier, ISSN: 0951-8320
- [9] Vasileva M., N.Velikova, N.Nikolaev, Model Study of Lightning Protection of 110 kV Substation, 14-th IEEEIC, Poland, 2014, pp. 113-115, ISBN 978-1-4799-4661-7
- [10] Vavrilov N., A. Stoyanova, Model of Accelerated Degradation Test for LED Module, International Journal of Engineering Innovation & Research, Vol. 3, Issue 6, 1/2014, pp. 765-768, ISSN: 2277 5668
- [11] Nelson W., Accelerated Testing, Statistical Models, Plans, and Data Analysis, Wiley Inc., 2004, ISBN: 978-0-471-69736-7
- [12] Liu D., H. Leidecker, T. Perry, F. Felt, Accelerating factors in life testing of high-voltage multi-layer ceramic capacitors, CARTS USA, March 21-24.2005, Palm Springs, CA
- [13] Liu X, W. Qiu, Modeling and planning of Step-stress accelerated life tests with independent competing risks, IEEE Trans. on Reliability, Vol.60, No:4, 2011, pp.712-720, ISSN 0018-9529
- [14] MIL-HDBK-781. Reliability Testing for Engineering Development, Qualification, and Production, US Department of Defence, USA
- [15] Escobar L., W. Meeker, A Review of Accelerated Test Models, Statistical Science, Vol. 21, No.4, 2006, pp.552-577, <http://projecteuclid.org/euclid.ss/1177334529>
- [16] Meeker W., G. Sakarakis, A. Gerokostopoulos, More Pitfalls of Accelerated Tests, Journal of Quality Technology, Volume 25, Issue 3 (July 2013), pages 213-222, ISSN 0022-4065
- [17] IEC 61649 Ed. 2.0: Weibull Analysis, UK, 2008
- [18] Park J, S. Bae, Direct Prediction Methods on Lifetime Distribution of Organic Light-emitted Diodes From Accelerated Life Test, IEEE Transactions on Reliability, Vol. 59, No: 1, March 2010, pp. 74-90, ISSN 0018-9529
- [19] Levenbach G. Accelerated life testing of capacitors, IRE Trans. on Reliability and Quality Control, June 1957, pp. 9-20, ISSN 0097-4552
- [20] Chang M., D. Das, P.V. Varde, M. Pecht, Light emitting diodes reliability review, Microelectronics Reliability 52 (2012) 762-782, Elsevier, ISSN: 0026-2714
- [21] Murthy D., M. Rausand, S. Virtanen, Investment in new product reliability, Reliability Engineering and System Safety 94 (2009) , pp. 1593-1600, Elsevier, ISSN 0951-8320
- [22] Murthy D., T. Østera, M. Rausand, Component reliability specification, Reliability Engineering and System Safety 94 (2009) pp. 1609-1617, Elsevier, ISSN 0951-8320
- [23] Murthy D., P.-E. Hagmark, S. Virtanen, Product variety and reliability, Reliability Engineering and System Safety 94 (2009), pp. 1601-1608, Elsevier, ISSN 0951-8320
- [24] Папанчев Т., А. Георгиев, Приложение на методи за разпознаване на образи и анализ на данни в класификацията на електронни елементи по надеждност, Електронно списание „Компютърни науки и комуникации”, брой 1/2012, стр. 23-31, ISSN: 1314-7846, <http://ojs.bfu.bg/index.php/knk>
- [25] Георгиев А., Т. Папанчев, Анализ на надеждността на електронни изделия, основаващ се на данни от ускорени изпитвания, Proceedings of 3th International Scientific Congress, TU-Varna, 04.-06.10.2012, Vol. 2, pp. 128-133, ISBN 978-954-20-0551-3.
- [26] Papanchev T., A. Georgiev, G. Todorinov, „Analysis of Parameter Estimation Methods for Weibull Distribution and Interval Data”, International Journal of Engineering and Innovative Technology (IJEIT), Vol. 3, Issue 9, March 2014, pp. 230-235, ISSN: 2277-3754



# Indices for Reliability Assessment of a Star Structured Complex Electronic System

Nikolay Nikolov<sup>1</sup>, Anton Georgiev<sup>2</sup> and Toncho Papanchev<sup>3</sup>

**Abstract** – This paper concerns reliability assessment of a star structured complex electronic system. The common approaches to reliability indices selection and reliability requirements determination are presented. Some specific reliability indices and basic dependences valid for reliability assessment of a star structured electronic system are presented and described.

**Keywords** – Structural reliability, Reliability indices, Reliability of complex systems.

## I. INTRODUCTION

A complex electronic system (CES) usually comprises a large number of functional units and blocks (i.e. system elements) interconnected in such a way that the system is able to perform a set of required functions including its basic system function as well as all auxiliary functions. The system elements, together with the links between them form the system structure. Often the CES layout is shaped in line with a network topology. For SCADA (Supervisory Control and Data Acquisition) systems most often this is a centralized star topology [5].

The system and its elements have operational modes including normal operating modes, test modes and contingency modes induced by failures, faults or operator errors.

In reliability perspective each particular CES structure determines the system reliability characteristics.

The system reliability assessment is a very important task standing in front of reliability engineers during the system design process and also during the system lifespan [1]. Such assessment is able to be performed on the base of a set of reliability indices selected in such a way that it becomes possible the system reliability characteristics to be fully revealed.

## II. SELECTION OF RELIABILITY INDICES FOR RELIABILITY ASSESSMENT OF CES

Selection of a proper set of reliability indices for reliability assessment of a complex electronic system (CES) is a task strongly depending on the type and the purpose of the system and also on the common functional requirements to it.

<sup>1</sup>Nikolay Nikolov is with the Faculty of Automation and Computing at Technical University of Varna, 1 Studentska str., Varna 9010, Bulgaria, E-mail: nikolay.nikolov@tu-varna.bg

<sup>2</sup>Anton Georgiev is with the Faculty of Automation and Computing at Technical University of Varna, 1 Studentska str., Varna 9010, Bulgaria, E-mail: gorgiev\_an@tu-varna.bg

<sup>3</sup>Toncho Papanchev is with the Faculty of Automation and Computing at Technical University of Varna, 1 Studentska str., Varna 9010, Bulgaria, E-mail: t.papanchev@tu-varna.bg

The indices for reliability assessment of CES can be divided on operational and technical ones depending on the degree of system defragmentation [2]. The operational indices characterize the system from customer point of view, as the technical ones carry some more technological sense.

The technical indices are suitable for to be evaluated statistically. These are necessary for reliability assessment of system elements, i.e. subsystems.

Selecting the reliability indices for reliability assessment of a CES, the rules listed below is advisable to be followed:

- The total number of indices have to be the minimal one;
- The applications of complex reliability indices, presenting combinations of criteria have to be avoided;
- The reliability indices chosen have to carry a simple physical sense;
- The reliability indices chosen have to make possible performance of analytical reliability estimation during system design process;
- The reliability indices chosen have to allow a statistical evaluation of it, based on the reliability tests results or system operational data;
- The reliability indices chosen have to make possible quantitative reliability limits to be set-on.

## III. REQUIREMENTS REGARDING RELIABILITY OF A CES

For to declare reliability requirements technical objects at three different levels have to be distinguished. These are systems, subsystems and components.

### A. Requirements Regarding System Reliability

Determination of reliability requirements to a CES can be achieved, following three approaches. It might be based on:

- Expert advice, design engineer experience and practice;
- Prototype analysis, or statistical data for a CES similar as purpose, structure and/or component base to the current one;
- Reliability level which is optimal for the current system.

The latter approach is applicable only in case when:

- The system function is measurable by usage of the same units of measure as the expenditure of its production;
- Reliable data for components reliability are at the disposal;
- The system structure, system functioning and maintenance process are fully determined.

In this case the system function can be maximized, as follows:

$$F_k(R) = E_k(R) - C_k(R) \quad (1)$$

where  $R$  is a system reliability index depending on the  $k$ -th variant of the system structure  $S_k$  chosen and also on the reliability of elements of the  $i$ -th kind -  $r_i$ , or:

$$R = R(S_k, r_i, k = 1, \dots, m, i = 1, \dots, n) \quad (2)$$

where  $m$  is the number of system structure variants;  $n$  is the number of system components;  $E_k(R)$  is the system function of  $k$ -th system variant as a number – value, valid for reliability level  $R$ ;  $C_k(R)$  are expenses for reliability level equal to  $R$  of  $k$ -th system variant to be ensured [3].

For each fixed  $k$  a solution can be found, based on the condition below:

$$\frac{\partial E_k(R)}{\partial R} = \frac{\partial C_k(R)}{\partial R} \quad (3)$$

After that the variant of the highest absolute value have to be chosen from among of the optimal solutions  $E_k(R)$ .

In cases when the complex effect of the system function is incommensurable with the expenditures, only the former two approaches for system reliability requirements determination are applicable.

#### B. Requirements Regarding Subsystems Reliability

Normally the requirements regarding subsystem reliability are set-up when the system reliability requirements are already at the disposal.

##### 1. Uniformly division approach.

If the system consist of a number of  $N$  elements, which are similar or identical in complexity and structure it is possible a reliability index given ( $R$ ) to be equally divided in accordance with the rule below:

$$E_k = \sqrt[N]{R}, \quad i = 1, \dots, N. \quad (4)$$

In this case MTTF for the  $i$ -th subsystem is approximately equal to:

$$T_i = NT, \quad i = 1, \dots, N, \quad (5)$$

where  $T$  is the average system MTTF given.

##### 2. Proportional division approach.

If  $n_i$  is the elements number of  $i$ -th subsystem, then:

$$T_i = \sqrt[n_i]{R}, \quad i = 1, \dots, N, \quad a_i = n_i \left( \sum_{1 \leq i \leq N} n_i \right)^{-1} \quad (6)$$

If the failure rates of the elements (or prototypes) of the  $j$  – th type  $\lambda_j$  are known, then this approach is able to be modified by substitution in Eq.(6) of  $a_i$  by:

$$a_i = \sum_{1 \leq j \leq M} \lambda_j n_{ij} \left( \sum_{1 \leq i \leq N} \sum_{1 \leq j \leq M} \lambda_j n_{ij} \right)^{-1} \quad (7)$$

##### 3. Optimal division approach.

In case when at the time of general system reliability requirements ( $R$ ) set-up the system structure ( $S$ ) is known as well as the techniques for subsystems reliability improvement, i.e. functions  $R_i(C_i)$ , where  $C_i$  are the resources spent for provision respective subsystem of desired reliability, it becomes possible an optimal reliability requirements division to be found in two cases as follows:

a. At the maximum of the system reliability index, when the total resource  $C^0$  is limited:

$$\max_C \left\{ R(S, R_i(C_i)) \mid \sum_{1 \leq i \leq N} C_i \leq C^0 \right\}, \quad C = (C_1, C_2, \dots, C_N) \quad (8)$$

b. At the minimum of the system maintenance spends, when the reliability index value given  $R^0$  is achieved:

$$\min \left\{ C(S, R_i(C_i)) \mid R^0 \right\} \quad (9)$$

#### C. Requirements Regarding Components Reliability

The reliability requirements for electronic components usually are set-up by experts, or these are based on reliability data obtained during prototypes testing.

### IV. RELIABILITY ASSESSMENT OF A STAR STRUCTURED CES

#### A. The Star Structure as a Monotonous Structure

Some systems, including the star structured systems, have specific characteristics in regard to system reliability. Their reliability characteristics monotonous worsen when the reliability characteristics of system elements are getting worse[7]. The structure of such systems is called monotonous structure.

A simple logical system analysis is necessary to be performed in order to be identified a monotonous structure as such one.

Let introduce a logical random variable, which can take two different values, as follows:

$$x_i = \begin{cases} 1, & \text{if the } i\text{-th element is functioning,} \\ 0, & \text{if the } i\text{-th element is failed.} \end{cases} \quad (10)$$

The probability that the  $i$ -th system element is in a workable condition is determined as a math expectation:

$$p_i = Mx_i \quad (11)$$

An  $n$ -component vector, denoted as

$$X = (x_1, \dots, x_n),$$

would characterize the system condition. When the system structure is fixed, the system condition depends on the condition of its elements ( $n$  in number). In case when a failed element is occurred, the vector takes the form:

$$X_i = (x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n) \quad (12)$$

or it becomes an  $(n-1)$ -component vector. The  $i$ -th component is missing. Generally for all missing components is valid

$$i \in \alpha,$$

and the number of vector components is equal to

$$n - |\alpha|.$$

The system condition can also be described by a logical random function. In this case this will be a structural system function. This takes the value as follows:

$$\varphi(X) = \begin{cases} 1, & \text{if the condition } X \text{ is equal to workable system condition;} \\ 0, & \text{if the condition } X \text{ is equal to non-workable system condition.} \end{cases}$$

The probability about the system to be in a workable condition is determined as a math expectation of the structural system function:

$$h = M\varphi(X) \quad (13)$$

This index can be expressed by reliability indices of system elements, as follows:

$$h(p) = h(p_1, \dots, p_n) \quad (14)$$

The system structure is assessed as a monotonous one in cases when the equations listed below are valid for the current system. These are:

$$\varphi(1) = 1, \text{ where } 1 = (1, 1, \dots, 1); \quad (15)$$

$$\varphi(0) = 0, \text{ where } 0 = (0, 0, \dots, 0); \quad (16)$$

$$\varphi(X) \geq \varphi(Y), \text{ if } X > Y \quad (17)$$

The latter equation aggregates a number of  $n$  conditions of the type

$$x_i \geq y_i, \text{ for } i = 1, \dots, n,$$

and at least one of it have to be strictly fulfilled.

Taking into consideration the arguments expressed so far it becomes clear that a star structured systems have to be assessed as a monotonous one. This conclusion gives the opportunity a logical structural reliability analysis to be performed in regard to a star structured CES.

The reliability function of such system can be presented as:

$$h(p) = p_i M\varphi(X_i, x_i = 1) + q_i M\varphi(X_i, x_i = 0) \quad (18)$$

where

$M\varphi(X_i, x_i = 1)$  is the probability the system to work proper under condition that the  $i$ -th element is absolutely reliable.

$M\varphi(X_i, x_i = 0)$  is the probability of the same, but under condition that the  $i$ -th element is definitely failed.

The reliability system function can also be similarly expanded in regard to two system elements  $i$ -th and  $j$ -th

ones[6]. In this case the system reliability function takes the form as follows:

$$\begin{aligned} h(p) = & p_i p_j M\varphi(X_{ij}, x_i = 1, x_j = 1) + \\ & + p_i q_j M\varphi(X_{ij}, x_i = 1, x_j = 0) + \\ & + q_i p_j M\varphi(X_{ij}, x_i = 0, x_j = 1) + \\ & + q_i q_j M\varphi(X_{ij}, x_i = 0, x_j = 0) \end{aligned} \quad (19)$$

The latter equation gives the opportunity to assess the structural reliability of a star structured CES by taking into consideration the peculiarity of the star topology, i.e. non-equality of elements positioned at different hierarchic levels of this structure. It becomes obligatory the system reliability function to be expanded in regard to these system elements, which stand at the structural hierarchic levels higher than the lowest one, or by the other words which are not peripheral system elements. Their influence and impact on the system reliability can be fatal, because a failure of each of it can cause a failure of a system branch or a total system failure (when the system element at the highest level failed).

#### *B. The Star Structured CES as a System with an Additive Factor of Effectiveness*

The structure of some specific electronic systems (star structured systems are also included) consists of functional redundancy [4]. This makes the system able to functioning even in case when one or some partial failures are been occurred. Then the system continues to work with decreased quality and efficiency of functioning, but it is not totally failed.

For qualitative assessment of functioning of such systems it is advisable a quantitative index to be introduced, i.e. quality of system functioning. This takes into consideration the influence and also the impact of partial failures on system functioning. The effectiveness of system functioning is a quantitative characteristic of quality and quantity of work performed by the system.

Star structured CES are systems which are characterized by a relatively simple effectiveness factor. Each peripheral element of such system brings its separate and independent contribution to the effectiveness of the entire system. These systems are known also as systems with an additive effectiveness factor. Such behavior is typical for most of the SCADA systems.

If the contribution of the  $i$ -th element to the system effectiveness is  $\varphi_i$ , then the system effectiveness of a system intended for short term of operation at a time  $t$ , would be:

$$E = \sum_{1 \leq i \leq n} \varphi_i r_i(t) \quad (20)$$

where  $r_i(t)$  is the probability the  $i$ -th element to be in a workable condition at the moment of  $t$ .

Consider a CES which is built up in line with a centralized star topology [8]. The system structure also can be described as a star within a star. This is shown on Fig.1. The system is

spread over two sites of service. The system structure consists of a number of  $n + 1$  elements and three hierarchic levels. One of the system elements is positioned at the I-st (the highest) hierarchic level, two elements are at the II-nd level and a number of  $n - 2$  peripheral (end) elements are positioned at the III-th (the lowest) level. A number of  $m - 2$  of it are installed at the first service site and the rest of it (a number of  $n - m$ ), are respectively installed at the second site of service.

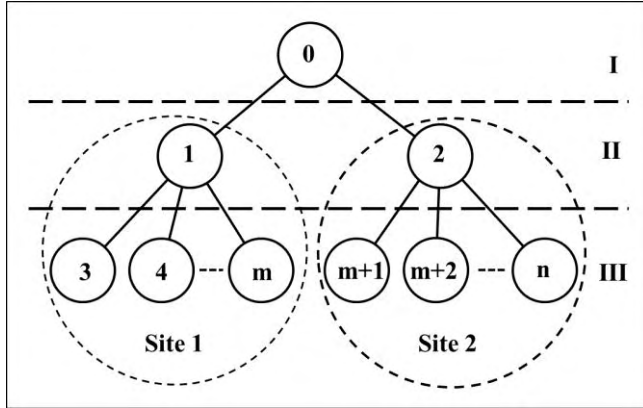


Fig.1. Star structured CES topology

Obviously the peripheral elements would be able to function effective only in case when the elements positioned at the I-st and II-nd level are in a workable condition. If  $K_i$  is availability of the  $i$ -th element, then for the system effectiveness is valid:

$$E = K_0 \left( K_1 \sum_{i=3}^m K_i \varphi_i + K_2 \sum_{i=m+1}^n K_i \varphi_i \right). \quad (21)$$

In case when all peripheral elements at a site of service are similar, or identical and bring an equal contribution to the system effectiveness (which is typical for most of the SCADA systems), the system effectiveness can be presented as:

$$E = K_0 \left[ K_1 (m-2) \varphi_1 \sum_{i=3}^m K_i + K_2 (n-m) \varphi_2 \sum_{i=m+1}^n K_i \right] \quad (22)$$

Consider the same system but intended for a long term operation. Let the contribution of the  $i$ -th element to the system effectiveness is  $\varphi(t)$ . In case of a failure at the moment  $t \leq t_i \leq t_0$ , for the system effectiveness is valid:

$$E(t, t+t_0) = \sum_{i=1}^n \left[ r_i(t, t+t_0) \varphi_{0i} + \int_t^{t+t_0} f_i(x_i) \varphi_i(x_i) dx_i \right], \quad (23)$$

where  $\varphi_{0i}$  is the contribution of  $i$ -th element to the entire system effectiveness if it was in a workable condition during the time interval  $(t, t+t_0)$ .

Now it is possible to determine the effectiveness of the SCADA system already described. Let it was intended for a long term operation. The failure rate of the  $i$ -th system element is denoted as  $\lambda_i$ . The reliability function of each element allows an exponential distribution. In this case the

average system effectiveness at a random moment of time  $t$ , can be determined as:

$$E(t) = e^{-\lambda_0 t} \left[ e^{-\lambda_1 t} \sum_{i=3}^m \varphi_i e^{-\lambda_i t} + e^{-\lambda_2 t} \sum_{i=m+1}^n \varphi_i e^{-\lambda_i t} \right]. \quad (24)$$

The equation above gives the opportunity to determine the system effectiveness by the specific failure rate and also by the individual contribution factor of each system element. The latter depends not only by the element function but also by the characteristics of the object served.

## V. CONCLUSION

The main problem in reliability assessment of a star structured CES appears to be the evaluation of the contribution of each peripheral element to the entire system effectiveness, even in cases when this contribution is similar or identical for all peripheral elements or only for these installed at the same site of service. These might be estimated upon an expert advice for each specific CES application. The rest reliability indices like the elements availability and the elements failure rate can be evaluated using data obtained by testing of prototypes, or might be estimated upon data for similar or identical elements at disposal. It is also possible for this purpose to be used data obtained during operation of the same or similar elements for long enough time. Based on this it becomes possible to estimate system effectiveness for a star structured CES for each specific application.

## ACKNOWLEDGEMENT

This paper presents results obtained during a scientific research performed within the PD11 / 2016 project "Study on the reliability of a network structured electronic system", sponsored by Technical University-Varna scientific research fund.

## REFERENCES

- [1] Georgiev A.Sl. "New concept for increase in operative reliability of electronic systems". Monography, Aquaprint Ltd., ISBN 978-954-92824-6-7, Varna, 2013. (Bulgarian)
- [2] Rausand M. "System reliability theory". John Wiley & Sons, Inc., ISBN 0-471-47133-X, New Jersey, 2004.
- [3] Gnedenko B., I.Ushakov, "Probabilistic reliability engineering", John Wiley & Sons, Inc., USA, 1995, ISBN-0-471-30502-2
- [4] Svensson S., D.Spira. "Framework to enable reliability analysis of SCADA solutions". Elforsk. 2012.
- [5] "SCADA Systems". White paper. Motorola Inc. 2007
- [6] Lloyd E., Lederman, W. "Handbook of applicable mathematics". Volume 4. New York, 1997.
- [7] Ushakov I.A., "Reliability of technical systems". Manual, Publishing "Radio i sviaz", Moscow, RF, 1986, BBK 30.14, H 17, UDK 62-192: 52(031). (Russian)
- [8] Georgiev, A., Nikolov, N. "Structural reliability analysis of a star structured complex electronic system". E-magazine Computer science and communications, No 4, 2015 pp.44 – 51, ISSN: 1314 – 7846.



# Using Wireless Interfaces in a Smart Home Model

Valentina Rankovska <sup>1</sup>

**Abstract** – In this paper the application of wireless interfaces in embedded systems for monitoring and control is examined. A model of smart house is synthesized, applying wireless interfaces, designed for educational purposes. The core of the system is Arduino board with a possibility of remote Android based control.

**Keywords** – Embedded Systems, Wireless Interfaces, Smart Home Model, Arduino.

## I. INTRODUCTION

The architecture and the features of embedded systems develop continuously - they are becoming more diverse in composition and functions, more flexible, adaptable and reliable.

Their development through the years can be summarized to the following three phases:

- In the early days of occurrence and development of the microprocessor systems they are single- or multiple-board systems, based on microprocessors, memories and peripheral units - timers, parallel, synchronous and asynchronous serial interfaces, etc. as separate ICs.
- The occurrence of microcontrollers in the early 70s of the last century leads to increased reliability, speed, flexibility, miniaturization.
- Nowadays, together with the growing variety of microcontrollers with enhanced features and parameters, on one hand the number and type of the peripheral intelligent programmable devices grows increasingly - sensor units, actuators, etc., and on the other control is not limited only to the use of microcontrollers as control devices.

Wireless interfaces allow building networks consisting of heterogeneous in nature components. In the process of controlling the intelligent programmable modules, collection and processing of information from them and other activities, easily can be incorporated a variety of "standard" mobile devices such as smartphones, tablets, laptops and more.

In general the methods and tools for monitoring, control, collection and processing of data using wireless interfaces evolve rapidly and are used in all spheres of life: building automation, industrial applications, automotive and medical industry and many others.

Wireless communication offers a flexible and reliable approach to object control, obtaining information on its parameters and state, receiving data and organize them in a database, etc.

There are various forms of wireless communication based

on different frequency bands, modulation methods and protocols, which are briefly presented in Part II.

A smart home model implementing wireless interfaces is presented in part III.

The model is intended for educational purposes in Microprocessor Circuits and Embedded Systems for the Bachelor degree students in Electronics in the Technical University of Gabrovo.

## II. A SHORT SURVEY ON WIRELESS INTERFACES

As wireless interfaces provide a range of advantages in data transfer in modern embedded systems, more and more devices that have the ability to connect wirelessly with others are produced. That is why the terms "Internet of Things" and "Wireless Connectivity of Things" are presently widely spread. [1], [2], [3], [4]

Using the infrared range for data transfer is implemented long ago, for instance as a remote control of various devices. But due to a number of advantages it gained a rapid development and is widely used to connect various devices such as PC and peripherals.

Advantages:

- Infrared signals are easily generated and identified;
- As the generated signals are out of the limits of the visible spectrum, it is easy to apply optical filters, excluding totally the visible light and therefore the interference;
- High level of channel safety;
- Low price of the hardware and also lack of law regulation.

Drawbacks:

- Need of direct visibility and therefore ability for short distance connection. That limits its application in implementing larger networks;
- The data transfer is peer to peer.

*Bluetooth* interface is especially made to replace the cable connections at office and home appliances. Frequencies in the 2,4 - 5 GHz range are used, a frequency band initially reserved for industrial scientific and medical (ISM) purposes. Nowadays it is widely used for data transfer in local wireless networks. The Bluetooth specification is maintained by Bluetooth Special Interest Group (SIG), founded in 1998.

Main features:

- Low power radio connection – typical power consumed about 1 mW;
- Typical range – 10 m;
- Data baud rate, initially 1 Mb/s, but for instance Bluetooth 2.0 - to 3 Mb/s at distances to 100 m;
- Simultaneous connection with up to 8 devices.

The advantages of the Bluetooth technology are reduced size of the equipment, simple usage, safety of the data transferred, and good maintenance of the standard.

<sup>1</sup>Valentina Rankovska is with the Faculty of Electrical Engineering and Electronics at Technical University of Gabrovo, 4 H. Dimitar str., Gabrovo 5300, Bulgaria, E-mail: rankovska@tugab.bg.

Some disadvantages are the comparatively high power consumption and impossibility to organize complex configuration networks.

Unlike other wireless technologies, where the aim is to provide high data transfer rate, long distances, etc., *ZigBee* (IEEE 802.15.4) [3] is created with opposite requirements for small range, low price, low power consumption, low data transfer range and low size of the equipment used, and also low requirements to the software. It uses the standard „Low-Rate Wireless Personal Area Networks - LR-WPAN - IEEE 802.15.4“. Like Bluetooth, the radio band ISM is used.

It is suitable especially for home automation systems and some measurement and control systems together with using not expensive microcontrollers. The data transfer rate is comparatively low and the power consumption - minimal. Main application areas are as follows: receiving data from moving or rotating parts of conveyers, robots, etc., industrial systems for monitoring and control, wireless connection with sensors, tracing the route of the movement and location of property and equipment, security systems, etc.

The ZigBee standard, like Bluetooth, uses the baud transfer range of 2,4 GHz. The largest baud rate is 250 kbit/s. Although it provides typical data transfer distance 10 m, there are no requirements to the transmitter power. The most common transmitters are 1 mW (to 10 m), 10 mW for distance to 80 m indoors and to 1 km at line of sight. Increasing the distance can be achieved using antennas with a special design.

Minimized power consumed is a result of the fact that the slave devices are in idle state most of the time. They get active for short periods of time only to confirm their presence in the network.

Implementing ZigBee networks with various topologies could be configured – star, tree, mesh.

Connecting the components of an embedded system to Internet, including via Wi-Fi, results in additional advantages as the user can access the network components almost from everywhere [5]. Furthermore no special knowledge and skills are required to configure the network.

The main advantages of the Wi-Fi are the following:

- No licensed radio range is used and no law regulation for individual users;
- Quickly developing area, which allows usage of more and more devices communicating via such interfaces;
- Relatively low price;
- Roaming is possible, which allows the user with mobile devices to move from one access point to another;
- There are different levels of data encryption to protect traffic, etc.

A brief comparison of wireless interfaces for embedded systems according to different criteria is given in Table 1 and Table 2 [1], [3].

TABLE I

	Wi-Fi	Bluetooth	Bluetooth LE	ZigBee	Sub GHz
Internet-Connected Device	Best	OK	OK	Poor	Poor
Smartphone Accessory	OK	Best	OK	Poor	Poor
Range	OK	OK	OK	OK	Best
Low Power Consumption	OK	OK	Best	OK	Best
High Data Rate	Best	OK	Poor	Poor	Poor
Interoperability Between Vendors	Best	Best	OK	Poor	Poor

### III. THE SMART HOME MODEL

The smart home model is a suitable complex example to study various components of embedded systems - control devices and actuators, sensor modules, interfaces for data exchange and in particular the interface "man-machine". It also provides wide opportunities for implementation and testing of a number of modern wireless interfaces.

An example of such a system is shown in Fig. 1 [6].



Fig. 1. A smart home example

The main criteria, defining a building as “smart”, are the following [6]:

- Input system, which receives information through a corresponding receiver. This can be done in four ways: by sensors (in real time), internally stored and recovered data, manually entered (programming and reprogramming) by the users and available online (Internet).
- Processing and analysis of information - performed by control system;
- Output system which responds to input data in the form of some actions;
- Response time requirements;
- Self-teaching ability.

The following stages can be identified in the development and services of the smart home:

1. First stage – the components are autonomous devices which communicate with other devices in home. These first devices are: bought and installed (and a part of) a home security system; a home cinema system or Do It Yourself (DIY). Using a variety of communication technologies, they usually do not communicate with each other directly but with their controllers.

**TABLE II. A COMPARISON OF THREE LEADING WIRELESS TECHNOLOGIES**

Parameter, m	Bluetooth/IEEE 802.15.1	ZigBee/ IEEE 802.15.4	Wi-Fi/ IEEE 802.11
Distance, m	~ 10 (50-100)	10	~100
Baud rate, Mbit/s	0.723	0.250	1-2 to 54
Max number of devices in the network	8	245	unlimited
Power consumption, mW	10	1	50
Operational time supplied with 2 batteries AA	-	6 months in standby mode	-
Price/complexity, conditioned units	10	1	20
Main purpose	A connection PC - peripheral devices	Wireless sensor networks	Wireless Ethernet extention

2. The second stage is the Internet connectivity - once connected with Internet the devices could be controlled via a smartphone or other mobile devices.

3. In the third stage these smart devices can communicate and exchange information between each other, without necessity of a human intervention. [7]

The wireless interfaces begin to be widely applied during the second and third stage.

The architecture of the smart home model is shown in Fig. 2.

For now the following functions have been implemented:

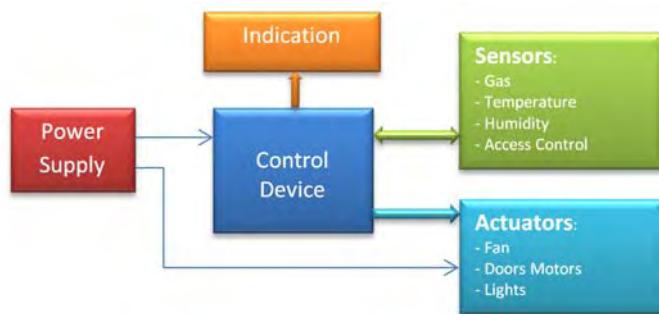


Fig. 2. The architecture of the smart home model

- Access control with Radio Frequency Identification (RFID);
- A garage door control via Bluetooth interface;
- A smoke sensor;
- Detecting of presence and turn on / off lights;
- Measurement, indicating and regulating the environment temperature;
- Humidity measurement and indicating;
- Remote control of indoor devices via Bluetooth interface
- LCD indication.

The model components - the control unit and the sensors and actuators have been selected taking into account the compromise between low cost and desired features, because it is intended for educational purposes [8].

The open hardware development board Arduino Uno is chosen as a control unit (Fig. 3) [9]. It is suitable because of its compact size and low cost. It is based on the 8-bit RISC

Atmel microcontroller – ATmega328. The microcontroller includes 32 KB Flash program memory, 2 KB SRAM data memory, 2 8-bit and 1 16-bit timer/counter, RTC, 8-channel 10-bit ADC, serial interfaces (USART, SPI, IIC), etc.

Arduino is a good choice for educational and hobby purposes, as it can be used as a ready to use core for an embedded system:

- A possibility is provided to power supply and program the microcontroller and peripherals via USB from a PC, and also independently - from a separate connectors for power supply and in-circuit serial programming;
- Clock frequency;
- Reset circuit;
- Expansion connectors accessing all the general purpose input/output pins of the microcontroller. It is possible easily to connect various peripheral modules to them and also to put on the Arduino board and connect so called shields with various purpose produced by many companies.

In future improvements including web-based control Arduino Uno could be used with an Ethernet shield or to be easily changed with another compatible development board, for instance Arduino Leonardo.

The access control is implemented using radio-frequency identification (RFID) module, which is intended for access control, personal identification, home automation, etc. A mini-RFID module has been chosen, with 125 kHz frequency, distance for reading to 35 mm. It is used for reading passive RFID tags, cards, etc.

For the remote control of devices such as garage door, lights, etc. bluetooth interface is implemented by the inexpensive module HC-05. Some features: 2.4GHz ISM band, supports baud rates 2400 -1382400, default COM setting: 9600, N, 8,1.

The doors of the room and the garage are closed/opened by small motors.

Mini PIR sensor HC-SR505 is implemented to detect motion for distance to 3 m, induction angle: <100 degree cone angle and induction distance - 3 meters.

Temperature and humidity are measured by sensor module DHT-22 with accuracy: humidity  $\pm 2\%RH$  (max  $\pm 5\%RH$ ); temperature  $\pm 0.5$  Celsius; resolution: humidity 0.1%RH; temperature 0.1 Celsius.

With increasing/ decreasing the temperature a fan is switched on/ off and also its speed changes when changing the temperature range.

The sensor MQ-2 module is useful for detecting gas leakage - LPG, i-butane, propane, methane, alcohol, hydrogen and



Fig. 3. Arduino Uno Development board



Fig. 4. Ardudroid application

smoke for home and industrial applications.

To remotely control some devices via bluetooth interface and Android based device, like tablet, smartphone, etc., the free application Ardudroid is used (Fig. 4) [10]. It allows a direct control of the development board pins - digital, analog and serial interface.

#### IV. CONCLUSION

The implementation of wireless interfaces in embedded systems for various applications is presented.

A Smart Home model is designed. It is used in teaching "Microprocessor Circuits" and "Embedded Systems" for the Bachelor degree students in Electronics in the Technical University of Gabrovo. The model can be used to study the various components of embedded systems, in particular the wireless interfaces. It is convenient because, unlike other hardware development tools used in the laboratory classes, it includes several devices interfacing wireless in one project.

The further development of the project includes remote monitoring and control via web-based interface (Ethernet and Wi-Fi), and the development of additional modules to the system.

#### ACKNOWLEDGEMENT

The present work is supported by the Science Research Fund at the Ministry of Education, Youth and Science under contract № Д1606Е/2016.

#### REFERENCES

- [1] Dale, Mitch. Using Wi-Fi to Connect Embedded Systems with Mobile Devices. MICROCHIP TECHNOLOGY, October 2013. <http://www.rtcmagazine.com/articles/view/103315>
- [2] Уилмсхерст, Т. Разработка встроенных система с помощью микроконтроллеров PIC. Принципы и практические примеры. СПб, Корона-Век, 2008 г.
- [3] [www.zigbee.org](http://www.zigbee.org)
- [4] [www.irda.org](http://www.irda.org)
- [5] Kandov I., G.Goranov, A.Aleksandrov, Web based system for control and measuring parameters of sensor network, ITHET International conference on information Technology based Higher Education and Training 2014, 11-13 September, pp.1-4.
- [6] George Haynes. Beyond Smart Cities. <https://www.linkedin.com/pulse/beyond-smart-cities-george-haynes-5948850577994760192> (Accessed: april, 2016)
- [7] Cees Links. The New Smart Home is the Really Smart Home. 09/10/2012. <http://www.wirelessdesignmag.com/blog/2012/09/new-smart-home-really-smart-home> (Accessed: april, 2016)
- [8] Jhonson, R., S. Cotton. Electronic Sensors: Making the Connection. The Technology Teacher, May/June 2008, pp. 5-9.
- [9] <http://www.arduino.org/products>
- [10] <http://www.techbitar.com/ardudroid-simple-bluetooth-control-for-arduino-and-android.html> (Accessed: april, 2016)



# **Energy Systems and Efficiency**



# Determination of Optimal Modules Number in Photovoltaic Strings for Inverter Power Maximization

Ljupco Trpezanovski<sup>1</sup>, Dimitar Dimitrov<sup>2</sup>

**Abstract** – In this paper is presented a way for determination the optimal number of photovoltaic (PV) modules connected in a string. The aim for optimal number of modules determination is to obtain the optimal working voltage on inverter DC side and activate the device for Maximal Power Point Tracker – MPPT. The calculations are performed according the PV module and inverter data for Standard Test Conditions and defined specific working scenarios. After that, a way for determination of optimal number of strings connected to the inverter is shown. The calculations are performed under condition to gain the maximum active power from the inverter in all possible working scenarios (working states of the inverter with activated MPPT device). To avoid the inverter damage, during the calculations the inverter limiting working performances are taken into account. The proposed methodology in this paper is practically applied on PV modules from Yingli Solar Company of China and inverter from the SMA Company of Germany.

**Keywords** – Photovoltaic modules, DC/AC inverter, maximum active power, photovoltaic power plants.

## I. INTRODUCTION

The total energy from the solar radiation which yearly reaches the Earth is about  $10^{18}$  kWh/year. This energy is 60.000 times greater then the yearly consumed electricity and 25 times greater from the all fossil energy potentials on the Earth. Because of the energy losses in the atmosphere (absorption and reflection) the average daily energy reaches the Earth surface is about  $5,52 \text{ kWh/m}^2$  [1]. For the specific location on the Earth the exact value of solar radiation will depend of a latitude, longitude and elevation of the location, season (position of the Sun over horizon), temperature, presence of: fog, clouds, pollution, wind etc. Consequently, it is obvious that the Sun is the biggest and main source of renewable energy. The geographical position and a climate in the Republic of Macedonia offer very good perspective for solar energy utilization. The average daily radiation per year is between  $3,4 \text{ kWh/m}^2$  in the north part of the country (Skopje) and  $4,2 \text{ kWh/m}^2$  in the south-west part (Bitola). Total solar radiation per year varies from minimum  $1250 \text{ kWh/m}^2$  in the north to maximum  $1530 \text{ kWh/m}^2$  in the south-west part of the country. Therefore, the yearly average solar radiation is  $1390 \text{ kWh/m}^2$  year [2]. In the Study for renewable sources of

energy in the Republic of Macedonia [2] and Strategy for energy sector development in the Republic of Macedonia in the period 2008-2020 [3] (published by the Macedonian Academy of Sciences and Arts) is predicted construction of 10 to 30 MWp photovoltaic power plants (PVP). These PVP will produce electricity from 14 to 60 GWh/year.

Solar energy transformation in the electricity is directly performed in the photovoltaic (PV) solar cells grouped in solar modules. For more electrical power, the modules are connecting in serial and form a string. More strings can be connected in parallel. The electricity produced in PV modules is on direct current (DC). The electricity conversion from DC to alternative current (AC) is performing in the inverters. With the set of PV modules (connected in strings) and inverter(s) altogether connected in proper way, the PVP are constructed. The purpose of PVP is production of electricity directly with solar energy transformation. The produced electricity can be used individually from separated consumers (Off-Grid) or be injected into a power distribution system (On-Grid).

During the PVP design and construction it is necessary to take into account a way for strings forming and their connection to the inverter(s). These procedures are required because the aim is to get maximum produced electricity from the solar radiated energy on the PV modules surface in combination with inverter(s) optimal working performances.

In this paper from theoretical and practical aspect on the real case of applied PV modules and inverter a way for determination of optimal photovoltaic modules number in a string and optimal strings number connected into inverter for active power maximization is presented.

In the section II the main characteristics for photovoltaic modules and inverters necessary for proposed calculations are given. The calculations for certain types of PV module and inverter are shown in section III. The conclusions are drawn in the section IV.

## II. PHOTOVOLTAIC MODULE AND INVERTER CHARACTERISTICS

The determination of the optimal PV modules number in a string depends of the module characteristics and optimal working performances of the inverter in which the string(s) is/are connected. The theoretical calculations are practically applied on the PV module Yingli Solar type YL235P-29b with power output of 235 Wp [4].

The current-voltage ( $I$ - $V$ ) diagram of this type of PV modules is given on Fig. 1. On this figure can be seen that for different intensity of solar radiation exists only one point on the  $I$ - $V$  diagram for which the product  $V \cdot I$  has maximal value. More exactly, for each curve exists only one working point (named Maximal Power Point – MPP) in which the PV

<sup>1</sup>Ljupco Trpezanovski is with the Faculty of Technical Sciences at University St. Kliment Ohridski, Bitola, Makedonska Falanga 33, 7000 Bitola, Republic of Macedonia,  
E-mail: ljupco.trpezanovski@gmail.com

<sup>2</sup>Dimitar Dimitrov is with the Faculty of Electrical Engineering and Information Technologies at University Ss. Cyril and Methodius, Skopje, Rugjer Boshkovik bb, 1000 Skopje, Republic of Macedonia.

module voltage is  $V_{MPP}$  and flows current  $I_{MPP}$  with production of maximal power  $P_{MPP} = V_{MPP} \cdot I_{MPP}$ .

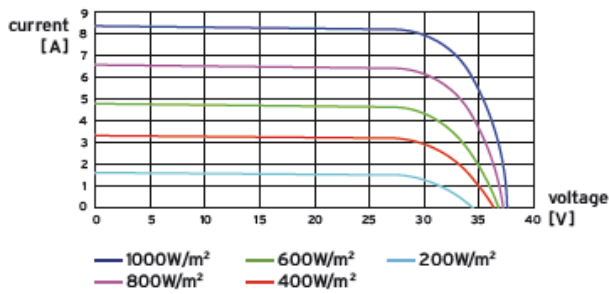


Fig. 1. I-V diagram for PV module type YL235P-29b.

In case for Standard Test Conditions – STC (radiation  $1000 \text{ W/m}^2$ , air mass coefficient  $AM=1,5$  and temperature  $T=25^\circ\text{C}$ ) applied on PV module YL235P-29b, the I-V and  $P = f(I, V)$  diagrams are shown on Fig. 2. The function  $P = f(I, V)$  shows how PV module output power depends from module voltage and current. Open circuit voltage, signed as  $V_{OC}$  is a voltage on not connected plugs of the radiated PV module on STC. A short circuit current, signed as  $I_{SC}$  is current through the PV module output when the plugs are short connected.

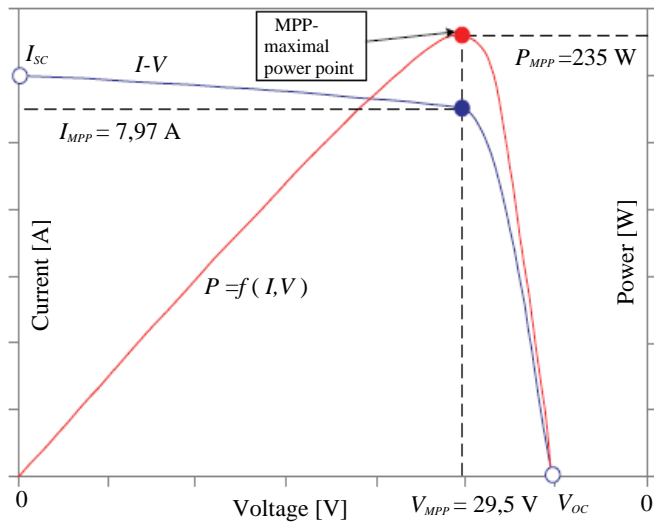


Fig. 2. I-V and  $P = f(I, V)$  diagrams for module YL235P-29b.

The main aim for calculation of optimal PV modules number in a string is to gain maximal active power from the string. In other words, obtained string voltage should be in such interval that with application of Maximum Power Point Tracker (MPPT) device built in the inverter provides inverter working voltage  $U_{MPPT} = V_{MPP}$  and current through inverter  $I_{MPP}$ . The other electrical and thermal characteristics for the PV module type YL235P-29b for STC conditions are given in Table I.

The electricity produced by a string or group of strings connected in parallel is on DC voltage. It is necessary to

convert the electricity from DC to AC voltage if PVP will be connected on a grid. For that purpose group of strings is connected on DC/AC inverter(s).

TABLE I  
ELECTRICAL AND THERMAL PV MODULE CHARACTERISTICS [4].

PV module type	YL235P-29b	
Power output	[W]	235,0
Power output tolerance	[%]	$\pm 3$
Module efficiency	[%]	14,4
Voltage at $P_{max}$ $V_{MPP}$	[V]	29,5
Current at $P_{max}$ $I_{MPP}$	[A]	7,97
Open circuit voltage $V_{OC}$	[V]	37,0
Short circuit current $I_{SC}$	[A]	8,54
Nominal operation Cell Temp. NOCT	[ $^\circ\text{C}$ ]	$46 \pm 2$
Temperature coefficient $\Delta I_{SC}$	[1/ $^\circ\text{C}$ ]	+ 0,0006
Temperature coefficient $\Delta V_{OC}$	[1/ $^\circ\text{C}$ ]	- 0,0037
Temperature coefficient $\Delta V_{MPP}$	[1/ $^\circ\text{C}$ ]	- 0,0045

This type of inverter converts the electricity from DC voltage on AC voltage with minimal losses of electricity. If in the inverter has built in the MPPT device in certain intensity of solar radiation the group of strings will delivered maximum DC power to the inverter. For practical analysis in this paper, inverter type SMA SMC 11000TL [5] is taken in consideration. This inverter is from German producer SMA Solar Technology AG and has very good characteristics. The efficiency curves [5], for different AC powers  $P_{AC}$  (on the output) and different DC voltages  $V_{DC}$  (on the DC input) are shown on Fig 3.

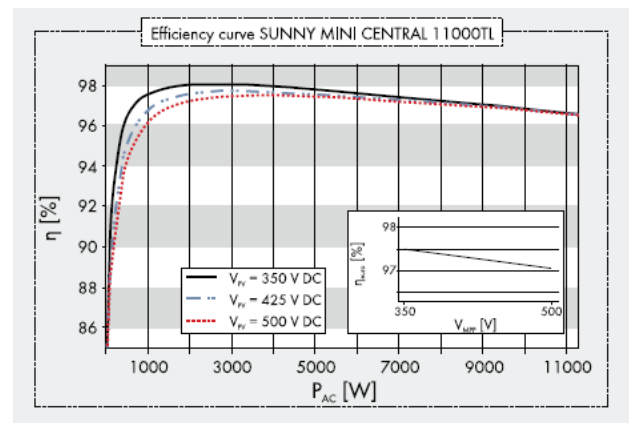


Fig. 3. SMC 11000TL inverter efficiency  $\eta$ [%] curves [5].

All technical data for the inverter necessary for determination of optimal PV modules number in a string and optimal number of strings connected into inverter DC input are given in Table II. It should be taken into account that practically only 5 strings can be connected to the inverter DC input with one MPPT device.



TABLE II  
TECHNICAL DATA FOR INVERTER TYPE SMC 11000TL [5].

Input (DC)		
Maximum DC power	[W]	11.400
Maximum DC voltage	[V]	700
MPP voltage range	[V]	333 – 500
DC nominal voltage	[V]	350
Minimum DC voltage / start voltage	[V]	333/400
Maximum input current / per string	[A]	34/34
Number of MPPT / strings per MPPT	No	1/5
Output (AC)		
AC nominal power (230 V, 50 Hz)	[W]	11.000
Nominal AC voltage range	[V]	180 – 260
AC grid frequency range	[Hz]	50-60±4,5
Maximum output current	[A]	48
Power factor (cosφ)	[p.u.]	1
Efficiency / Euro-eta	[%]	98,0/97,5

### III. DETERMINATION OF OPTIMAL NUMBER OF PV MODULES IN A STRING AND STRINGS IN INVERTER

For certain intensity of solar radiation the string or group of strings (PV generator) will delivered maximum DC power to the inverter if it's MPPT device is activated. According to the technical data of the inverter SMA SMC 11000TL (Table II), the MPPT device will be active if inverter input DC voltage  $U_{PV}$  is between 333 and 500 V, or:

$$U_{MPPT,min} = 333 V \leq U_{PV} \leq 500 V = U_{MPPT,max} \quad (1)$$

This voltage range should be achieved during the long time of PV generator operation. According to the recommendations of inverter producer it is necessary to check the PV generator voltage in the Maximal Power Point  $V_{MPP}$  for PV cells temperature of 15 °C and 70 °C. The  $V_{MPP}$  for a PV module on temperature T can be calculated by Eq. (2), [6]:

$$V_{MPP,T} = V_{MPP,STC} \left[ 1 + \frac{\Delta V_{MPP}}{100} (T - 25) \right], \quad (2)$$

where  $\Delta V_{MPP}$  is temperature coefficient for voltage changing in MPP and  $V_{MPP,STC}$  is PV module voltage in MPP for STC.

These values for PV module type YL235P-29b are:

$\Delta V_{MPP} = -0,45 \% / ^\circ C$  and  $V_{MPP,STC} = 29,5 V$ . Applying the Eq. (2) for the proposed PV module obtained values are: for cells temperature  $T=15 ^\circ C$  the MPP voltage is  $V_{MPP,15} = 30,8 V$  and for  $T=70 ^\circ C$   $V_{MPP,70} = 23,5 V$ .

Taking into account recommendations and also parameters of inverter and PV module the permitted number of PV

modules in a string can take values from  $n_{min}$  to  $n_{max}$ . For the proposed inverter and PV module these numbers are obtained through Eqs. (3) and (4) taken from [6]:

$$n_{min} = \frac{U_{MPPT,min}}{V_{MPP,70}} = \frac{333}{23,5} = 14,17 \Rightarrow n_{min} = 15 \text{ modules/string} \quad (3)$$

$$n_{max} = \frac{U_{MPPT,max}}{V_{MPP,15}} = \frac{500}{30,8} = 16,23 \Rightarrow n_{max} = 16 \text{ modules/string} \quad (4)$$

Inverter producers proposed maximal value for input DC voltage. For inverter SMC 11000TL this voltage is 700 V. During the calculation process for PV module number in a string it is necessary to check if string voltage exceeds maximal input DC voltage. For this check the cells temperature should be  $T=-10 ^\circ C$ . The worst case for voltage increasing is when the PV string (or PV generator) is working in open circuit. If the cells temperature is different than that in STC, the open circuit voltage can be calculated with Eq. (5), [6]:

$$V_{OC,T} = V_{OC,STC} \left[ 1 + \frac{\Delta V_{OC}}{100} (T - 25) \right]. \quad (5)$$

Temperature coefficient for open circuit voltage changing for PV module YL235P-29b is  $\Delta V_{OC} = -0,37 \% / ^\circ C$ . According to this data and Eq. (5) for cells temperature  $T=-10 ^\circ C$ ,  $\Delta V_{OC,-10} = 41,8 V$ .

For 15 serial connected PV modules input DC voltage to the inverter will be  $U_{PV,15} = 15 \times 41,8 = 627,0 V$  whereas for 16 PV modules will be  $U_{PV,16} = 16 \times 41,8 = 668,8 V$ . It is obvious that in both cases the input DC voltage to the inverter don't exceed 700 V.

According to the technical data for the inverter SMA SMC 11000TL maximal power on the DC input (from PV generator side) is 11400 W. In real working conditions PV module maximum (peak) power cannot be greater than power on STC. For the proposed PV module YL235P-29b this power is 235 Wp. In case for 15 PV modules, the peak power of a string is  $15 \times 235 = 3525 Wp$  whereas in case for 16 PV modules  $16 \times 235 = 3760 Wp$ . Technically it is possible to connect 5 strings to the proposed inverter. However, taking into account before explained voltage and maximal DC input power limits for the chosen type of PV modules, the optimal number of strings that can be connected to the inverter is 3. Therefore, the PV generator should be constructed with 3 strings connected in parallel with same number of PV modules per string.

In case for 15 PV modules per string the peak power to the inverter DC input will be  $3 \times 3525 = 10575 Wp$  whereas in case for 16 PV modules per string  $3 \times 3760 = 11280 Wp$ . The main purpose in PVP designing is to obtain the maximum active power from the inverter as it is possible. According the above conducted considerations the optimal number of PV modules per string is 16 (Fig. 4 a)) and optimal number of parallel strings connected on inverter DC input is 3 (Fig. 4 b)).

Finally, it is necessary to check if the current of calculated number and configuration of modules and strings doesn't exceed limited maximal current of the inverter. According to the technical data in Table II, for the proposed inverter value of this current is  $I_{PV,max} = 34$  A. Maximal current which can flow through a string is short current of PV modules.

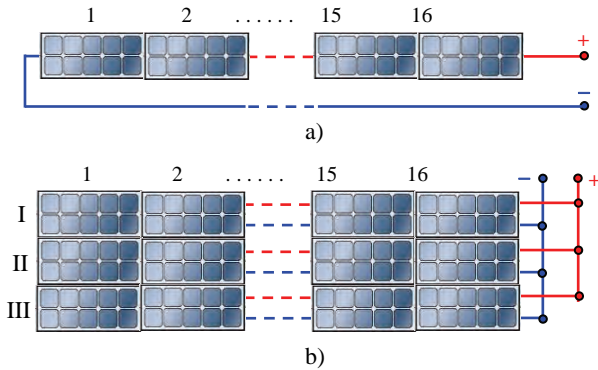


Fig. 4. a) String with 16 PV modules b) PV generator with 3 strings.

During the calculations it can be taken into account that the value of this current rises when the cells temperature rise. For the proposed PV module the short current is  $I_{sc,STC} = 8,54$  A on STC and temperature coefficient is  $\Delta I_{sc} = 0,06$  % / $^{\circ}$ C. Short current value for cells temperature  $T=70$   $^{\circ}$ C can be calculated according Eq. (6), [6]:

$$I_{sc,T} = I_{sc,STC} \left[ 1 + \frac{\Delta I_{sc}}{100} (T - 25) \right]. \quad (6)$$

With the above adequate quantities and applying Eq. (6) the PV module short current (which is the same as a string short current)  $I_{sc,70} = 8,77$  A is calculated.

Because the PV generator is consist of 3 parallel connected strings maximum current in the inverter DC input can be short current of the PV generator or  $3 \times 8,77 = 26,31$  A. Therefore, the inverter maximum DC input current of 34 A (Table II) will not be exceed. Taking into account all considerations it can be

conclude that 3 parallel connected strings with 16 PV modules per string is optimal combination from the aspect of maximum gained active power for certain solar radiation intensity. This combination of PV modules and strings totally satisfy all limitations and recommendations for accepted PV modules YL235P-29 and inverter SMA SMC 11000TL.

#### IV. CONCLUSION

During the designing and construction of PVP it is necessary to gain maximum power from the chosen equipment for certain solar radiation intensity. Beside selection the quality equipment, it is required to match optimal PV generator parameters with the inverter parameters. In this paper is presented a way for determination of optimal photovoltaic modules number in a string and optimal strings number connected into inverter for active power maximization. All limiting factors of the inverter as: MPPT device working voltages, maximal permitted voltage and current are taken into account. On a real case of chosen types of PV module and inverter are performed calculations. Step by step is explained way of calculations applying technical data of PV module and inverter. The obtained results are analyzed and commented.

#### REFERENCES

- [1] V. Mijalović, "Distribuirani izvori energije – principi rada i eksploatacioni aspekti", Akademska misao, Beograd, 2011.
- [2] Team of ICEIM-MANU, *Base study for use of renewable energy sources in R. Macedonia until 2020*, Government of R. Macedonia, Skopje, 2010.
- [3] Team of ICEIM-MANU, *Strategy for energy sector development in R. Macedonia until 2030*, Government of R. Macedonia, 2010.
- [4] Yjngli Solar YL 235 P-29b / 1650x990 Series, Technical Data, Yingli Green Energy Holding Co. LTD.
- [5] Sunny Mini Central SMC 11000TL, Technical Data, SMA Solar Technology AG, Germany.
- [6] DGS Berlin, "Planning and Installing Photovoltaic Systems. A guide for installers, architects and engineers", Second Edition, Earthscan, ISBN 978-1-84407-442-6, London, 2008.

# Transient Stability of Asynchronous Generator on Distribution Network

Metodija Atanasovski<sup>1</sup>, Mitko Kostov, Nikola Acevski, Blagoj Arapinoski, Elena Kotevska

**Abstract** – Dispersed generation (DG) brings together wide range of technologies for electricity production. Most frequently used electric machines for electricity production at each technology that belongs to DG group are: synchronous generators, asynchronous generators (AG) and power converters. This paper deals with transient stability of AG. Typical model of distribution network (DN) is developed with AG connected to it. The modelling and simulation are performed with NEPLAN and MATLAB/SIMULINK/Simpowersystems toolbox software packages. AG with squirrel cage is used. Several short circuits are simulated in order to investigate dynamic behaviour of AG. Following parameters are analyzed: rotor speed, active and reactive power, and currents of AG. Critical clearing time is calculated for keeping stable operation of AG. Several useful and practical conclusions are obtained.

**Keywords** – Asynchronous generator, Transient stability, Distribution network.

## I. INTRODUCTION

Last decade of past century bring remarkably revival of interest for connection generation units to DN. This phenomena in power systems is called dispersed generation. The term dispersed is introduced to make a difference with conventional centralized generation usually connected on transmission network. The story of DG is not a new one. Construction of small generating units is well known long ago from the beginning of power systems industry [1].

The installed capacity of DG is a key factor for obtaining voltage level for its connection to DN. Voltage levels for DG connection are the typical distribution voltage levels which vary from 400 V up to 110 kV [2].

DG category definition is not based on primary source used for electricity generation, but it is based on its technical performances from power system point of view. DG brings together wide range of technologies for electricity production. Most frequently used electric machines for electricity production at each technology that belongs to DG group are: synchronous generators, asynchronous generators (AG) and power converters (dc/ac or vice versa). AGs are mostly used at wind farms, small and medium size hydro power plants [2].

The problem of transient stability on DN becomes very interesting with increased presence of DG. Reasons about this can be summarized as: investigation of generators behaviour

on DN and their impact on it, calculation of critical clearing time for remaining generator stable operation and impact of generators transient stability on networks protection.

This paper deals with transient stability of AG. Typical model of distribution network is developed with AG connected to it. The paper is consisted of four sections. Concept and model of AG for transient stability analysis is explained in section II. Also critical clearing time for stable operation is defined. The modelling and simulation are performed with NEPLAN and MATLAB /SIMULINK/ Simpowersystems toolbox software packages. AG with squirrel cage is used. Obtained results from performed simulations are depicted in section III. Three phase and two phase short circuits are simulated in order to investigate dynamic behaviour of this type of generators. Following parameters are analyzed: rotor speed, active and reactive power, and currents of AG. Critical clearing time is calculated for keeping stable operation of AG. Several useful and practical conclusions are obtained and elaborated in section IV.

## II. DISTRIBUTION NETWORK MODEL WITH ASYNCHRONOUS GENERATOR

DG transient stability studies are similar to large scale power system transient stability studies, except that DG capacity is normally very small relative to the bulk system and has no significant influence on its frequency or stability. The research is performed for DG that uses AG as electric machine. Practically the purpose is to investigate the ability of AG unit to remain synchronized and determining its critical clearing time, when feeder disturbance occurred on the DN.

The concept of critical clearing time in the case of synchronous generator is well known. Unlike synchronous generators, AGs do not have field windings to develop the required electro-magnetic field in the machine's air-gap. Therefore, AGs can not work with out external power supply. The electro-magnetic torque ( $T_e$ ) developed inside an induction machine at any given speed is proportional to the square of the terminal voltage as follows [3]:

$$T_e = K \cdot s \cdot U^2 \quad (1)$$

Where  $K$  is constant value depends on the parameters of the machine,  $s$  is the machine slip.

Electro-magnetic torque is, therefore, bound to reduce following a fault condition, proportionally to square of voltage. On the other hand, the dynamic behavior of the rotor is governed by the swing equation given below:

<sup>1</sup>Metodija Atanasovski, Mitko Kostov, Nikolace Acevski, Blagoja Arapinoski and Elena Kotevska are with the Faculty of Technical Sciences-Bitola, University St. Kliment Ohridski, Republic of Macedonia, E-mail: metodija.atanasovski@tfb.uklo.edu.mk.

$$J \frac{dw}{dt} = T_m - T_e \quad (2)$$

Where:  $J$  is moment of inertia of the rotating mass,  $T_m$  is the mechanical torque applied on rotor of the associated wind or hydro turbine,  $w$  is rotor speed.

It can be concluded from (2) that assuming the mechanical torque is kept constant, then any reduction in the electromagnetic torque, for instance due to fault condition, causes the rotor to accelerate. This in turn leads to an increase in the kinetic energy of the rotating mass. When the fault is cleared and consequently system voltage recovers, the magnetic field inside the air-gap of the machine starts to build up. This causes high inrush current to be drawn by the machine from the network which in turn causes a voltage drop across the interfacing link between the AG and the substation leading to a reduction in the voltage at the generator terminals. The resulting electro-magnetic torque acts on the rotor in a direction opposite to that of mechanical torque applied by turbine. If the energy stored in the newly established rotating magnetic field becomes higher than that stored in the rotating mass, rotor speed is forced to slow down and the generator eventually retains its normal operating condition following few oscillations, otherwise, its speed continues to increase until it is tripped by appropriate protection devices. When this is happening generator terminals usually experience sustained voltage dip. This investigation has shown that there is a maximum time for the fault to be cleared, otherwise AG losses its stability. Such time will thereafter be referred to as the critical clearing time for AG [4].

### III. SIMULATION AND RESULTS

Study case analyzed in this paper is shown on Fig. 1. It consists of DN with AG of 1 MVA rated power. The modelling and simulation are performed with NEPLAN [5] and MATLAB /SIMULINK/ Simpowersystems toolbox [6] software packages. AG with squirrel cage is used which is represented with fourth and sixth order model for transient stability. Mechanical moment (power) of prime mover (turbine) is considered constant to achieve greater generics of results. Part of the reactive power consumed by AG is supplied locally with model of capacitor.

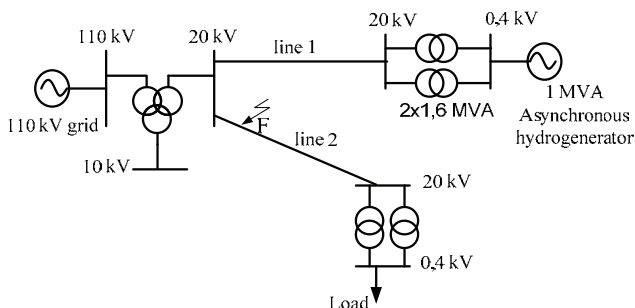
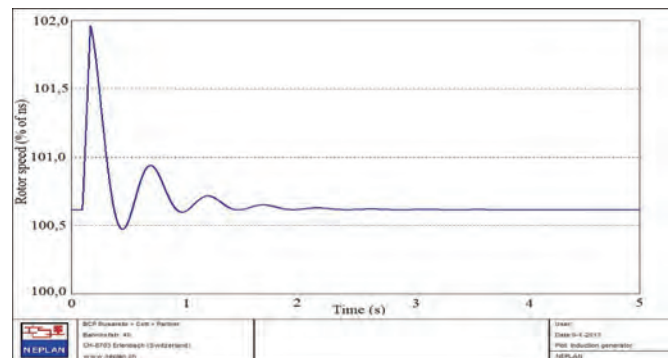


Fig.1. Schematic diagram of the case study DN with AG

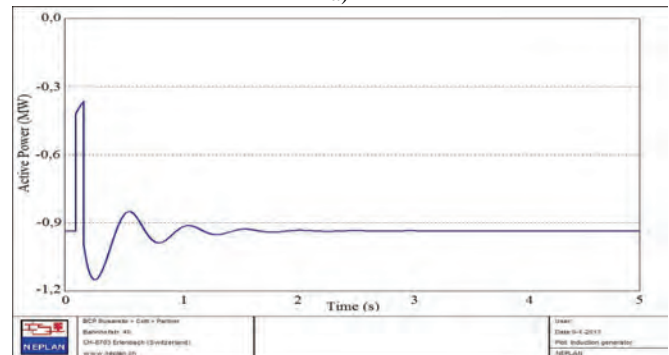
Generator is integrated into the DN at 20 kV voltage level through 2x1.6 MVA 0.4/20 kV/kV step-up transformers and

20 kV distribution line 1. 110 kV network is represented by infinite model with voltage source behind its Thevenin's equivalent impedance. The fault level of the 110 grid is assumed 5000 MVA. HV/MV substation is represented by three windings transformer 110/20/10 kV/kV/kV and 31,5/31,5/10,5 MVA/MVA/MVA. The load is connected to the substation through distribution line 2 and 20/0,4 kV/kV transformers. Load characteristic is represented with constant impedance model. Both lines 1 and 2 are simulated with  $\pi$ -equivalent circuit with impedance of  $(0.413 + j0.36) \Omega / km$ . All transformers are modeled in a same way as in short circuit calculations.

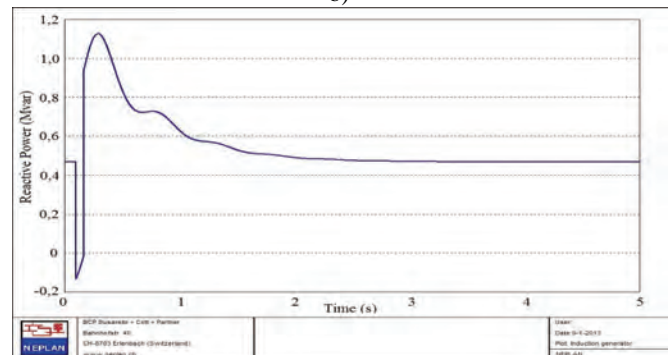
Simulations are performed for three and two phase short circuits at location F (see Fig. 1), located on 20 % of line 2 length measured from substation HV/MV/MV. Results for three phase fault duration of 70 ms are shown on Fig. 2. The corresponding AG rotor speed is shown on Fig. 2(a), active power of AG on 2(b) and reactive power on 2(c).



a)



b)



c)

Fig. 2. Simulation of three phase fault at location F with duration of 70 ms; a) Rotor speed of AG in %, b) Active power of AG, c) Reactive power of AG.



Fig. 2a) shows that rotor speed retains stable position through few oscillations after fault clearance. Analyzing Fig. 2b) it is obvious that AG active power has dropped to 40 % of its nominal value during fault duration, but after short circuit is cleared active power retained its normal operational value. Fig. 2c) depicts reactive power variation of AG.

The different behavior of each generator can be explained by analyzing the response of the reactive power exchanged between the generator and the network for each situation. In the case of the induction generator, the reactive power exchanged takes into account the reactive power supplied by the capacitors. When fault occurs (Fig. 2c)) AG for a fault duration injects reactive power into the network due to self-excitation phenomenon, but, soon after fault clearance, AG consumes a large amount of reactive power, which can lead the system to a voltage collapse if it is not disconnected quickly with fast response of protection.

Due to this response of reactive power of AG, its small critical clearance time can be explained. Namely, after several repeated simulations by increasing fault duration time, critical clearance time of AG is calculated to be 100 ms. Fig. 3 shows rotor speed behaviour due to fault duration of 130 ms. It is obvious that AG is unstable and its rotor speed continuously increase until protective devices disconnect AG from the network.

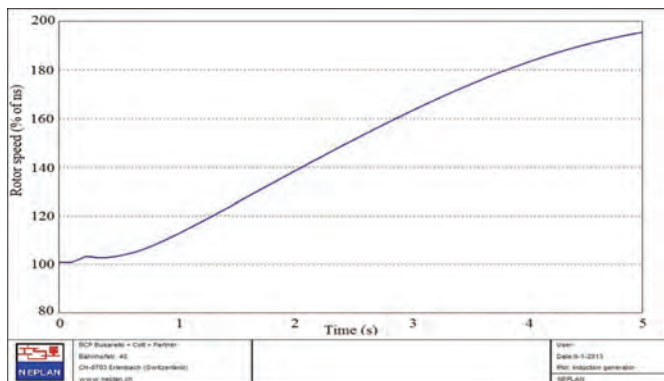


Fig. 3. AG rotor speed variation in % for simulation of three phase fault at location F with duration of 130 ms

The same DN from Fig. 1 is also modeled using program for analysis electromagnetic transients. MATLAB SIMULINK/Simpowersystems toolbox is used for investigating the response of AG phase currents to symmetrical and asymmetrical short circuits in DN. Three phase model of network elements is developed and simulations are performed in continuous time domain. Previous simulations in NEPLAN are done in phasor domain which is usually used for transient stability analysis.

For purposes of these simulations at location F three and two phase short circuits are applied. Fig. 4 and Fig. 5 show AG stator phase currents behavior for three and two phase short circuit appropriately. Analyzing stator phase currents of AG for three phase short circuit (Fig. 4) at location F occurred at  $t=50$  ms, it can be seen that initially current magnitude is high but, it decrease quickly because this machine has no capacity to provide sustained short-circuit currents during three-phase faults. Voltages (phase or line to line) on AG

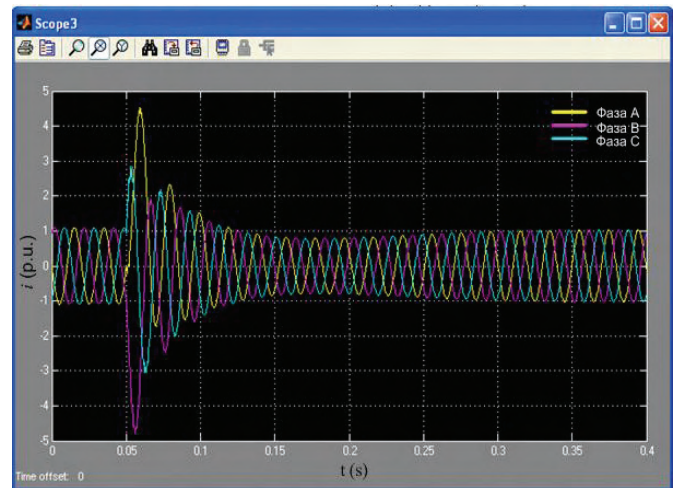


Fig. 4. Stator phase currents of AG for three phase short circuit at location F

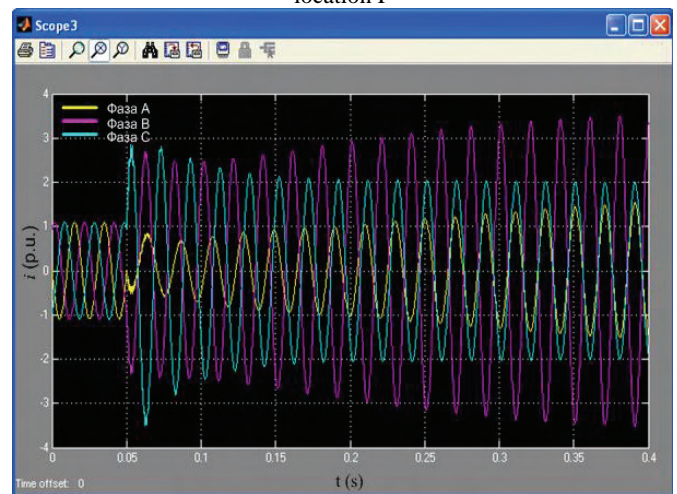


Fig. 5. Stator phase currents of AG for two phase short circuit at location F

terminals drop on 30 % of their nominal value. If short circuit occurs on AG terminals voltages drop to zero and the capacitor bank becomes unloaded. Consequently, there is no external excitation source for the generator, and it becomes unable to produce voltage. Theoretically, this fact could become the detection of faults by protection systems based on over-current relays more difficult. However, in this case, voltage-based relays could be used.

Analyzing AG stator phase currents for two phase short circuit (Fig. 5) at location F occurred at  $t=50$  ms, it can be concluded that AG current response demonstrates capacity for sustainable supply of short circuit current. All three phases remain excited by the network. Currents at faulted phases B and C present sustained response, although current at healthy phase A also shows sustainable increase of its amplitude.

#### IV. CONCLUSION

Dynamic behaviour of AG connected on DN is investigated in the paper. It can be concluded that AG has very low critical clearing time to remain its stability during fault in the

network. Reasons for this phenomenon are in reactive power response during fault occurrence and after its clearance.

Performed simulations have clearly shown that current response of AG is with high amplitudes of currents at all phases in the moment of three phase short circuit (symmetrical short circuit) occurrence. However, currents rapidly decrease because AG has no capacity to provide sustained supply of short-circuit currents during three-phase faults. This fact imposes conclusion that in networks which have reached circuit breakers breaking current limit, AG can be used in DG installations. For asymmetrical short circuits, current response of AG has capacity to provide sustained supply of short-circuit currents.

Comparison of AG dynamic behavior with synchronous generator is very important for determination of there advantages and disadvantages from DN stand point of view. This conclusion suggests further investigation about this matter. Synchronous generator can be applied on the same network model for obtaining its dynamic performance and critical clearing time for same fault conditions in DN.

## REFERENCES

- [1] Atanasovski M., Taleski R., "Power Summation Method for Loss Allocation in Radial Distribution Networks With DG," *IEEE Trans. on Power Systems*, Vol. 26, No. 4, pp. 2491–2499, November 2011.
- [2] Nick Jenkins, Ron Allan, Peter Crossley, Daniel Kirschen and Goran Strbac: *EMBEDDED GENERATION*, Published by The IEE, London, United Kingdom, 2000.
- [3] Salman K. S., Ibrahim M. Rida. "Investigating the Impact of Embedded Generation on Relay Settings of Utilities Electrical Feeders." *IEEE Trans. on Power Delivery*. Vol. 16, No. 2, pp. 246-251, April 2001.
- [4] Walmir Freitas, J. C. M. Vieira, A. Morelato, L. C. P. da Silva, V. F. da Costa, F. A. B. Lemos. "Comparative Analysis Between Synchronous and Induction Machines for Distributed Generation Applications". *IEEE Trans. on Power Systems*, Vol. 21, No.1, pp. 301–311. February 2006.
- [5] NEPLAN User's Guide V5, BCP BCP, [www.neplan.ch](http://www.neplan.ch).
- [6] User's Guide MATLAB/SIMULINK/SimPowersystems Toolbox, version 3, The MATHWORKS, 2003.

# Economic Energy Scheduling of an Islanded Microgrid

Galia Marinova<sup>1</sup> and Vassil Guliashki<sup>2</sup>

**Abstract** – This paper considers the calculation of an effective energy schedule in an islanded microgrid. GridLab-D open source simulation tool is used to for simulation of microgrid elements. Matlab environment is used to run an optimization solver. The product GridMat is used as an interface tool between Matlab and GridLab-D. An economic scheduling optimization problem on the considered microgrid is formulated and solved.

**Keywords** – Microgrids, GridLab-D, GridMat, Matlab, Energy scheduling optimization.

## I. INTRODUCTION

A microgrid represents a low-voltage distribution system consisting of distributed energy resources (DERs) or renewable energy resources (RES) and controllable loads, which can be used/controlled in either islanded or grid-connected mode. The microgrid should be robust in controlling supply, demand, voltage, and frequency. The DERs/RES production plan can be evaluated by using meteorological forecasts, which have an intrinsic uncertainty. In such a setup, energy storage can help in meeting the hourly production plan [2].

In this paper the microgrid economic scheduling is studied, i.e. the problem to optimize the energy storage/battery schedule, as well the schedule of the diesel generator used, covering the time-varying energy demand and operational constraints while minimizing the costs of internal generated energy. The experimental microgrid includes a photovoltaic system, a wind turbine, a diesel generator and three houses. The microgrid's point of common coupling (PCC) is disconnected from the main grid and the microgrid operates in an island mode. Formulating an optimization task the amount of power demand and supply for the next 24 hour period may be assumed to be known without any change. This is an unrealistic setup, especially in real world applications.. Also the solar radiation forecasts could be inexact and could vary essentially. For this reason the energy, generated by the diesel generator should include a reserve rate (see [5, 6]) and the forecasted data for the renewable energy resources (wind turbine and photovoltaic system), as well as for the loads (houses) should be taken adding a safe margin for each microgrid element.

The open source GridLab-D (see [3]) is used to simulate all the elements of the microgrid. The software product GridMat (see [1]) is used as an interface tool between Matlab (see [4])

and GridLab-D. Climate data, available on the official website of GridLab-D, are used for the simulations. The optimization problem is formulated and solved in an efficient way by using the Matlab optimization toolboxes/solvers.

## II. THE EXPERIMENTAL MICROGRID

The microgrid studied in this work, is composed by several units which produce, exchange and consume energy. Essentially, the microgrid operates with a three-phase medium voltage alternating current (AC) transmission system, which can be connected or not to the main grid (Network) through a transformer system, in order to buy the energy necessary to cover the demand, or sell the surplus energy produced by the RES. When the microgrid is used in an Island mode (disconnected from the Network), a diesel generator is considered in order to supply, together with the RES, the energy necessary to cover the loads. Two type of RES are considered connected in the Microgrid: 1) a photovoltaic system composed by an inverter and a group of solar panels, and 2) a wind turbine. A group of batteries (energy storage system) is also interconnected to the microgrid through a DC/AC bi-directional inverter. This is exactly the part that makes the microgrid under study a *smart* microgrid, since the schedule of batteries is based on the behavior of the loads and the energy production by the RES. The system configuration of the proposed microgrid is presented on Fig. 1:

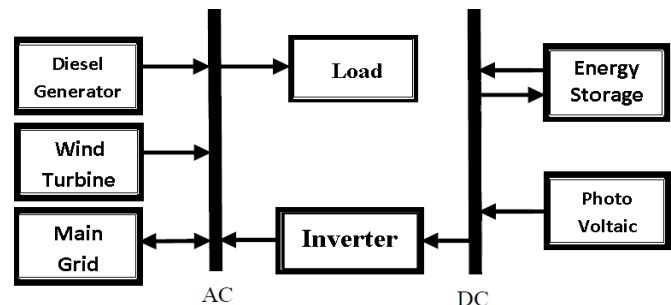


Fig. 1. Microgrid system configuration

The microgrid with all its components is shown on Fig 2.

## III. ECONOMIC SCHEDULING OPTIMIZATION MODEL

In this study the behavior of the RES and houses has been simulated from historical climate data of a particular geographical position: Seattle (USA); The data for solar radiation and wind speed, as well for the houses energy consumption are real data for a given winter day. They are taken as a forecasted data.

In [5] are given energy safety margins necessary to cover the uncertainty of the forecasted data. Taking into account these margin values, in the created optimization model are assumed the following values: Wind turbine: (–30%);

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

<sup>2</sup>Vassil 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.



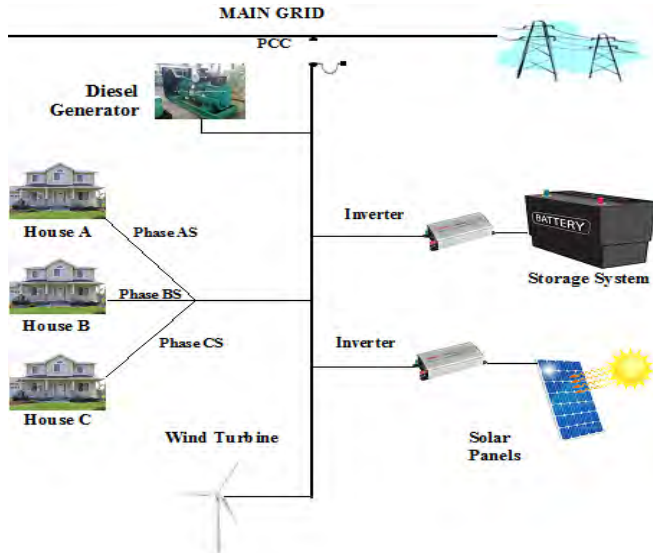


Fig. 2. The experimental microgrid

Photovoltaic: (-37%); Houses: (+25%); Diesel generator: (+20%). Having available correct forecasted data for the RES production and houses consumption one day before for the next day, it is possible to optimize the microgrid behavior for a whole year, solving one day ahead the correspondent scheduling optimization problem for the next day.

The time interval being analysed (one day and one night) is divided by 24 time steps, each with 1 hour length. The balance power  $P_B$  of the studied microgrid should satisfy the following equations (see [6]):

$$P_{RES} + P_B = P_L \quad (1)$$

$$P_B = P_{Bat,d} + P_{DG} \quad (2)$$

where  $P_{RES}$  is the output power of renewable energy sources,  $P_B$  is the balance power,  $P_{Bat,d}$  is the power from discharging the battery system,  $P_{DG}$  is the output of the diesel generator, and  $P_L$  is the microgrid load, equal to houses consumption energy plus battery system charging energy. The parameters and the decision variables used are presented in Table 1.

TABLE 1. PARAMETERS

Parameter	Description
CS	Capital cost for interval of one hour
OM	Operation maintenance for one hour
RC	Replacement cost (of the battery)
FC	Fuel cost for interval of one hour
EC	Emission cost for interval of one hour
CRF	Capital recovery factor for one hour
SFF	Sinking fund factor for one hour

Taking into account that the photovoltaic area, the wind turbine capacity, as well as the house voltage consumption cannot be subject to optimization since their schedules are independent, the objective function includes the balance power:

$$\begin{aligned} \min F = & \sum_{t=1}^{24} (C_t \cdot P_{Bt}) = \sum_{t=1}^{24} CC_{DG}(t) + OM_{DG}(t) + FC_{DG}(t) + EC_{DG}(t) + \\ & + \sum_{t=1}^{24} OM_{Bat}(t) + RC_{Bat}(t) + CC_{Inv}(t) \end{aligned} \quad (3)$$

where  $P_{Bt}$  is the balance power for hour  $t$  and  $C_t$  is the cost of this power. In  $C_t$  are included the depreciations costs of each microgrid energy generation element (unit), of operational costs of individual units, of the fuel cost (for the fuel consumed by the diesel generator), and of emission cost. Calculating  $F$  only the hours, when the diesel generator operates and when the battery system is charging/discharging are taken into account. In [9, 10] are given formulas for calculating the correspondent annual values. Hence the one hour capital cost of microgrid units, which do not need a replacement during the project life time, such like diesel generator and inverter, is calculated as follows:

$$CC_{DG} = \frac{Ccap_{DG} \cdot CRF(i, y)}{5375} \quad (4)$$

Assuming, that the diesel generator is used average 15 hours in a 24 h period, the denominator is:  $5375 = 15 \times 365$ ;

$$CRF(i, y) = \frac{i \cdot (1+i)^y}{(1+i)^y - 1} \quad (5)$$

Here  $Ccap_{DG}$  is the capital cost (US\$),  $y$  is the project life time, and  $i$  is the annual interest rate [11]:

$$i = \frac{i' - f}{1 + f} \quad (6)$$

where:  $i'$  is the loan interest (%), and  $f$  is the annual inflation rate (%).

The one hour operation maintenance cost is:

$$OM = \frac{Ccap_{DG} \cdot (1 - \lambda)}{5375 \cdot y} \quad (7)$$

for the diesel generator, and

$$OM = \frac{Ccap_{Bat} \cdot (1 - \lambda)}{6570 \cdot y} \quad (8)$$

for the battery, where:  $\lambda$  is the reliability of correspondent unit.

Assuming, that the battery bank is used average 18 hours in a 24 h period (i.e.  $365 \times 18 = 6570$  hours annually), the one hour battery bank replacement cost is:

$$RC = \frac{Crep_{Bat} \cdot SFF(i, y_{rep})}{6570} \quad (9)$$

where:  $Crep$  is the replacement cost of battery bank, and SFF is the sinking fund factor, which is calculated as follows [11]:

$$SFF = \frac{i}{(1+i)^y - 1} \quad (10)$$

The one hour fuel cost of diesel generator for hour  $t$  is:

$$FC = Cf \cdot G(t)$$

where:  $Cf$  is the fuel cost per liter, and  $G(t)$  is the hourly consumption of diesel generator [7, 8, 9, 10] as follows:

$$G(t) = (0,246P_{DG}(t) + 0,08415 \cdot P_R) \quad (11)$$

where:  $P_{DG}(t)$  is the diesel power at time  $t$ , and  $P_R$  is the rated power of the diesel generator.

The hourly emission cost ( $CO_2$  emission) is:

$$EC(t) = \frac{E_f \cdot E_{cf} \cdot P_{DG}(t)}{1000} = 0,0187 \cdot P_{DG}(t) \quad (12)$$

where:  $E_f$  is the emission function (kg/kWh), and  $E_{cf}$  is the emission cost factor (\$/ton)

The necessary economic data are given in Table 2:



**Table 2. THE ECONOMIC DATA**

Interest rate $i^*$ (%)	3
Inflation rate (%)	1,6
Inverter life time (years)	20
Battery life time (years)	10
Reliability of inverter (%)	0,98
Reliability of battery (%)	0,98
Reliability of diesel (%)	0,9
Cost of diesel generator (US\$/KW)	500
Cost of battery bank (US\$/KWh)	200
Cost of inverter (US\$/KW)	1000
Fuel cost ( $C_f$ ) (US\$/l)	0,75
Emission function (kg/kWh)	0,34
Emission cost factor (US\$/ton)	55

Other parameters to be defined are the  $P_{bt\_max}$ , fixed to 10 kW for charging and discharging, and  $E_{bt\_max}$ , fixed to 100 kWh. The data in Table 3 are taken from [9], only the fuel cost value is taken from [10]. Since  $P_R = 38$ , hence  $Ccap_{DG} = 19000$  \$. In [6] is stated, that the high speed (3600 r/min), air-cooled diesel can be used for about 20 000 h. Hence  $y$  in formulas (5), (7) and (8) is:  $y = 3,721$ . The annual interest rate  $i = 0,53846154$ . Hence  $CRF(i, y) = 0,6742$ .  $CC_{DG} = 2,38$  \$/h.  $OM_{DG} = 0,095$  \$/h.  $OM_{Bat} = 0,0061$  \$/h.  $Crep_{Bat} = 20000$  \$.  $SFF = 0,1357$ .  $RC = 0,413$  \$/h.  $Ccap_{Inv} = 10000$ . The inverter one hour capital cost is:  $CC_{Inv} = 1$  \$/h.

Hence the objective function (3) is presented in the form:

$$\min F = \sum_{t=1}^{24} 2,38_{DG}(t) + 0,095_{DG}(t) + 0,1845.P_{DG}(t) + 2,398_{DG}(t) + 0,0187.P_{DG}(t) + \sum_{t=1}^{24} 0,274_{Bat}(t) + 0,413_{Bat}(t) + 1_{inv}(t) \quad (13)$$

The constraints concerning the diesel generator are:

$$0,3.P_R \leq P_{DG}(t) \leq P_R \quad (14)$$

Taking into account the modified values from [5], the following constraint is obtained:

$$P_{DG}(t) = \begin{cases} 1,2.(1,25.P_L - 0,63.P_{PV} - 0,7.P_{WT} - P_{Bat\_d}) \\ \text{if } 0,63.P_{PV} + 0,7.P_{WT} + P_{Bat\_d} < 1,25.P_L \\ 0. & \text{otherwise} \end{cases} \quad (15)$$

The constraints concerning the battery system are:

$$-P_{bt\_max} \leq P_{Bat}(t) \leq +P_{bt\_max} \quad (16)$$

$$SOC_{min} \leq SOC(i) \leq SOC_{max} \quad (17)$$

$$\sum_{t=1}^{24} P_{Bat}(t) = 0; \quad t = 1, \dots, 24; \quad (18)$$

where:  $P_L(t)$  is the power absorbed by the houses during the hour " $t$ " [kW];  $P_{PV}(t)$  is the power delivered by photovoltaic panels during the hour " $t$ " [kW];  $P_{WT}(t)$  is the power delivered by wind turbine during the hour " $t$ " [kW];  $P_{Bat\_d}(t)$  is the power delivered by the battery block (discharging) during the hour " $t$ " [kW].  $P_{bt\_max}$  is the maximum power that the battery system can deliver/absorb [kW];  $SOC(t)$  is the State Of Charge of the battery during the hour " $t$ " [%]  $SOC_{min}$  = lower limit for the State Of Charge of the battery [%]  $SOC_{max}$  = upper limit for the State Of Charge of the battery [%].

Finally taking into account the energy balance of the microgrid (see equations (1)-(2)), the last constraint obtained is:

$$P_{Bat}(t) + P_{DG}(t) \geq P_H(t) - P_{PV}(t) - P_{WT}(t), \quad t = 1, \dots, 24; \quad (19)$$

where  $P_H(t)$  is the house consumption energy. The energy  $P_{Bat}(t)$  is considered positive when the battery is discharging and negative when is charging. Therefore, the equation (16) represents the power limit, which can be delivered or absorbed by the inverter tie to the battery system; the system cannot supply or absorb a power more than the  $P_{bt\_max}$ .

The SOC of the battery represents the amount of energy stored in the battery system. Therefore, the equation (17) means that, for each time step, the SOC must be included between a minimum and a maximum value depending by the system used to storage the energy and agree with physical limit of maximum SOC of 100%. In this case, the minimum and maximum level of SOC are fixed to 20% and 100% respectively.

The SOC is depending on the value of  $P_{Bat}(t)$  for each time step; the relation between these variables is shown below:

$$SOC(t) = SOC(t-1) - \frac{P_{Bat}(t)}{E_{bt\_max}} \Delta t \quad (20)$$

where:  $\Delta t$  is the time step [h],  $SOC(0)$  = Initial charge of the battery (it is an input value of the problem). In this optimization problem, the initial value of the SOC is fixed to 50%. It means that, at the begin of the optimization, the battery system is charged to the half of its full charge.

The constraint, shown in equation (18), is used in order to get, at the end of the 24h period, the same value of SOC like at the begin of the period.

#### IV. TEST RESULTS OBTAINED BY MEANS OF THE SIMULATION AND THE OPTIMIZATION TOOLS

The simulations with GridLab-D give the results about the consumption of the houses, the production of the solar panels and wind turbine. The results obtained from the simulations, for a winter day, are shown on Fig. 3.

The problem (13)-(19) has 48 variables:  $P_{DG}(t)$  and  $P_{Bat}(t)$ ,  $t=1, \dots, 24$ ; To solve this optimization problem, the Matlab solver *fmincon* has been used. In the "Help" menu in Matlab, in the "Optimization Toolbox" the description and explanation how to run this solver is given.

The first 24 variables on Fig. 4 represent the battery schedule, and the next 24 variables correspond to the diesel generator schedule.

The calculated optimal schedule for the battery system is:

[-10. -10. -100 1.3786 10. -10. -1.1083 10. 10. -10. -10. -10. 9.7297 -10. -10. 10. 10. 10. 10. -10. 10. 10. -10.]

The calculated optimal schedule for the diesel generator is:

[11.4 11.4 11.4 11.4 0. 11.4 22.0169 32.5701 22.3065 17.9062 14.4096 11.5544 0. 11.4 11.4 0. 18.4714 25.1381 29.1214 28.5803 26.6152 24.9310 20.7732 13.0999].

The optimization result shows, that the total cost for 24h period for a winter day, based on the objective function (13) is:  $F = 205,037$ . Without optimization for an initial battery schedule:  $P_{Bat}(0) = [0 \ 0 \ 0 \ -0.8 \ -0.8 \ -0.8 \ -0.8 \ -0.8 \ -1 \ 0 \ 1 \ 1 \ 1 \ 1 \ 1]$

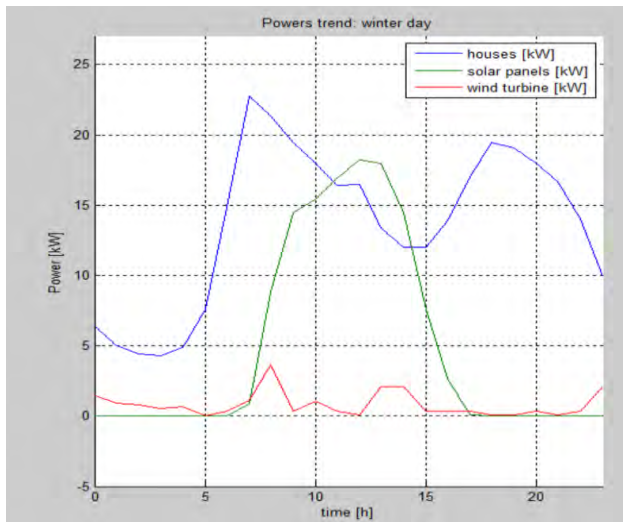


Fig 3. Houses consumption, solar panels and wind turbine energy generation for a winter day

The results of running "*fmincon*" for a winter day using the "*optimtool*" command in Matlab are shown on Fig. 4.

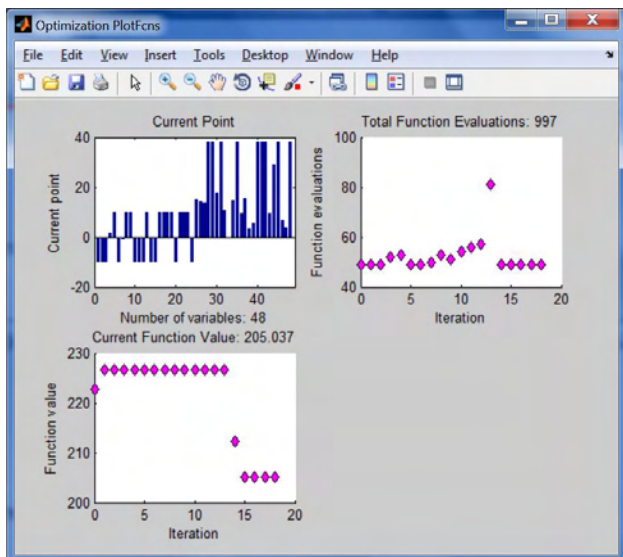


Fig. 4. Optimization result from "*fmincon*" solver

$-0.8 -0.8 -0.8 -0.8 -0.8 -1 \ 1 \ 2 \ 2]$  the objective function value is  $F = 226,717$ . Hence the optimization of battery schedule and diesel generator schedule simultaneously leads to about 10,574% reduction of necessary costs.

## V. CONCLUSION

An optimization of a battery and diesel generator schedule in a microgrid is presented in this paper. Using the solution for the schedules two goals are achieved: 1) It is guaranteed that the load demand is covered by an enough high reserve. 2) The optimization leads to 10,574% reduction of the necessary costs. At the same time the costs for the end user are essentially reduced. Minimizing the objective function, the fuel consumption by the diesel generator is minimized, as well and the harmful impact on the environment is also reduced.

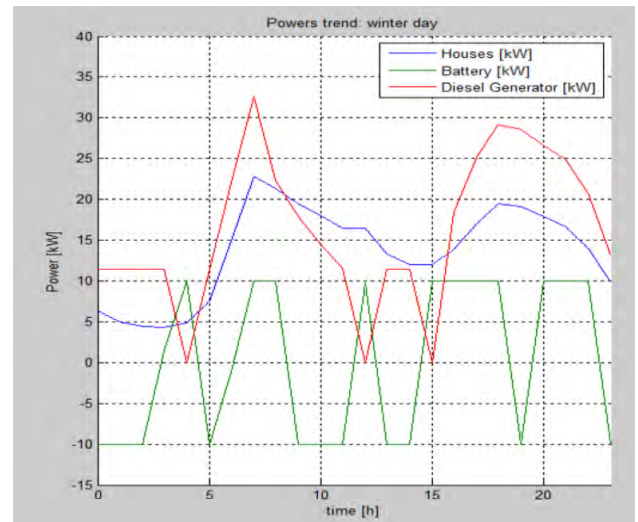


Fig. 5. Graphics of Houses consumption, optimized Battery schedule and optimized Diesel generator schedule

## REFERENCES

- [1] Al Faruque M. A., F. Ahourai : "GridMat: Matlab Toolbox for GridLAB-D to Analyse Grid Impact and Validate Residential Microgrid Level Energy Management Algorithms", IEEE PES Conference on Innovative Smart Grid Technologies (ISGT'14), Washington DC, USA, February 2014. <http://aicps.eng.uci.edu/papers/GridMat-ISGT-2014.pdf>
- [2] Marinelli M., "Testing of a Predictive Control Strategy for Balancing," *IEEE TRANSACTIONS ON SUSTAINABLE ENERGY*, 2014.
- [3] Chassin, D. P., "GridLAB-D: An agent-based simulation framework for smart grids," 13 May 2014.
- [4] Matlab, [Online]. <http://www.mathworks.com/products/matlab/>
- [5] Chang G. W, H. J. Lu, H. J. Su, "Short-term Distributed Energy Resource Scheduling for a DC Microgrid," *Energy and Power Engineering*, 2013, vol. 5, pp. 15-21.
- [6] Xiao J., L. Bai, F. Li, H. Liang, and C. Wang, "Sizing of Energy Storage and Diesel Generators in an Isolated Microgrid Using Discrete Fourier Transform (DFT)", *IEEE Transactions on Sustainable Energy*, Vol. 5, No. 3, July 2014
- [7] Dufo-Lopez R. and J. L. Bernal-Agustin, "Multi-objective design of PV-Wind-Diesel-Hydrogen-Battery systems", accepted for publication in *Renewable Energy* (<http://www.sciencedirect.com/science/journal/09601481>)
- [8] Skarstein O, Ulhen K., "Design Considerations with Respect to Long-Term Diesel Saving in Wind/Diesel Plants", *Wind Engineering* 1989; 13(2):72-87.
- [9] Luu, N. A., "Control and management strategies for a microgrid", Ph.D. Thesis, Université de Grenoble, France, 18.12.2014, <https://tel.archives-ouvertes.fr/tel-01144941/document>
- [10] Seryoatmojo H., A. A. Elbaset, Syafaruddin and T. Hiyama, Genetic Algorithm based optimal sizing of PV-Diesel-Battery System, Considering CO2 Emission and Reliability, *International Journal of Innovative Computing, Information and Control*, Vol. 6, No. 10, October 2010, 1-09-0844, ISSN: 1349-1198.
- [11] Diaf S., M. Belhamelb, M. Haddadic, A. Louchea, Technical and economic assessment of hybrid photovoltaic wind system with battery storage in Corsica Island, *Energy Policy*, 36 (2) (2008), pp. 743-754

# Assessing the Profitability of New Investment Projects on Construction of Small HPP on HPS Crn Drim

Goce Bozinovski<sup>1</sup> and Atanas Iliev<sup>2</sup>

**Abstract** – The Crn Drim basin is the richest with water from all of Macedonia. The paper gives the overview of the analyses made on practical example of the energy contribution and economical evaluation of the new planned hydro power plants of the cascade on hydro energy system of Crn Drim River in Macedonia. The existing hydro energy system consists of two cascade hydro power plants and three reservoirs, but the operation experiences and the site location give the possibility to upgrade the system with additional hydro power plants. The technical calculations and power improvement are the base platform to make the economical evaluation of the NPV, B/C, PBP values for each power plant from all over integrated hydro energy system. The geographical configuration of the terrain and hydrological conditions of the region, especially the tributaries of Radika river, allow construction of more hydropower plants. These hydro power plants would improve the energy situation in Macedonia, and also would regulate the operating mode of existing HPP.

This paper will analyze more the hydropower projects in Crn Drim basin which are analyzed and planned for long time. The results presented in the paper are the energy production and economic viability of the construction and operation in power system.

**Keywords** – Economy, Hydro Power Plant (HPP), Energy.

## I. INTRODUCTION

The main feature in the analysis of projects of hydroelectric plants is a high degree of uncertainty and imprecision of input parameters. Therefore, in order to perform an economic analysis of the projects, it is necessary to analyze multiple scenarios to determine the likelihood of certain option and meet investor risk investments of the HPP. As a challenge to show the many projects of hydroelectric determine those that are economically viable and feasible to consider those who the developer would guarantee the quickest return on investment. For this purpose will be determined and analyzed the financial performance of the planned investment of small hydropower on HS Black Drin.

The economic value of the power plants representing the following categories: NPV (Net Present Value), BCR (Benefit Cost Ratio), PBP, (Pay Back Period), IRR (Internal Rate of Return).

<sup>1</sup>Goce Bozinovski is with the AD ELEM-Skopje, Branch HES "Crn Drim"-Struga, Plostad na revolucijata bb, 6330 Struga, p.fax 83, R. Macedonia, e-mail: goce\_boz@yahoo.com

<sup>2</sup>Atanas Iliev is with the Faculty of Electrical Engineering and Information Technologies – Skopje, Karpos II, 1000 Skopje, R. Macedonia, e-mail: ailiev@feit.ukim.edu.mk

## II. BASIC PARAMETERS FOR THE SMALL HPP

This includes all projects of hydro power plants with installed capacity below 10 MW, which according to their projected energy parameters are treated as small hydropower plants with a reduced price of electricity in high tariff from 50 to 65 € / MWh and the low tariff of 45 € / MWh in correlation with preferential prices from Energy Regulatory Commission (ERC) of RM. In that group of potential small planned include: HPP Tresonce, HPP Gary, HPP Selce, HPP St. Petka and HPP Kosovrasti. These new hydropower plants are mostly tributaries of the supply to be used for filling the reservoir of potential HPP Boskov Most of Mala reka, and the existing reservoir of Debar Lake.

TABLE I  
TECHNICAL PARAMETERS FOR THE PLANNED SMALL HPP ON HYDRO SISTEM CRN DRIM

	$P_{inst}$	$W_{year}$	CF- Capacity Factor	Economic analysis
	[MW]	[GWh]		
<b>Tresonce</b>	8	21	<b>0.30</b>	Indiv. project
<b>Gari</b>	2.7	9	<b>0.38</b>	Indiv. project
<b>Selce</b>	3.3	9	<b>0.31</b>	Indiv. project
<b>Sv.Petka</b>	1.2	4	<b>0.38</b>	Indiv. project
<b>Kosovrasti</b>	5.4	18	<b>0.38</b>	Indiv. project

Expected electricity production or capacity factor depends on annual production:

$$CF = \frac{W_{year}}{8760 \cdot P_{inst}} \quad (1)$$

Where is:

$W_{year}$  - annual production of electricity,

$P_{inst}$  - installed power units.

## III. FINANCIAL INDICATORS FOR THE PLANNED SMALL HPP OF HS CRN DRIM

Small hydropower plants are have the following common input parameters: operation life of 20 - 25 years of the loan repayment of 10 years. The following table 2. presents an

overview of the economic indicators (NPV, B/C, CF,  $T_{kred}$ ,  $T_{oper}$ ,  $C_{ht}$  and  $C_{lt}$ ) of the project analyzed the Small HPP hydro system Crn Drim. This presents economic indicators after all discounted benefits and costs. Where:  $T_{kred}$  - time of repayment of the loan in years;  $T_{oper}$  - operating life of the plant, number of years of operation for economic profit;  $C_{ht}$  - work in high tariff (peak load);  $C_{lt}$  - work in low tariff (base load).

Net Present Value (NPV) is obtained as:

$$NPV = NPV(B)_{1-n} - NPV(C)_{1-n} \quad (2)$$

The B/C ratio is obtained as:

$$BCR = \frac{NPV(B)_{1-n}}{NPV(C)_{1-n}} \quad (3)$$

Where is:

$NPV(B)_{1-n}$  - benefits for the whole operation life,

$NPV(C)_{1-n}$  - costs during HPP operation life.

TABLE II  
ECONOMIC INDICATORS FOR PLANNED INVESTMENT PROJECTS OF  
SMALL HYDRO POWER PLANT ON CRN DRIM BASIN

HPP	NPV (€)	B/C	CF	$T_{kred}$ - $T_{oper}$ (year_year)	$C_{ht}$ - $C_{lt}$ (%_%)
Tresonce	2.316.40	1,21	0,30	10_25	60_40
Gari	2.435.79	1,33	0,38	10_20	60_40
Selce	1.217.89	1,33	0,38	10_20	60_40
Sv.Petka	465.89	1,10	0,31	10_20	60_40
Kosovrasti	792.99	1,45	0,38	10_20	50_50

On figure 1. and figure 2. presents graphical representations of economic indicators NPV and B/C for evaluation of new

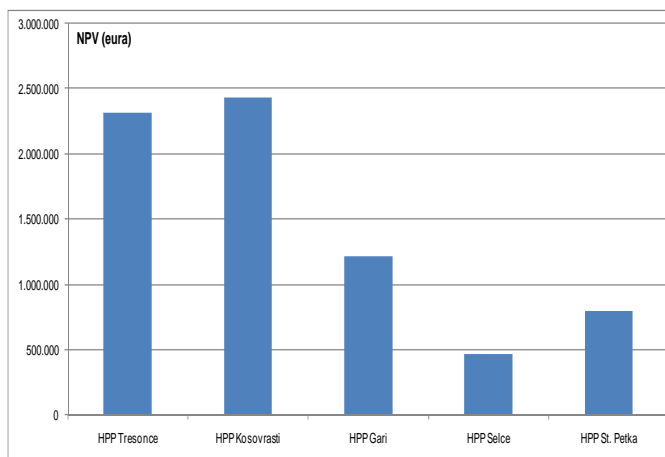


Fig. 1. NPV value for hydro projects on Crn Drim

hydro power projects HPS Crn Drim.

The table and graphical presentation clearly show which projects are the most attractive for investors, as the group of small HPP, such as: HPP Gari, HPP Kosovrasti and HPP St. Petka. These plants have the highest factors of NPV and  $B/C > 1,3$ . Fig.3 shows the operational energy parameter of CF (capacity factor) for each hydro power plant which is directly connected to the economical indicators. CF is certainly linked to economic indicators, or those projects that have high economic indicators have high energy production with CF about 0,4. The peak load power plants have low CF which is about 0,3, such as: HPP Tresonce, HPP Selce.

The following figure 3. presented is a graphic indicator of exploitation CF for all hydropower projects analyzed, which is of course directly related to the the economic indicators for each project.

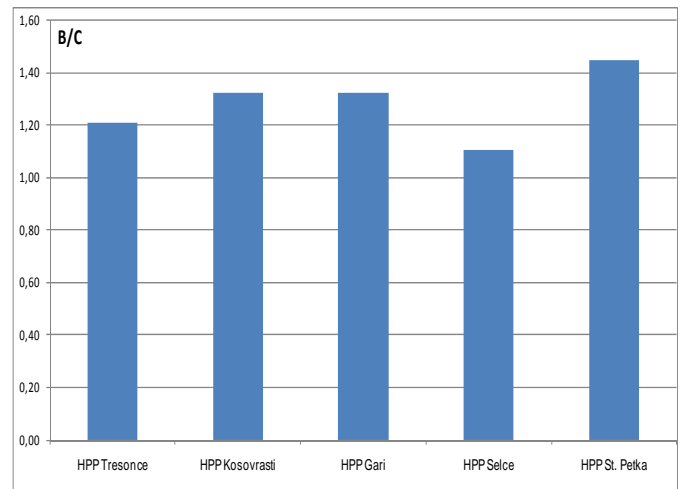


Fig. 2. B/C ratio for hydro projects on Crn Drim

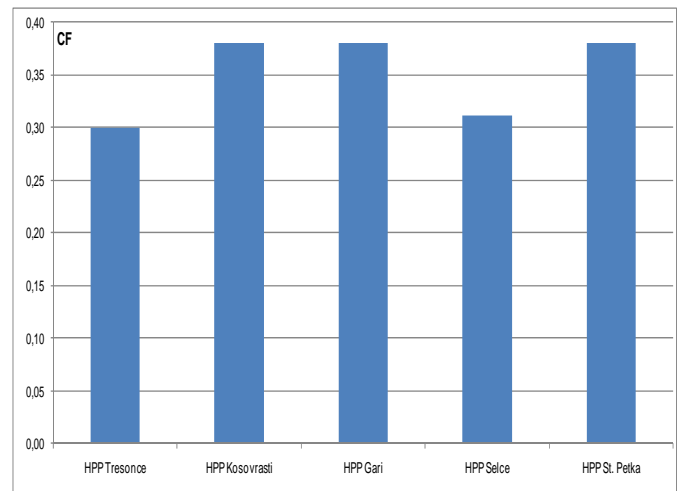


Fig. 3. CF for new small hydro projects on Crn Drim

Capacity factor CF gives the expected energy production which is strongly connected with economic indicators. HPP Kosovrasti, HPP Gary and HPP Sveta Petka have CF value about 0.38 which is actually highly acceptable value. HPP Tresonce and HPP Selece have low CF which is



approximately below 0.3 which means that they have to operate in the peak load.

#### IV. CONCLUSION

The specific natural geographical position of the Crn Drim basin with all facilities (power plants, reservoirs and supply channels ) plays an important role in the power system of Macedonia. With new hydro power facilities, the goal is to improve the energy value of the whole hydro system of Crn Drim as the followings:

- increased installed capacity,
- higher production of electricity,
- reduce spillovers,
- increased number of operation hours in peak load,
- reduced number of operational work as base load power plant.

Improving the energy value of hydropower production complex will certainly contribute to the financial benefit of the whole hydro complex. The main benefit of the system is the increased installed capacity and increased electricity production as well as reducing the spillover. Operating on the hydro power plants in the high tariff of peak load is particularly important in dry hydrology, and wet hydrological conditions can be avoided or minimized the overflows.

#### REFERENCES

- [1] Olsson, M., Söder. L.: "Hydropower Planning including Trade-off between Energy and Reserve Markets", Published in Proceedings of IEEE Bologna PowerTech 2003 June 23-26th Bologna Italy.
- [2] Amir Pasha Zanjani Nasab; Financial Analysis of Small-Hydro Power Project in Malaysia from the Investor Perspective, 2012 IPCBEE Vol.33 (2012) Singapore.
- [3] Technical documentation of hydro energy system of Crn Drim.
- [4] Carmen L.T. Borges, Senior Member, IEEE, and Roberto J. Pinto: Small Hydro Power Plants Energy Availability Modeling for Generation Reliability Evaluation, IEEE, Vol 23, No.3 August 2008;
- [5] G. Bozinovski, A. Iliev, "Energy Contribution and Economical Evaluation of Hydro Energy Complex System of Cascade Power Plants" – Renewable Energy Sources (OZE) 5th International Scientific Conference OZE 2014, High Tatras - Tatranske Matliare, May 20 – 22, 2014, Slovakia (467\_OZE\_2014, ISBN 978-80-89402-74-8).

This Page Intentionally Left Blank

# Determination of Harmonics Level in Local Electrical Distribution System

Ginko Georgiev, Silviya Letskovska, Kamen Seymenliyski and Pavlik Rahnev

**Abstract** – In this work the calculated results for harmonics level are presented. The results are for local electrical distribution system – company “Plastic Products” – the town of Sredetz – Bulgaria. The experimental investigation is done. The goal of this determining the reason for damages and to propose variants of technical – economical decision for solving of the arising problems.

**Keywords** – Power factor, THD, Harmonics.

## I. INTRODUCTION

The character of production line needs the usage of power nonlinear (electrically) charges.

This leads to generation of harmonics in electrical distribution (ED) system of the enterprise [1-3].

There is difficultness for compensation of power factor and that is why the losses in the network go higher (increase).

As a result there is very often damage of compensation devise.

Very often these are its capacitor batteries.

The problems with similar character were revealed in electrical distribution system of “Plastic Products” - town of Sredetz, Bulgaria.

With the lack of possibility compensation devise to be introduced in normal exploration mode, the company pays significant taxes to Electrical Distributing Companies - EVN. It is due to low values of power factor.

The additional financial losses are realized and from increased electrical losses, depending on circulated harmonics in ED-system.

It was necessary the investigation of electrical distribution system to be performed.

It was done using power analyzer, which gives a possibility to receive data as for powers (active “P”, passive “Q”), as well as for power factor  $\cos\varphi$ .

Some of the parameters for the quality of electrical energy could be determined using coefficients THD (total harmonic distortion) by voltage and current and amplitude – frequency spectrum of the harmonics [4-5].

<sup>1</sup>Ginko Georgiev - Burgas Free University, San Stefano 62, 8000 Burgas, Bulgaria, E-mail: ginkoele@abv.bg

<sup>1</sup>Silviya Letskovska - Burgas Free University, San Stefano 62, 8000 Burgas, Bulgaria, E-mail: silvia@bfu.bg

<sup>1</sup>Kamen Seimenliyski – Burgas Free University, San Stefano 62, 8000 Burgas, Bulgaria, E-mail: silvia@bfu.bg

<sup>2</sup>Pavlik Rahnev – Technical College, As. Zlatarov University, Y. Yakimov 1, Burgas 8010, Bulgaria, E-mail: pavlikrahnev@abv.bg

## II. DETERMINATION OF THE COEFFICIENT OF VOLTAGE NONSINUSIADILITY ON THE BUS OF SUBSTATION

### 1.1. Short description of the object.

The industrial unit “Polystirol” is supplied from the transformer TM-630 with short circuit voltage  $u_{sc}=6\%$ .

The short circuit power on the primary side of the transformer is  $S_{sc}=25$  MVA.

From the busses of the substation are supplied:

- Current rectifier connected in “Larionov” with power 120 kW and working with  $\cos\varphi=0,7$ ;
- Five asynchronous motors with next characteristics:
  - nominal power 22 kW;
  - nominal power factor  $\cos\varphi_n = 0,92$ ;
  - multiplicity of starting moment  $t_s = 1,2$ ;
  - multiplicity of starting current  $k_s = 6$ ;
  - efficiency with nominal charge  $\eta_n = 0,92$ ;
  - nominal slip  $s_n = 2,66\%$ ;
  - ratio between power losses in stator and total losses in nominal load  $\gamma = 0,3$ .

Average monthly  $\cos\varphi$  of industrial unit “Polystirol” is 0,7.

### 1.2. Methodology for analytical determination of the THD and capacitor's power.

- Creating of equivalent circuit for the fifth harmonic for one phase, as a thyristor rectifier is replaced with equivalent current source, but power transformer, asynchronous motors and supplying system are replaced with their resistances – (Fig. 1).

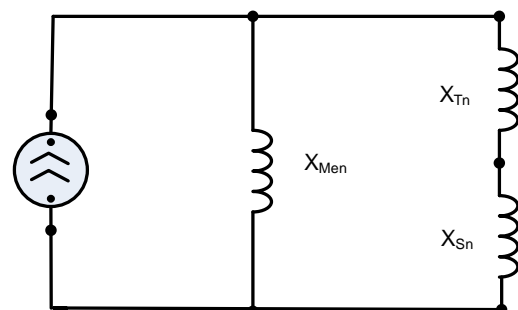


Fig. 1. Equivalent circuit.

- Determination element reactances in the electrical power supplying circuit [6]:

- *System reactance:*

$$x_{S_n} = k_x \frac{U^2}{S_{S_{cs}}} n = 4,16 \times 10^{-3} n \quad (1)$$

Where:

- U – voltage, U = 400 V;
- S<sub>S<sub>cs</sub></sub> - short circuit power;
- n - harmonic number;
- k<sub>x</sub> – coefficient, k<sub>x</sub> = 0,65.

- *Transformer reactance:*

$$x_{T_n} = \frac{k_x u_k U^2}{S_T} n = 0,013409 n \quad (2)$$

Where:

- k<sub>x</sub> – coefficient, k<sub>x</sub> = 0,88;
- S<sub>T</sub> – transformer Apparent power;
- U – nominal voltage of the motors (V).

- *Equivalent reactance of five asynchronous motors:*

$$x_{Me_n} = \frac{1}{5} k_x \frac{U^2 \cos \phi_n}{k_s P_n} \sin \phi_s n = 0,1647 n \quad (3)$$

Where:

- k<sub>x</sub> – coefficient, k<sub>x</sub> = 0,78;
- P<sub>n</sub> – rated power;
- k<sub>s</sub> – ratio between starting current I<sub>start</sub> and rated current I<sub>n</sub>;  $k_s = I_{start} / I_n$ ;
- sinφ<sub>s</sub> – starting sinφ;
- sinφ<sub>n</sub> = 0,9469. It is determined in dependence on power factor in the start of motor;

$$\cos \varphi_S = \cos \varphi_n \left[ \frac{t_s}{(1 - S_n) k_s} + k_s (1 - \gamma) \right] = 0,3215 \quad (4)$$

- *Total equivalent reactance for relevant harmonic:*

$$x_{en} = x_{Me_n} \frac{(x_{S_n} + x_{T_n})}{(x_{Me_n} + x_{S_n} + x_{T_n})} n \quad (5)$$

The sequence number of the harmonic of Larionov rectifier are 5, 7, 11, 13, 17, 19 etc.

- Determining of current of the base harmonic, which is determined by the equation:

$$I_n = \frac{I_1}{n} \left[ \frac{\sin\left(\frac{n\gamma}{2}\right)}{\frac{n\gamma}{2}} \right] \quad \gamma = 12^\circ (0,2094 \text{ rad}) \quad (6)$$

Where:

- I<sub>n</sub> – relevant harmonic current;
- I<sub>1</sub> – fundamental harmonic current;
- γ – thyristor commutation angle.

I<sub>1</sub> is a current of the base harmonic, which determined by the equation:

$$I_1 = \frac{P_n}{\sqrt{3} U_n \cos \varphi} = 247,44 \text{ A} \quad (7)$$

- Determining of voltage of the n- harmonic:

$$U_n = I_{e_n} x_{e_n} \quad (8)$$

- Determining of coefficient THD:

$$THD_n = \sqrt{3} k_1 k_2 \frac{\sqrt{\sum_{n=5}^{19} U_n^2}}{U_1} 100 = 3,96\% \quad (9)$$

- k<sub>1</sub>, k<sub>2</sub> - correction coefficients.

- Determining of the necessary compensation capacitive power:

For a moment load of the industrial unit is:

$$Q_{cap} = P (tg \varphi_1 - tg \varphi_2) = 180 \text{ kVA} \quad (10)$$

In this case:

$$x_{cap} = 2,66 \Omega \quad (11)$$

Comparing the calculating inductive reactance with capacitive it is seen that is possible appearance for the 13 – th harmonic.

### 1.3. Experimental results.

Using power analyzer *Circutor* are determinate:

- Coefficient total harmonic distortion (THD). The derived result is closed as a value to the theoretical calculation (Fig. 2).



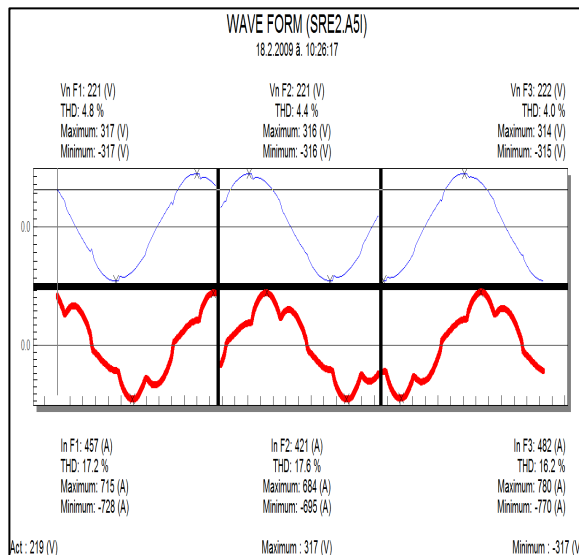


Fig. 2.

- Harmonious composition of the current (Fig. 3,a and Fig.3,b):

Phase current3		
Irms (A):408		THD (%):16.2
Fundamental (A):482		Disphase (°):85.7
Harmonic	Amplitude (%)	Disphase (°)
7 (-)	3.143	64.3
8 (-)	0.136	119.3
9 (+)	0.332	265.1
10 (-)	0.101	120.1
11 (-)	3.798	322.0
12 (+)	0.210	287.4
13 (+)	0.366	128.9
14 (+)	0.075	196.5
15 (-)	0.276	309.9
16 (-)	0.116	145.4
17 (+)	1.792	358.0
18 (-)	0.110	331.3
19 (-)	0.218	276.3
20 (-)	0.085	234.4
21 (-)	0.191	340.9

Fig 3,a.

Phase voltage3		
Vrms (V):222		THD (%):4.0
Fundamental (V):222		Disphase (°):127.1
Harmonic	Amplitude (%)	Disphase (°)
7 (-)	0.535	329.8
8 (-)	0.107	211.0
9 (+)	0.086	128.5
10 (-)	0.070	82.4
11 (-)	1.172	110.5
12 (+)	0.067	67.3
13 (+)	0.078	211.4
14 (+)	0.055	17.0
15 (-)	0.155	298.4
16 (-)	0.044	207.5
17 (+)	0.763	249.4
18 (-)	0.060	213.9
19 (-)	0.120	50.5
20 (-)	0.055	145.0
21 (-)	0.132	80.3

Fig. 3,b.

From the Figure 2 is clear that the biggest energy influence of the current harmonics as disturber 5, 7, 11, and 17 harmonic.

Although that 13-th harmonics has not big weight only 0,366% from the base, it is the main factor for arising resonances.

### III. CONCLUSION

The investigations performed onto the work of the electrical energy system of the local company "Plastic Products" give the reason to make a decision, that there is a big pollution with harmonics.

They are generated from the part of the technological electrical equipment, used in the production of plastic material.

The equipment has strongly expressed nonlinear character as an electrical charge.

The problems, which arise in the work of the compensation capacitor device for increasing  $\cos\phi$  are as a result of the resonance current processes.

They arise between the capacitors of compensation device and inductances as elements of electric network.

As a result the capacitive batteries of the compensation device is overload by current and from this by power over the limited levels.

This is the reason for their damages in going out from the working condition.

The calculation which are done are closed to the experimentally received results.

The failure of the capacitor batteries is connected with the arising current resonance phenomena as a result of the harmonics and following current and from this power overloads over limited levels.

It is proofed the possible resonance for the 13 -th current harmonics.

Dissolving the existing problem is connected with dissolving of the problem with the pollution of electrical supplying system with harmonics.

This can be reached in the way their filtering or changing of the resonance areas.

### REFERENCES

- [1] Ion Boldea, Syed A. Nasar, The induction machines design handbook - 2nd ed., ISBN 978-1-4200-6668-5 (hardcover: alk. paper), 2002.
- [2] Klimov V. P., Moskalev A. D., Harmonics problems in modern power systems // Practical power electronics. - 2003. - № 6.
- [3] Zhezhelenko I. V., Higher harmonics in the power supply systems of industrial enterprises - M.: Energoatomisdat, 375 p., 2010.
- [4] Zhezhelenko I. V., V. Bozhko, Rabinovich M. L., Quality elektroenergies in industrial plants. - Kiev: Vishcha School, 1986. - 120 p.
- [5] Alexandrov, D. S., The reliability and quality of power supply enterprises: study, A 46 Benefit / DS Alexandrov, EF Scherbakov.- Ulyanovsk: Ulstu, 2010 - 155 p.
- [6] Fyodorov Anatoly, Vassilev N., S. Siderov, Electricity of industrial, Sofia: Technique 1979.

This Page Intentionally Left Blank

# Model study of Lighting Protection of 110/20 kV Substation

Margreta Vasileva<sup>1</sup> Neli Velikova<sup>2</sup> and Danail Stanchev<sup>3</sup>

**Abstract** – The purpose of the current report is to be examined a model study of the influence of the lightning strokes, which occur in the electric distributional lines, on the electrical equipment of the substation and on the protective effect of the metal-oxide surge arresters (MOSA).

**Keywords** – Lightning strokes, Metal-oxide surge arresters, Protective effect

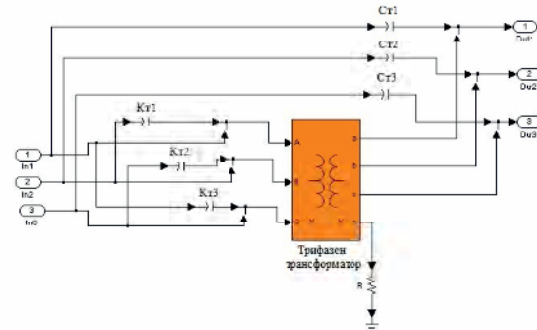


Fig. 1 Model of a power transformer for the examination of wave processes

## I. INTRODUCTION

The electrical equipment in the electrical substations for middle and high voltage, which are connected to the overhead power lines, are protected of lightning strokes through metal-oxide surge arresters [1]. The type and place of the connection of the protective means for surges are in immediate connection with the coordination of the isolation of the equipment and their stable work.

## II. SIMULATIONAL MODEL OF A POWER TRANSFORMER 110/20kV

The input data for the model of the power transformer in MATLAB SIMULINK must be in relative units in relation to the nominal data of the equipment. The following data should be entered: active resistance R1 [o.e.] and R2 [o.e.] and the self-induction of the coils L1 [o.e.] and L2 [o.e.]; the active resistance Rm [o.e.] and the self-inductance Lm [o.e.] in the magnetizing branch of the substitute circuit. The standard block for modeling of the power transformer is being used for the research of the stationary modes of work.

By the examination of the wave processes is necessary to be considered the influence of the longitudinal  $K_T$  and the transversal capacitances  $C_T$  of the coils. Therefore is formed a model scheme of the power transformer in SIMULINK, which could be used in the examination of the wave processes. Model scheme of a power transformer for the examination is shown on figure 1 [2].

The mathematical equations for the definition of the parameters of the substitute circuit of the power transformer are [4]:

$$z_T = z_k = \frac{U_k \%}{100} * \frac{U_{H.HOM}^2}{S_H}, \Omega \quad (1)$$

$$R_T = \Delta P_k * \left( \frac{U_{H.HOM}}{S_H} \right)^2 ; \Omega \quad (2)$$

$$X_T = \sqrt{Z_T^2 - R_T^2} ; \Omega \quad (3)$$

$$G_T = \frac{\Delta P_{n.x}}{U_{H.HOM}^2} ; S \quad (4)$$

$$B_T = \frac{I_{\mu} \%}{100} * \frac{S_H}{U_{H.HOM}^2} ; S \quad (5)$$

$$Z_T = Z_k = R_T + jX_T \quad (6)$$

$$Y_T = G_T - jB_T \quad (7)$$

$$Z_m = \frac{1}{Y_T} \quad (8)$$

$$Z_H = \frac{U_H^2}{S_H} \quad (9)$$

$$R_m = \text{real}(Z_m) \quad (10)$$

<sup>1</sup>Margreta Vasileva is with the Faculty of Electrical power engineering at Technical University of Varna, 1 Studentska str., Varna 9010, Bulgaria, E-mail: [m.vasileva@tu-varna.bg](mailto:m.vasileva@tu-varna.bg)

<sup>2</sup>Neli Velikova is with the Faculty of Electrical power engineering at Technical University of Varna, 1 Studentska str., Varna 9010, Bulgaria, E-mail: [nely.velikova@gmail.com](mailto:nely.velikova@gmail.com)

<sup>3</sup>Danail Stanchev is with the Faculty of Electrical power engineering at Technical University of Varna, 1 Studentska str., Varna 9010, Bulgaria, E-mail: [dstanchev1990@gmail.com](mailto:dstanchev1990@gmail.com)

$$X_m = \text{imag}(Z_m) \quad (11)$$

$$R_{m.o.e} = \frac{R_m}{Z_H} \quad (12)$$

$$X_{m.o.e} = L_{m.o.e} = \frac{X_m}{Z_H} \quad (13)$$

$$R_{T.o.e} = \frac{\Delta P_k}{S_H} \quad (14)$$

$$R_{1.o.e} = R_{2.o.e} = \frac{R_{T.o.e.}}{2} \quad (15)$$

$$X_{T.o.e} = \frac{X_T}{Z_H} \quad (16)$$

$$L_{1.o.e} = L_{2.o.e} = X_{1.o.e} = X_{2.o.e} = \frac{X_{T.o.e.}}{2} \quad (17)$$

$$\sigma = \sqrt{\frac{C_T}{K_T}} \quad (18)$$

$$C_{ex} = \frac{C_T}{\sigma} \quad (19)$$

$S_H$  – rated power of PT, VA;

$U_{H.HOM}$  – Secondary rated voltage, kV;

$U_H$  – Primary rated voltage, kV;

$\Delta P_{n.x}$  – no-load losses, kW;

$\Delta P_k$  – short circuit losses, kW;

$I_\mu$  % - magnetizing current, A;

$U_k$  % -short circuit voltage;

$\sigma$  - pulse factor of the coil

(for disc coils  $\sigma = 10 \div 20$ , for interwoven coils

$\sigma = 3 \div 5$ );

$C_T$  - transversal capacitance of the coil of CT;

$K_T$  – longitudinal capacitance of the coil of CT;

$C_{bx}$  – input capacitance of the coil ( $C_{bx}=800$  pF for CT 20/0,4 kV;

$C_{bx}=2500$  pF for CT 110/20 kV ) [6].

### III. VARIANT STUDIES

III.1. Research on the effect of the lightning strokes, which occur in the electrical distributional lines, on the electrical equipment in substation 110 kV without MOSA.

The voltages in the three phases are being controlled in the place of strike of the lightning stroke and on the power transformer. A model scheme of the substation is shown on figure 2 [3].

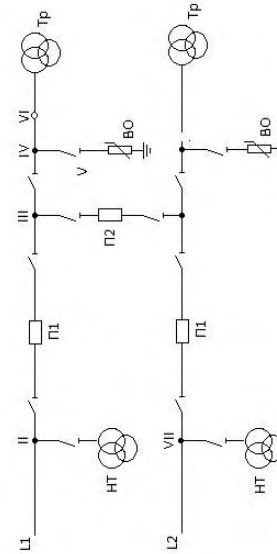


Fig. 2 Single line diagram of a substation

A. Strike of a lightning stroke on a phase conductor of power line L1.

The considered case is of a strike of a lightning stroke on phase B of an overhead power line L1. The struck section is at a distance of 100 m away from the substation. The power line L2 is connected to the system. The results of the research are given on figure 3.

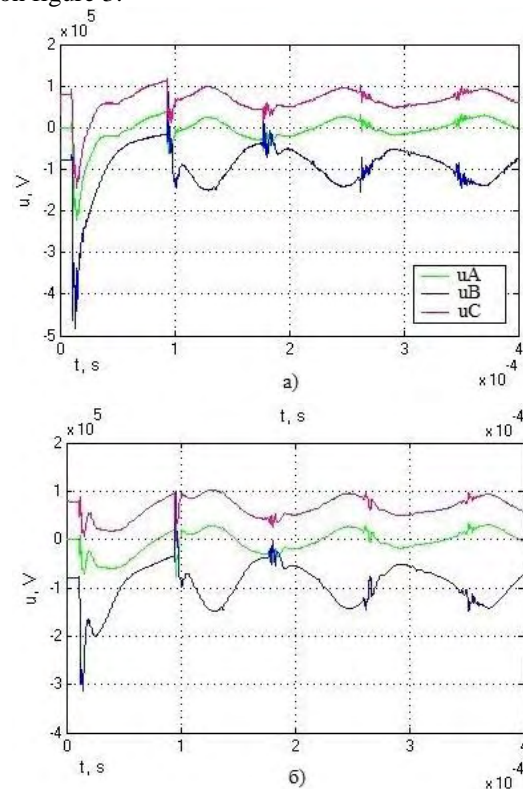


Fig. 3 Voltages in the case of strike of a lightning stroke on a phase conductor of a power line L1 without MOSA in an electrical system:

a) in the place of the strike; б) by the power transformer;



B. Strike of a lightning stroke on a phase conductor on power line L2.

The considered case is of a strike of a lightning stroke on phase B of an overhead power line L2. The struck section is at a distance of 100 m away from the substation. Power line L1 is connected to the system. The results of the research are given on figure 4.

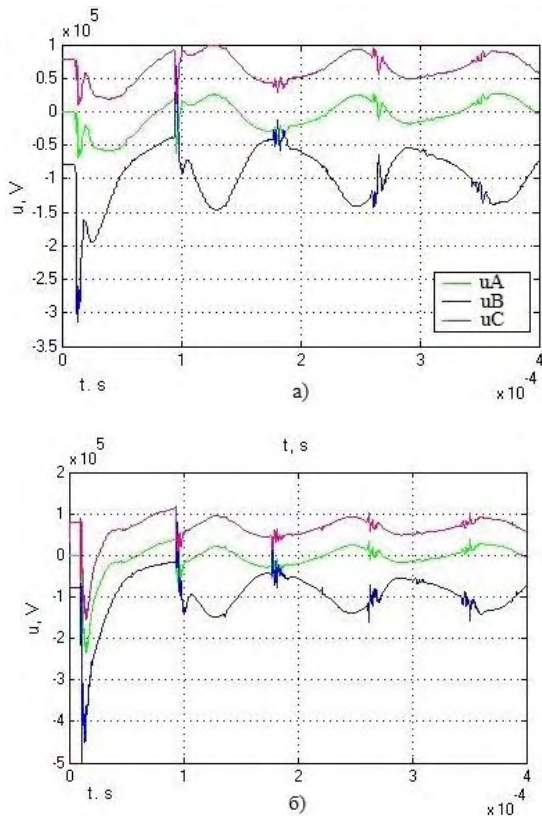


Fig. 4 Voltages in the case of strike of a lightning stroke on a phase conductor of a power line L2 without MOSA in an electrical system: a) by the power transformer; б) in the place of the strike;

The voltages, influencing the insulation of the power transformer, in case where it's not being protected, exceed its insulation level. Therefore they are dangerous for its insulation and should be restricted under 470 kV.

III.2. Research of the protective effect of MOSA in case of influence of the lightning strokes, which occur in the power lines, on the electrical equipment in substation 110 kV. Situations A and B are considered from p. III.1., but in case of existence of MOSA in electrical system. The results are given on fig. 5 and fig. 6.

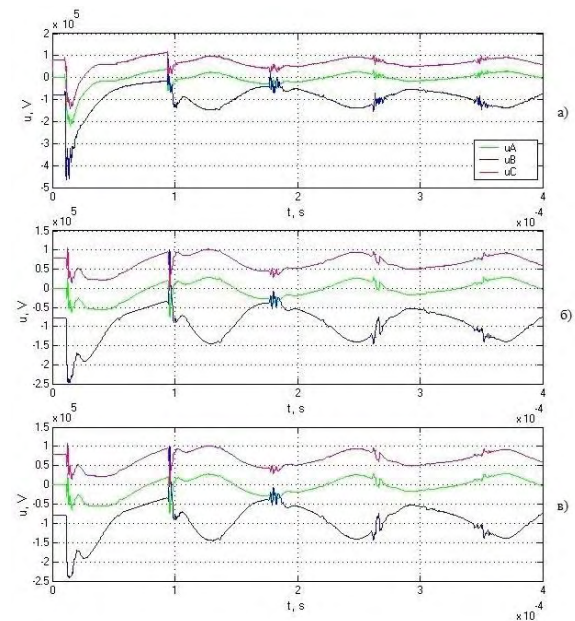


Fig. 5 Voltages in case of strike of a lightning stroke on a phase conductor of power line L1 with MOSA in an electrical system: a) in the place of the strike; б) by the MOSA; в) by the power transformer

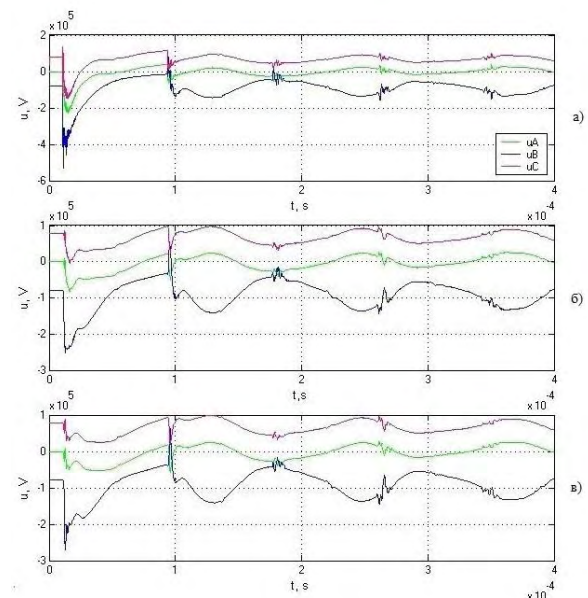


Fig. 6 Voltages in the case of strike of a lightning stroke on a phase conductor on a power line L2 with MOSA in an electrical system: a) in the place of the strike; б) by the MOSA; в) by the power transformer;

## IV. CONCLUSION

The simulation model of a power transformer, given in the report, can be used for a research of wave processes in the electrical substations.

The developed model provides precise data for research of the protective effect of the metal-oxide surge arresters.

## REFERENCES

- [1] Regulation 3, Structure of electrical systems and power lines, 2004 г.
- [2] Vasileva M., High Voltage Technics, Varna: TU-Varna, 2007.
- [3] Goldstein, M.E., Y. Korovin, AV Prokudin. Lightning overvoltages and switching surges in electrical systems. 2010
- [4] Gerasimov K., J. Kamenov, Modelling in power systems, Sofia: 2007..
- [5] Genov L. G., Wave processes in the coils of the transformers, T., S., 1974.
- [6] Gindullin F. A., Overvoltages in substation 6- 35 kV, M., Energoatom, 1989.

# Calculation of parameters of Three-Phase Induction Motor with Double Squirrel Cage with FEM 3D

Blagoja Arapinoski<sup>1</sup>, Mirka Popnikolova Radevska<sup>1</sup>, Vesna Ceselkoska<sup>1</sup>, Mile Spirovski<sup>1</sup> and Metodija Atanasovski<sup>1</sup>

*Abstract – The paper will be presented part of a research done over the model of three-phase induction motor with double squirrel cage. Described are methods for determining the parameters, and the emphasis is on the calculation using the three-dimensional magnetic field is obtained by the method of finite elements in the 3D domain. Made more calculations as follows: the equivalent inductance of the machine depending on the load currents and angular position of the rotor relative to the stator, and calculating their leakage reactance of certain parts of the machine as well as the calculation of the electromagnetic torque. The obtained results will be presented in tables and graphics.*

*Keywords – Three-Phase induction motor with double squirrel cage, FEM 3D, electromagnetic analysis, Maxwell 14- 3D Design.*

## I. INTRODUCTION

There are few determination methods for parameter for induction motor with double cage and they are:

- Classical (analytics) methods for determination of the parameters. Under these methods that are given in every vocational literature that is used for research and designing of electrical machines, parameter are determine bulky empirical equations obtained with idealization of electromagnetic processes in the engine. The calculations are not simple and the calculation itself is linked to the need for reading of a number of basic data diagrams and charts obtained by experimental measurements.
- Modern (numerical) methods for determination of parameters. Determination of the parameters of the motor with computer and adequate software package that in the basic has algorithms that can calculate certain characteristic magnetic magnitudes of the electric machine.
- Experimental determination of parameters. Is done in laboratory for testing of electric machines and can be done true indirect and direct method. This kind of determination required equipped laboratory and adequate machine that will be tested. In addition, the possibility to make a mistake in measuring and read is greater. Therefore numerical methods stand out as the most precise determination of the parameters.

<sup>1</sup>Blagoja Arapinoski, Mirka Popnikolova Radevska, Vesna Ceselkoska, Mile Spirovski and Metodija Atanasovski are with the Faculty of Technical sciences at University of Bitola, st. Makedonska Falanga 33, Bitola 7000, Macedonia.

As research subjects from which a portion is presented in the labor below, is selected three-phase asynchronous motor [1] with double cage and with the following nominal data:

$$P_n = 3.5kW, U_n = 240V, f = 50Hz$$

$$2p = 4, \cos \varphi = 0.85, \eta = 84\%, \text{ and } \Delta \text{ winding connection.}$$

Implemented's software package Ansoft Maxwell 14. 3D Design, which is made in a suitable model of the engine which is subject in this research and is obtained allocation of the magnetic field. Knowing the allocation of the magnetic field in three dimensional domain of induction motor with double cage, or the results of the magnetic vector potential in all nodes of a network of finite elements are sufficient for determining the parameters.

## II. CALCULATION OF EQUIVALENT INDUCTANCE DEPENDING OF CURRENTS LOAD AND ROTOR ANGULAR POSITIONS IN RELATION TO THE STATOR

The equivalent inductance between the stator and rotor is possible to determine if there are data for value of total co-energy  $W$  of magnetic field, which accumulate in the air gap in the induction motor with double cage. The term which is calculated co-energy of magnetic field, according to [1], [8], [9] by volume vector of current density  $J$  and the value of the magnetic vector potential  $A$ , in air gap of the engine, is given by:

$$W = \frac{1}{2} \int_V (J \cdot A) dV \quad (1)$$

The expression (1) is compared with known ratios to calculate the basic magnetic flux according to [1], [8], [9], which is determined by the density of the current loads of windings and magnetic vector potential is given with:

$$\Psi = \frac{1}{I} \int_V (J \cdot A) dV \quad (2)$$

After completion of processing, solving the relevant equations that equating model of induction motor with double cage into databases of outputs are available with the values of the magnetic vector potential, obtained after several iterative repetitions. When substituting the expression (2), in the expression (1) to calculate the total air gap co-energy is used expression:

$$W = \frac{1}{2} \Psi \cdot I \quad (3)$$

Because during a simulation to calculate the magnetic vector potential, excited current in the engine is kept constant, the expression of the basic magnetic flux can be written as the product of inductance and current:

$$\Psi = LI \quad (4)$$

According to the last expression (4), determining the equivalent inductance of the induction motor with a double cage is:

$$L = \frac{2W}{I^2} = \frac{\Psi}{I} \quad (5)$$

The total magnetic flux in the air gap of the induction motor with a double cage  $\Psi_\delta$ , depends on the currents of the motor load and position of the rotor relative to the magnetic axis of the stator. Therefore the value of the equivalent inductance of the motor depends on the same size, and that the expression that is calculated is given with:

$$L_\delta = \frac{\Psi_\delta(\theta, I)}{I} \quad (6)$$

The determination of the basic magnetic flux in the air gap of the induction motor with double cage in Maxwell 14 - 3D Design, has already been done and it made more simulations for analysis of electromagnetic processes in the engine. It is received the distribution of basic magnetic flux in the air gap of the induction motor with double cage for seven arbitrary operating points of the engine. The simulations were performed at rotation of the rotor relative magnetic axis to the stator from  $0^\circ$  to  $90^\circ$  mechanical, with mechanical step  $7.5^\circ$  of mechanical rotation. That is, for each operating point of making a pole twist of four full asynchronous motor with double cage. Term of points under which is made simulation electromagnetic analysis is given below:

Table 1 Description of operating points at which it is made simulation analysis

Work point	$s(\%)$	$P_1[W]$	$P_2[W]$	$I_1[A]$	$I_D[A]$	$I_G[A]$	$\cos \varphi$	$\eta(\%)$
1.	0.1	416	33	2.11	4.90	1.84	0.27	7.8
2.	0.7	120	804	2.69	33.9	12.8	0.62	66
3.	1.3	196	152	3.55	62.1	23.4	0.76	75
4.	1.9	268	219	4.51	89.2	33.8	0.82	81
5.	2.5	336	279	5.48	115	43.9	0.85	83
6.	3.3	419	349	6.74	147	56.7	0.86	83
7.	3.7	456	380	7.35	162	62.9	0.86	83

In Table 1 are given values for each operating point of: overrun percentage  $s(\%)$ , the power that the engine taken from the network  $P_1[W]$ , the power that as useful develops axle  $P_2[W]$ , stator coil currents  $I_1[A]$ , currents that flow in

the lower  $I_D[A]$  and currents that flow in upper cage rotor  $I_G[A]$ , power factor and coefficient of efficiency  $\eta(\%)$ .

It notes that five points for load below headline, point number six is at rated load and point seven is greater than the nominal load. In each analysis, each set point change is required loads in the model engine for each position of the rotor relative to the stator. The time required for the analysis of each operating mode is 9h 9min 43sek. In this network the generated finite element is identical for each individual computation [1].

The results of calculation are stored in a table (because of the sheer volume is represented), from which follows the graphic interpretation of the dependence of the equivalent inductance of the motor in function of the angular position of the rotor relative to the magnetic axis of the stator, for all seven examined setpoints and it is represented in Figure 1. Figure 2 shows the dependence of the equivalent inductance of the induction motor with double cage of load currents for each of the seven examined set points at constant position of the rotor relative to the stator magnetic axis.

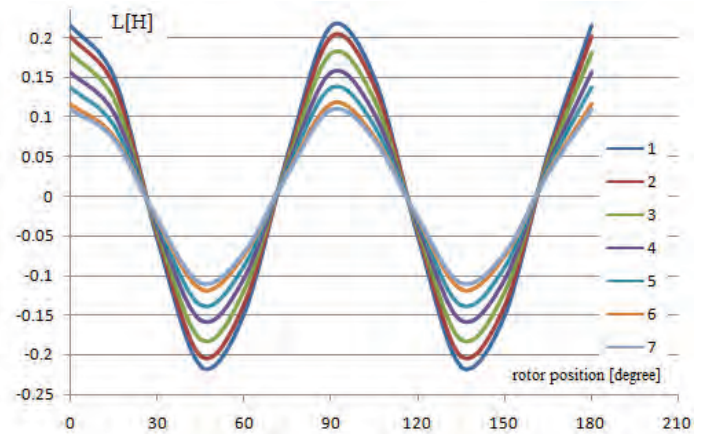


Fig.1. Equivalent inductance of air gap of induction motor with double cage at different angular positions of the rotor relative to the stator magnetic axis

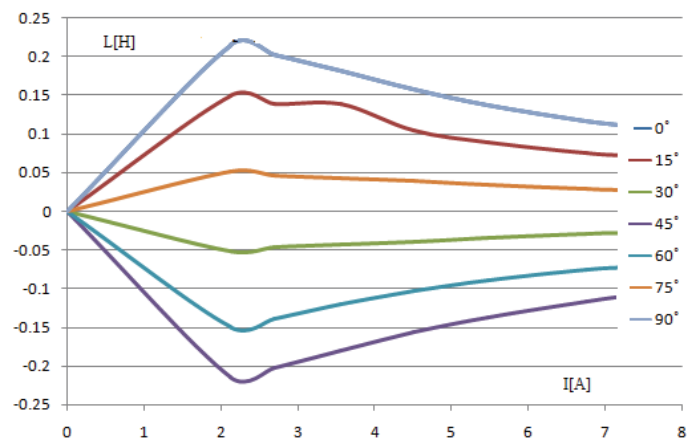


Fig. 2. Equivalent inductance of the induction motor with double cage for various load currents and constant position of the rotor relative to the stator magnetic axis



From Figure 1, it may be noted that the equivalent inductance monitors changes of magnetic flux in the air gap, and that at constant current loads have the greatest value when the rotor is facing with its magnetic axis. In particular rotation of rotor, the inductance value falls close to zero when the rotor act de magnetization or reaches a maximum value when the directions of exited flux of the stator and rotor coincide. Also, the feature is noted that the change in the equivalent inductance is greater when burdens stator coils are smaller and the engine operates in the linear part of the characteristic of magnetization. From Figure 2, it may be noted that the equivalent inductance less change with increasing load, so that in loads that are close to the nominal changes are more pronounced in the function of the position of the rotor. The impact of the load on the value of the equivalent inductance is minimized because of saturation of the magnetic circuit of stator and rotor.

### III. CALCULATION OF THE LEAKAGE REACTANCE

Own reactance of one phase of stator coil of induction motor with double cage is determined by the value of the magnetic vector potential in the three-dimensional domain of the engine when the nominal phase current flowing through only one phase winding example: phase A. Screening is the value of the total magnetic flux in the air gap to the engine under one magnetic pole. Use the value of the magnetic flux at the rated load which is:  $\Psi_{AA} = 0.792 \text{ Vs}$ , accordingly own inductance of stator phase is:  $L_{AA} = \Psi_{AA} / I_{inf} = 0.117 \text{ H}$ , while the stator reactance is:  $X_{AA} = \omega L_{AA} = 36.74 \Omega$ .

Leakage reactance of stator coil is the sum of the reactance of channel dissipation of magnetic flux in the active part of the conductor and the reactance of leakage magnetic flux in the top links:

$$X_{\sigma 1} = X_{k1} + X_{cv1} \quad (7)$$

In determining the reactance of break stator coil, according to the analytical empirical equations, the expression (7) is complemented by another set to leakage reactance of differential leakage up. It is the result of the existence of harmonics in the induced voltage of the stator coil. The magnetic flux distribution of mainly depends on the dimensions of the channels of stator magnetic circuit and performance of the coil. In determining the leakage reactance of the stator in the paper, the influence of one of the differential dissolution is taken into account in the scattered flux channels and one part of break flux in the frontal connections.

In this research, calculations are made using the values of leakage flux previously defined as: at rated load of the induction motor with double cage, leakage channel magnetic flux is  $\sum \Delta \Psi_{k1} = 0.0335 \text{ Vs}$ . Inductance of the channel dispersed magnetic flux has a value:  $L_{k1} = 0.00496 \text{ H}$ .

Reactance of channel dissipation of magnetic flux can be calculated from the known expression:  $X_{k1} = \omega L_{k1} = 1.56 \Omega$ .

To determine the reactance of break magnetic flux in front connections use the result given earlier. Analogously to the inductance of the prior break flux in the frontal connection

and reactance effect:  $L_{k1} = \frac{\sum \Delta \Psi_{cv1}}{I_{inf}} = 0.00611 \text{ H}$ . The

total value of the reactance break stator coil is:  $X_{\sigma 1} = X_{k1} + X_{cv1} = 3.48 \Omega$ .

Leakage reactance of the lower cage rotor  $X_{\alpha l}$ , is a sum of break reactance of the active part of the lower cage (the area covered by the rods of the lower cage) - and break reactance of the front links (ring of the lower cage) -  $X_{\alpha ls}$ , and leakage reactance from frontal connections (ring of the lower cage) -  $X_{\alpha lp}$ . For calculating of the leakage reactance of the active part of the lower cage of the rotor is using the value of leakage magnetic fluxion the rotor channels in the lower cage  $\sum \Delta \Psi_{2kd} = 2.57 \cdot 10^{-4} \text{ Vs}$ . The total value of leakage reactance of lower cage of the rotor is given by:  $X_{\alpha lk} = X_{\alpha ls} + X_{\alpha lp} = 0.000589 \Omega$ . The total value of leakage reactance of the lower cage of the rotor toward stator is:  $X_{\alpha lk}' = X_{\alpha lk} k_{12} = 9.8 \Omega$ .

For determining of the leakage reactance of the upper cage of the rotor is conducted identical procedure as in determining the leakage reactance of the lower cage. Are used already defined values of leakage magnetic flux of the active part of the upper cage (channels accommodating sticks) and magnetic flux in the ring (front link) which makes electrical connection of the active parts of the cage. Similar to the lower cage and in this case, the total leakage reactance in the upper cage  $X_{\sigma g}$ , is

a sum of leakage reactance of the rods of the upper cage  $X_{\sigma gs}$  and ring  $X_{\sigma gp}$ . Of the total leakage magnetic flux of channel

from the upper cage which amounts  $\sum \Delta \Psi_{2kg} = 0.26 \cdot 10^{-5} \text{ Vs}$ , are determined inductance and leakage reactance of the active part of the

$$\begin{aligned} L_{\sigma gs} &= \Delta \Psi_{2kg} / I_{2gk} = 0.045 \cdot 10^{-6} \text{ H}, \\ \text{upper cage: } X_{\sigma gs} &= \omega L_{\sigma gs} = 0.0000144 \Omega \end{aligned}$$

In the determination break reactance of the ring of upper cage rotor is used break flux in the ring of the upper cage  $\sum \Delta \Psi_{2pg} = 0.76 \cdot 10^{-5} \text{ Vs}$ , and analog in the sticks

calculation is based on:  $L_{\sigma gp} = 0.000000133 \text{ H}$ ,  $X_{\sigma gp} = \omega L_{\sigma gp} = 0.0000421 \Omega$ . Leakage reactance of the

upper cage of the rotor is given by:  $X_{\sigma gk} = X_{\sigma gs} + X_{\sigma gp} = 0.0000568 \Omega$  and its reduced value

to the stator in:  $X_{\sigma gk}' = X_{\sigma gk} k_{12} = 0.945 \Omega$ .

#### IV. STATIC PROPERTIES OF THE ELECTROMAGNETIC TORQUE OF THREE-PHASE ASYNCHRONOUS MOTOR WITH DOUBLE CAGE

Determining the static characteristic of the electromagnetic torque of induction motor with double cage is one of the most important goals in such research. Knowledge of this feature is particularly important because it provides a representation of the behavior of the engine in various operating modes. Determining the static torque of the engine with applied software package Maxwell 14 - 3D Design, is easy and fully automatic, and is performed before setting the parameter to determine the static torque in terms of before processing.

The calculations performed by the program in a way that uses the values of the magnetic vector potential in this specific case in the area covered in the air gap of the motor. Because complete electromagnetic analysis is made of seven different operating points, which in the previous chapters were given values of magnetic flux characteristic set points at different de-phased of the rotor relative to the stator, in this section will be determined the static characteristic of the electromagnetic torque for them seven set points, depending on the angular position of the rotor in the range of ( $0^\circ - 180^\circ$ ) electric. The actual determination of the torque which develops induction motor with double cage in Maxwell 14 - 3D Design, is based on the principle of virtual work, by applying the energy method, using already defined basic values of magnetic flux in the air gap of the motor.

Processing and obtaining the value of the electromagnetic torque is through the values of the magnetic vector potential in all the nodes of the three-dimensional domain quasi static model of asynchronous motor  $A_x, A_y, A_z$ , in particular given values of the currents of the load through the coils and a defined step of turning the rotor relative to the stator  $\Delta\theta$ . The accuracy of the determined value of the static electromagnetic point directly connected on one side to the size of the step that a change in the currents of the load on the engine and change the angle (position which covers the rotor relative to the stator), and on the other side with precision determination of the magnetic vector potential. The complete analysis is in direct relation to the accuracy of the mathematical model adaptation engine with real bike.

Data on the value of the electromagnetic torque are render the characteristics of the electromagnetic torque for various positions of the rotor under certain constant currents of load in stator coil and are presented in Figure 3. From the characteristic of Figure 3, under nominal current of the load on the induction motor with double cage (it's operating point marked with number 6) can be seen value of the electromagnetic torque  $M = 23.27 Nm$ , under angular de-phased of the rotor relative to the stator of  $15^\circ$  degrees mechanical or  $30^\circ$  electrical. Compared with the nominal value of the analyzed time asynchronous motor with double cage obtained by analytical calculations that is  $M = 23.67 Nm$ , it shows the great opportunity offered by

the method of finite element calculation of the electromechanical characteristics of electrical machines.

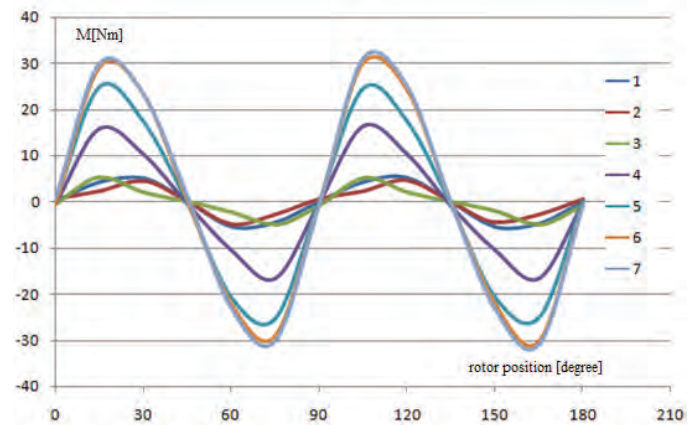


Fig.3. Characteristics of the electromagnetic torque

#### VI. CONCLUSION

All made calculations and simulation analyzes on the model of induction motor with double cage can be generalized and similarly be applied to any other electric machine. As a target in further research is required to develop the area transient analysis or consideration of dynamic characteristics and transient phenomena in induction motor with double cage with the method of finite elements in three dimensional domains. Generating of a mathematical model for dynamic analysis, getting the starting mechanical characteristics and optimizing the design of asynchronous motors.

#### REFERENCES

- [1] B.Arapinoski, M.Radevska, V. Ceselkoska and M.Cundev], "Modeling of Three Dimensional Magnetic Field in Three-Phase Induction Motor with Double Squirrel Cage " TEM Journal 2013.
- [2] Mirka Popnikolova Radevska, Blagoja Arapinoski, *Computation of solid salient poles synchronous motor electromagnetic characteristic*, 10<sup>th</sup> international conference of applied electromagnetic IIEC 2011, Nis, Serbia, 2011.
- [3] B. Arapinoski, M. Popnikolova Radevska, "Electromagnetic and thermal analysis of power distribution transformer with FEM" ICEST 2010, Ohrid, R.Macedonia 2010.
- [4] M. Popnikolova-Radevska, M. Cundev, L.Petkovska, "From Macroelements to Finite Elements Generation for Electrical Machines Field Analyses", ISEF International Symposium on Electromagnetic Fields in Electrical Engineering, Thessaloniki, Greece, 1995, p.p. 346-349.
- [5] B. Arapinoski, M. Popnikolova Radevska, D. Vidanovski "FEM Computation of ANORAD Synchronous Brushless linear motor" ELMA 2008, Sofia - Bulgaria.
- [6] M. Popnikolova Radevska: "Calculation of Electromechanical Characteristics on Overband Magnetic Separator with Finite Elements", ICEST 2006, p.p. 367-370, Sofia, Bulgaria 2006.
- [7] M. Cundev, L. Petkovska, M. Popnikolova-Radevska, "An Analyses of Electrical Machines Synchronous Type Based on 3D-FEM " ICEMA International Conference on Electrical. Machines and Applications, Harbin, China, September 1996, p.p. 29-32.

# Electromagnetic analysis on Three – Phase Induction Motor whit Cage with FEM 2D

Nikolce Talevski<sup>1</sup>, Mirka Popnikolova Radevska<sup>1</sup>, Blagoja Arapinoski<sup>1</sup>, Mile Spirovski<sup>1</sup>

**Abstract** – This paper deals with modelling of two dimensional magnetic field and both numerical computations and analysis of three phase induction motor with cage. The accuracy with which the electromagnetic quantities are computed in a great rate is dependent on the precision of calculations of electromagnetic field distribution in the machine. The 2D – Finite Element Method is very efficient for an accurate electromagnetic field solution and in this research is applied. This method is very efficient for an accurate electromagnetic field solution. The program package is used for performing automatic generation of finite elements mesh. After defining material construction, loading and excitation in both motor windings, the distribution of electromagnetic field in 2D motor domains is generated, the air gap flux and distributions of the flux density at the middle line of air gap is determinate.

**Keywords** – Three-Phase induction motor with squirrel cage, FEM 2D, electromagnetic analysis.

## I. INTRODUCTION

The three-phase induction machines are widely used in everyday life. Asynchronous motors stand out as characteristic in both constructive and in terms of technical performance. Such type of electric motor or asynchronous motor with cage rotor is selected as the subject of research from which one part presented in this paper. It will be made on electromagnetic analysis of specific asynchronous motor. This analysis will be made software, which calculations will be made based on well-known method of finite elements. As a results of the analysis will be presented: The distribution of the magnetic field in all regions of the engine, the magnetic induction in the air clearance category of the machine, and calculating the moment of the electromagnetic motor at nominal load. Software which has been applied to obtain the distribution of the electromagnetic field in 2D motor domain uses the famous and powerful finite element method (FEM). The three-phase induction motor with cage has the following rated data:  $P_n = 400kW$ ,  $U_n = 6000V$ ,  $f_n = 50Hz$ ,  $2p = 6$ ,  $\cos f = 0.86$ ,  $\eta = 0.84\%$ ,  $n_n = 987 \text{ min}^{-1}$  and delta winding connection.

From the main linkage flux results, the motor inductance and the electromagnetic torque are determined numerically, and their characteristics for various load and rotor angular positions are presented.

## II. GEOMETRY OF THE CONSIDERED PROBLEM

The subject of the research is three phase asynchronous motor with cage. As compared to the construction of the stator coil, the coil is identical to the standard three-phase asynchronous motors, in this case magnetic circuit of the stator is laminated and has 72 slots, they set a two-layered three-phase distributed coil with winding shortened step  $y = 11/12$ . The rotor has set a squirrel cage made of aluminum which has 90 channels. Complete geometry of three-phase induction motor with cage in 2D domain is presented on Figure1.

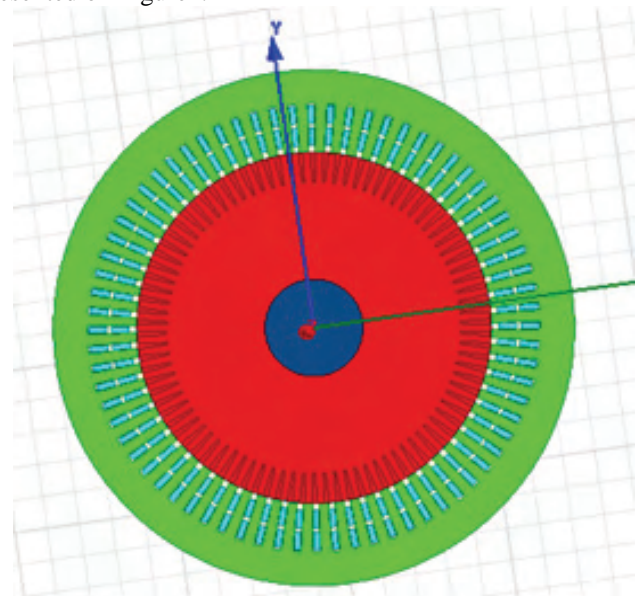


Fig.1. Geometry of model in 2D domain.

## III. MATHEMATICAL MODEL

The mathematical model is basically a program consisting of a system of Maxwell's equations well known in this analysis. Whit the settings of excitement of stator and rotor circuit process partially completes. It is necessary to make additional sections or places where they will present the socket burdens. By Figure 2 is represented excitement and terminals.

<sup>1</sup>Nikolce Talevski, Mirka Popnikolova Radevska, Blagoja Arapinoski, and Mile Spirovski are with the Faculty of Technical sciences at University of Bitola, Makedonska Falanga 33, Bitola 7000. Macedonia.



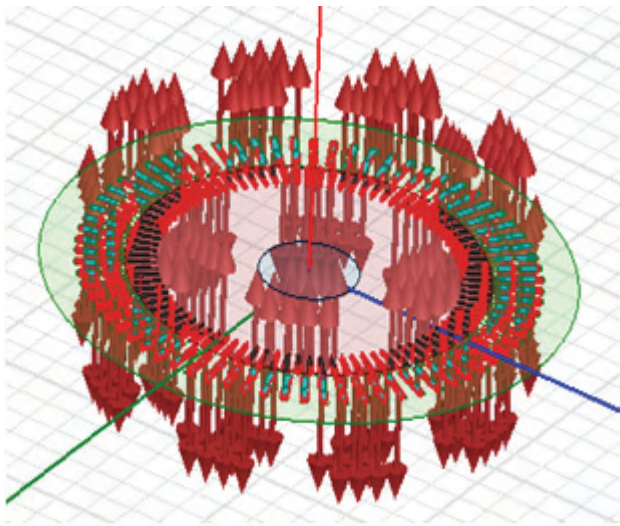


Fig.2. Excitements and terminals on asynchronous motor whit cage

The mathematical model provides an opportunity to solve equations for calculating the induction motor whit cage rotor. In such analysis is obtained from finite element network to test engine, and it can be seen that it is densest on air gap between stator and rotor. By increasing the density of the network of finite element increases the accuracy of calculation. Generated mash for this type of asynchronous motor whit cage rotor in 2D analysis down to the final number of 31538 items. This is represented in Figure 3.

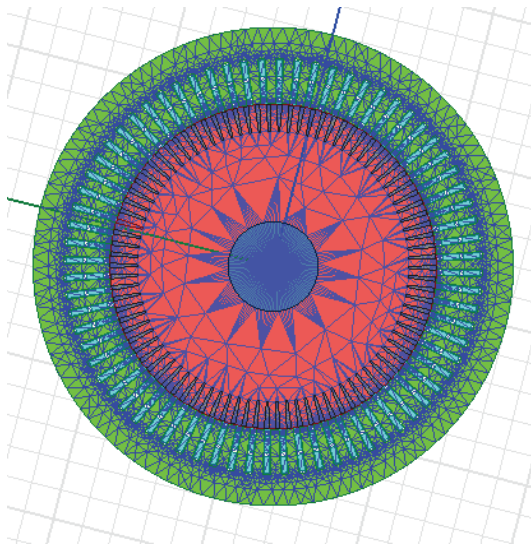


Fig.3. Finite element mash on asynchronous motor whit cage rotor complete geometry in 2D

#### IV. FINITE ELEMENT METHOD IN 2D FOR CALCULATION OF THE MAGNETIC FIELD

For performing the analysis a two dimensional numerical calculation of the magnetic vector potential and flux density in a two dimensional domain of the three-phase induction motor with cage is required. For that purpose the above mentioned computer program based on 2D Finite Element Method has

been used.. The numerical calculation is based on the Poissons' equation for magnetic field distribution in three dimensional domain:

$$\text{rot}(v(B) \cdot \text{rot}A) = J(x, y, z) \quad (1)$$

To realize a numerical solution of the equation (1) it is necessary to carry out a proper mathematical modeling of the machine. Then magnetic flux distribution can be plotted and this is presented on Fig.4 for excitation winding are energized with rated currents.

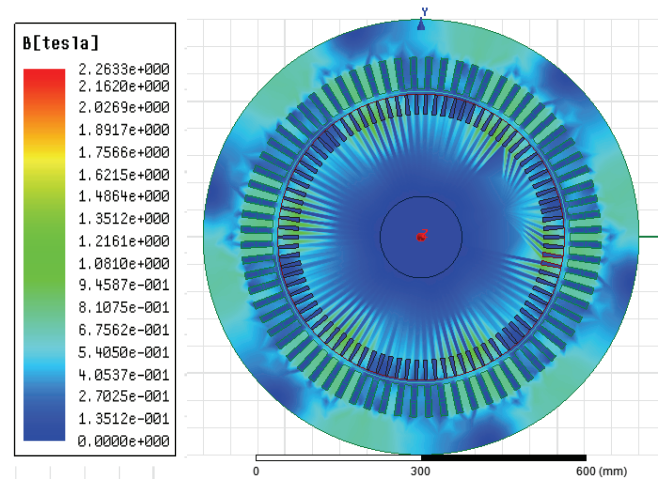


Fig.4. Magnetic field distribution in 2D motor domain

By using the procedure for numerical differentiation, the distribution of the magnetic flux density at the middle line of the air-gap in three-phase induction motor with cage rotor is determinate[1]. The characteristics of the magnetic flux density in dependence of the rotor position at different armature currents are presented in Figure 5

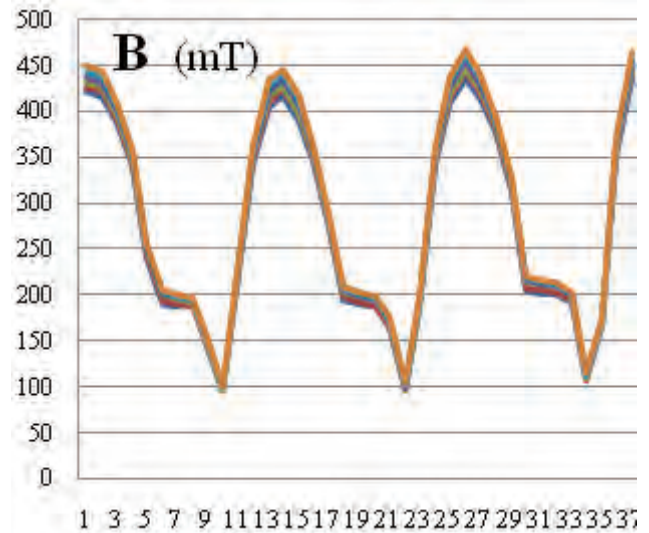


Fig.5. Characteristic of the magnetic flux density

Determination of the equivalent inductance between stator – rotor circuit can determine if there is value for the total energy for the magnetic field. The equivalent resistance is in direct relation to the magnetic flux, and determining the size necessary to know the load currents of the motor and the position of the rotor relative to the stator. Distribution of the



magnetic field through the three – phase asynchronous motor whit cage rotor is represented in Figure 5.

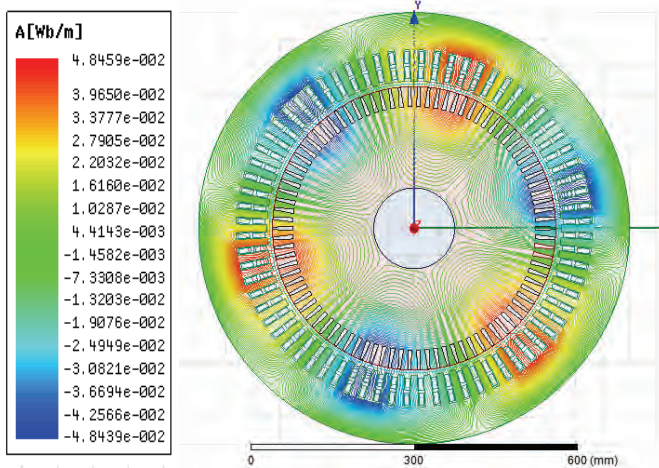


Fig.5. Magnetic field in asynchronous motor whit cage rotor

The characteristics of the magnetic field density in dependence of the rotor position at different armature currents are presented in Figure 6.

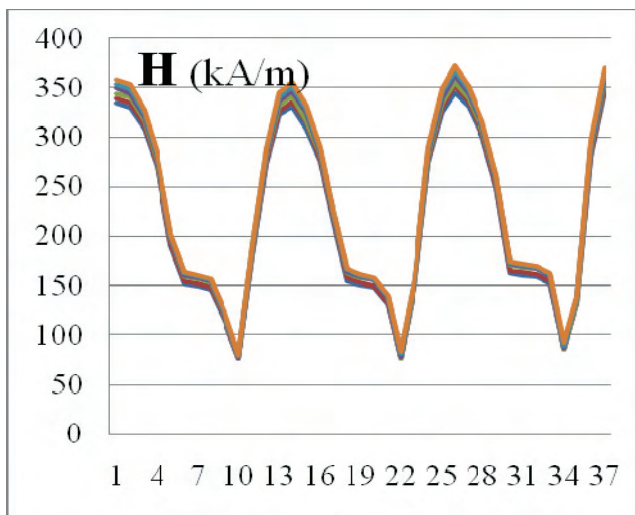


Fig.6. Magnetic field

Having the distribution of the magnetic vector potential in the whole investigate domain of the three phase asynchronous motor with cage rotor, the main flux in the air gap is determinate as well as leakage flux in the stator and rotor windings, and is going to be determined starting from the equation:

$$\Phi_{\delta} = \int_{\sum} rot A ds = \oint_c A dr = \int_{\sum} B ds = \iint_s (B \cdot n) ds \quad (2)$$

Then result is:

$$\Psi_{\delta} = w \iint_s (B \cdot n) ds = w \iint_s \left( \frac{\partial A_z}{\partial x} - \frac{\partial A_x}{\partial z} \right) dz dx \quad (3)$$

In the differentials replaced with differences and integrals with sums, for determining  $\Psi_{\delta}$ , the relation is as follows:

$$\Psi_{\delta} = w \cdot L \sum_{i=1}^{N_x} \Delta A_{zi} = w \cdot L (A_{zN_x} - A_{z1}) \quad (4)$$

The air-gap flux linkage  $\Psi_{\delta}$  for different constant rotor angular positions is presented on Figure 7.

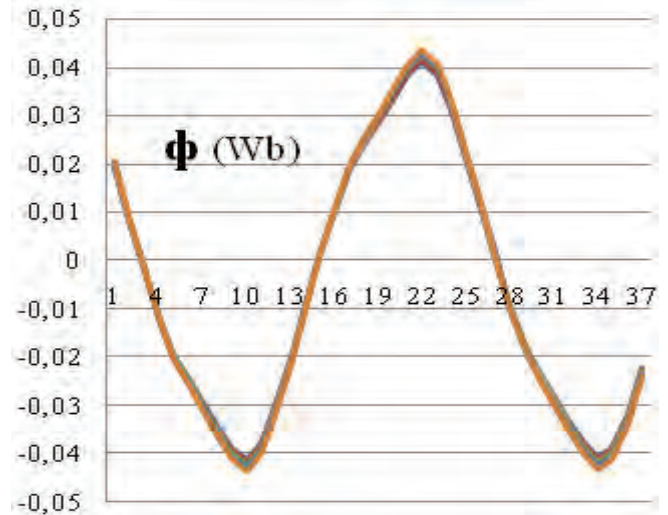


Fig.7. Air-gap flux linkage characteristics

The results of this crucial and addressing relevant matrices can afford to get the flux that occurs in the stator and rotor, for a period, and receive appropriate additions. Flux can afford to be distinguished as a flux which is distributed stator coil, stator part of the circuit, and flux which is distributed through rotor coil, rotor part of the circuit. Distribution of flux throughout stator and rotor circuit of asynchronous motor whit cage rotor is represented in Figure 8 and Figure 9.

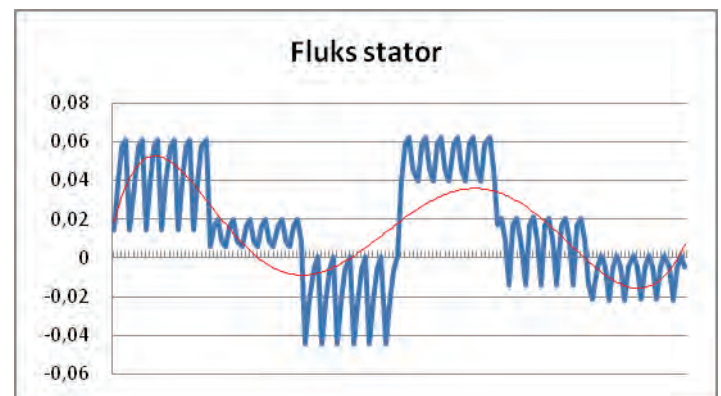


Fig.8. Distribution of flux – stator circuit

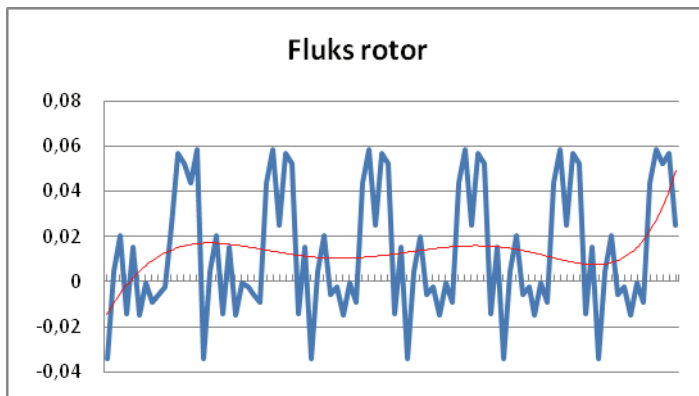


Fig.9. Distribution of flux – rotor circuit

## V. CALCULATION OF ELECTROMECHANICAL CHARACTERISTICS

The knowledge of electromagnetic torque characteristics is very important matter for analysis and performance of electrical motors. In this paper the energy concept for numerical calculation of electromagnetic torque is applied and for three phase asynchronous cage motor will be calculated by the change of the magnetic system co-energy at virtual angular displacement of rotor for different currents in the stator and rotor.

The torque characteristic of the three phase asynchronous motor with cage as a function of angular position of the rotor at rated load is given on the Figure 10.

The reliability of the calculated value of torque in this simulation analysis is confirmed, it is shown by the fact that the calculated value and the characteristic shape is identical to that obtained in experimental research done in real terms as an engine for the propulsion of a rubber band.

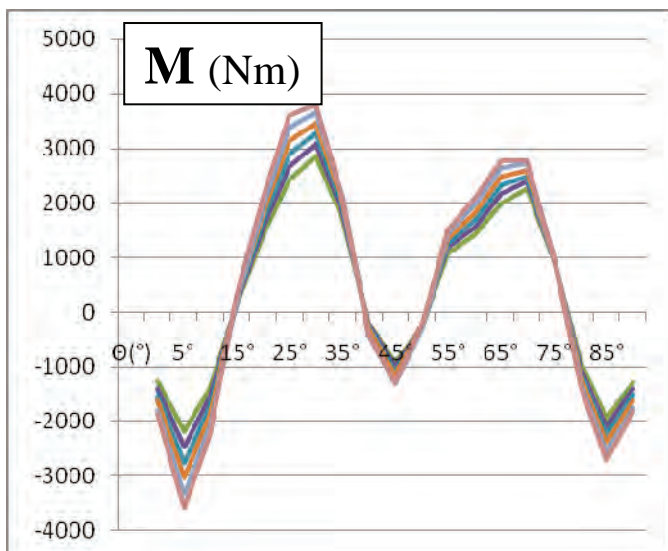


Fig.10. Torque characteristic of the three phase asynchronous motor with cage rotor

## VI. CONCLUSION

In this paper are presented some of the results obtained in an extensive research that aims to contribute to improving the performance of three phase asynchronous motor with cage rotor. Applied software which made simulation analysis is based on the finite element method implemented in two-dimensional domain. This contemporary method enables exact magnetic quantities such as air gap flux or flux density distribution to be evaluated in any part of the motor. The results of the field computations after they are used for calculations of electromechanical characteristics, as the static and dynamic torque.

## REFERENCES

- [1] B.Arapinoski, M.Radevska, V. Ceselkoska and M.Cundev, "Modeling of Three Dimensional Magnetic Field in Three-Phase Induction Motor with Double Squirrel Cage " TEM Journal 2013.
- [2] Mirka Popnikolova Radevska, Blagoja Arapinoski, *Computation of solid salient poles synchronous motor electromagnetic characteristic*, 10<sup>th</sup> international conference of applied electromagnetic IIEC 2011, Nis, Serbia, September, 2011.
- [3] B. Arapinoski, M. Popnikolova Radevska, "Electromagnetic and thermal analysis of power distribution transformer with FEM" ICEST 2010, Ohrid, R.Macedonia 2010.
- [4] M. Popnikolova-Radevska, M. Cundev, L.Petkovska, "From Macroelements to Finite Elements Generation for Electrical Machines Field Analyses", ISEF International Symposium on Electromagnetic Fields in Electrical Engineering, Thessaloniki, Greece, 1995, p.p. 346-349.
- [5] B. Arapinoski, M. Popnikolova Radevska, D. Vidanovski "FEM Computation of ANORAD Synchronous Brushless linear motor" ELMA 2008, Sofia – Bulgaria.
- [6] M. Popnikolova Radevska: "Calculation of Electromechanical Characteristics on Overband Magnetic Separator with Finite Elements", ICEST 2006, p.p. 367-370, Sofia, Bulgaria 2006.
- [7] M. Cundev, L. Petkovska, M. Popnikolova-Radevska, "An Analyses of Electrical Machines Synchronous Type Based on 3D-FEM " ICEMA International Conference on Electrical Machines and Applications, Harbin, China, September 1996, p.p. 29-32.

# Single-Line Ground Fault's Impedance Impact to Electrical Parameters in Ungrounded Electrical System

Milica Rašić<sup>1</sup>, Milan Radić<sup>2</sup> and Zoran Stajić<sup>3</sup>

**Abstract** – In this paper single-line ground fault in ungrounded electrical systems was considered. In accordance with one situation from practice that has been analyzed, appropriate model of ungrounded network and mathematical model that gives values of phase voltages and currents through single-lines to ground capacitances in the ungrounded system with single-line ground fault were given. In order to check all assumptions made for considered case from practice, an experiment in the laboratory environment was implemented and obtained results were presented.

**Keywords** – Ground fault, ungrounded power system, mathematical model, line and phase voltages, capacitive currents

## I. INTRODUCTION

As it is widespread known, the system grounding method does not have any impact to power system operation in normal conditions. However, in some non-regular regimes, such as ground faults, ground fault current's amplitude highly depends on the system grounding method. In addition, fault impedance value provides great impact to ground fault current. Generally, all electrical power systems can be classified in two categories: ungrounded and grounded systems. Furthermore, grounded systems can be classified in two groups: effectively (strongly) and none effectively (through reactance (Petersen coil) or through low-impedance) grounded power systems [1-3].

In Serbian electrical energy system all grounded methods are used according to different voltages level. Distribution systems at medium-voltage can be ungrounded, grounded through Petersen coil or low-impedance as opposed to distribution systems at high voltages ( $U_n \geq 110$  kV) that are usually effectively grounded.

In this paper, the emphasis was placed on ungrounded systems and impact that single-line ground fault provides to phase and line voltage's and current's values in those systems. Ungrounded systems are typical for distribution network 10 kV, as well as in industrial or mining electrical power systems (6 kV and 5 kV). For these systems it is specified that neutral points of all electric machines windings and elements in galvanic connection with distribution system are isolated from the ground. Single-lines to ground

capacitances represent the only connection between distribution system and ground in ungrounded power system [1]. Single-lines to ground capacitances are approximately equal each other, especially when conductors are transposed and they have same lengths. Therefore, the largest number of authors accepts the assumption that there is a capacitive symmetry in ungrounded systems and in accordance with that fact they make their mathematical models. This assumption was also adopted for research described in this paper.

## II. THEORETICAL BACKGROUND AND MATHEMATICAL MODEL

Considering the problem of single-line ground fault's appearance in electric power system at 10 kV (ungrounded system) in South Serbia, authors noticed some facts that looked atypical at the first sight. Actually, when single-line ground fault occurs in electric power system, electrical parameters, such as phase voltages, didn't have diagrams and values that are usually given in the literature for these situations. Namely, in literature is often said that in the case of single-phase ground fault, phase voltage in the "infected" phase drops to zero and in two other phases voltage increase to the value of line voltage. Line voltages in those cases keep same values as they had before the ground fault happened. However, after detailed literature review, authors determined the fact that described phase voltage's behaviour during single-line ground fault is typical only for non-load totally symmetrical power systems and for single-line ground faults with fault impedance  $Z_z = 0$  [1-4].

The problem that followed was how to prove the assumption that it was real single-line ground fault through fault impedance without appropriate mathematical model for this kind of power systems. In addition, situations that were assumed as single-line ground faults in this ungrounded power system were repeated by one usual pattern during the period of power system monitoring: the phase voltage  $U_C$  always dropped (but didn't reach 0), phase voltage  $U_B$  immediately increased (but didn't reach the line voltage  $U_{BC}$ ) and finally,  $U_A$  increased (but it had less increase than  $U_B$ ) or just kept the similar value as it had before single-line ground fault. Meanwhile, line voltages kept their values from the steady state. Therefore, according to the fact that there is a usual pattern by which are those situations characterized, it could be assumed that there are parts of transmission lines that had been overgrown by a different vegetation. This vegetation could cause situations assumed as single-line ground faults, during bad weather conditions, such as humidity or wind.

Another observed fact was that phase voltage's drop in one and asymmetrical increase in two other phases in the electric power system 10 kV did not have any impact to phase

<sup>1</sup>Milica Rašić is with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia  
E-mail: milicarasa@gmail.com

<sup>2</sup>Milan Radić is with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia  
E-mail: milan.radic@elfak.ni.ac.rs

<sup>3</sup>Zoran Stajić is with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia  
E-mail: zoran.stajic@elfak.ni.ac.rs



voltages at the secondary end of transformer 10/0.4 kV with  $D_{Y_n}$  connection what leads to suggestion that abnormal situations, as ground faults, in medium voltage networks are not transmitted to the low voltage side for this kind of transformer's connection.

In order to prove all mentioned assumptions, authors made electrical power system's model according to the real considered state that is shown in Fig. 1.

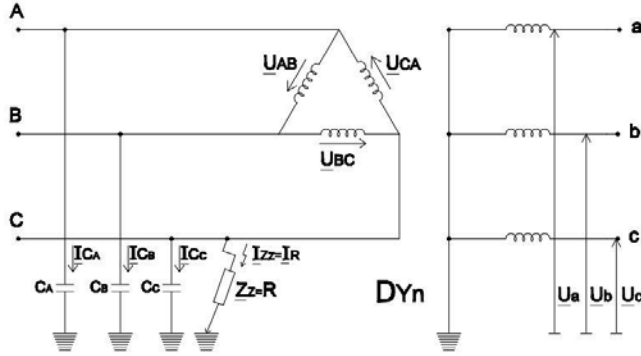


Fig. 1. Single-line ground fault in power distribution network

Primary side of 10/0.4 kV transformer is connected in D and its secondary side is connected in  $Y_n$ . Electric power system at 10 kV is ungrounded. Single-line ground fault occurred in phase C through fault impedance  $Z_z = R > 0 \Omega$ . This is main difference from models usually present in the literature, where  $Z_z = 0 \Omega$ .

From the model shown in Fig. 2 it can be set following system of equations:

$$I_{CA} + I_{CB} + I_{CC} + I_R = 0 \quad (1)$$

where  $I_{CA}$ ,  $I_{CB}$  and  $I_{CC}$  are capacitive currents that flow through single-lines to ground capacitances  $C_A$ ,  $C_B$  and  $C_C$ , respectively, and  $I_R$  is ground fault current;

$$U_A = Z_{CC} \cdot I_{CA} \quad (2)$$

$$U_B = Z_{CC} \cdot I_{CB} \quad (3)$$

where  $U_A$  and  $U_B$  are phase voltages and  $Z_{CC}$  is capacitive impedance that has the same value for all three phases, because it was assumed that  $C_A = C_B = C_C$ ;

$$U_A = U_{AC} + U_C \quad (4)$$

$$U_B = U_{BC} + U_C \quad (5)$$

where  $U_C$  is voltage of phase C, where single-line ground fault was happened,  $U_{AC}$  and  $U_{BC}$  are line voltages that have same values as before the ground fault;

$$U_C = Z_{CC} \cdot I_C \quad (6)$$

$$U_C = R \cdot I_R \quad (7)$$

Solving the system of equations Eqs. (1) – (7), formulas for phase voltages and currents through single-lines to ground capacitances calculation at the point of single-line ground fault for used power system model, were obtained. Those formulas are given in Eqs. (8) – (14).

$$U_C = \frac{-R \cdot [U_{AC} + U_{BC}]}{Z_{CC} + 3 \cdot R} \quad (8)$$

$$I_{CC} = \frac{-R \cdot [U_{AC} + U_{BC}]}{Z_{CC} \cdot (Z_{CC} + 3 \cdot R)} \quad (9)$$

$$U_A = U_{AC} + \frac{-R \cdot [U_{AC} + U_{BC}]}{Z_{CC} + 3 \cdot R} \quad (10)$$

$$I_{CA} = \frac{U_{AC}}{Z_{CC}} + \frac{-R \cdot [U_{AC} + U_{BC}]}{Z_{CC} \cdot (Z_{CC} + 3 \cdot R)} \quad (11)$$

$$U_B = U_{BC} + \frac{-R \cdot [U_{AC} + U_{BC}]}{Z_{CC} + 3 \cdot R} \quad (12)$$

$$I_{CB} = \frac{U_{BC}}{Z_{CC}} + \frac{-R \cdot [U_{AC} + U_{BC}]}{Z_{CC} \cdot (Z_{CC} + 3 \cdot R)} \quad (13)$$

$$I_R = \frac{-[U_{AC} + U_{BC}]}{Z_{CC} + 3 \cdot R} \quad (14)$$

Certainly, it should be mentioned that this model can also be applied for other different voltage levels with the condition that the power system is ungrounded and transformer is in  $D_{Y_n}$  connection. However, in this model line impedances and line-to-line capacitances were not taken into the account, because line impedances are much smaller than the fault impedance and analogously, line-to-line capacitances are much smaller than single-lines to ground capacitances. For this kind of analyses this model satisfies the required precision, but for highest precision calculations, mathematical model that takes them into account can also be made [5].

### III. EXPERIMENTAL INVESTIGATION

In order to check all assumptions mentioned in section 2, after mathematical model was created, authors decided to make laboratory experiment. Actually, electric power system model presented in Fig. 1 was built by laboratory equipment which connection diagram is shown in Fig. 2.

Two laboratory transformers 400/(100 ·  $\sqrt{3}$ ) V,  $D_{Y_n}$  connection were used. Transformer T1 represents transformer 10/0.4 kV in ungrounded system at which primary side single-line ground faults occurs, and transformer T2 represents utility source (it is connected to power system through three-phase regulation autotransformer (3fRAT)).

Capacitors  $C_A$ ,  $C_B$  and  $C_C$  represent single-lines to ground capacitance in ungrounded electric power system. Those capacitors are  $C_A = C_B = C_C = 1.5 \mu F$ , and they were chosen according to the allowed value of metal ground fault current in ungrounded systems. Actually, according to [6], ground fault current for electric power system 10 kV is about



0.03 A/km which is for ungrounded system of total feeder lines length 50 km (system that authors considered) 1.5 A. For transformer 10/0.4 kV, rated current 46 A (also parameters from considered system) total fault current of 1.5 A represents 3.26 %. Laboratory transformer has rated current 10 A, and, analogously, ground fault current of 3.26 % is about 0.326 A. In this case study, with capacitors of available manufacturer capacitances values, the closest metal ground fault current that was able to achieve was 0.342 A with capacitors of 1.5  $\mu$ F.

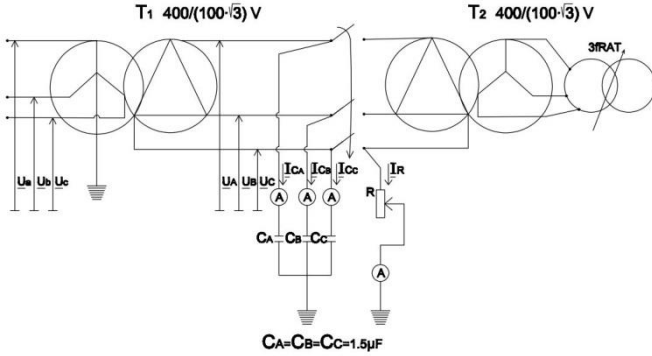


Fig. 2. Connection diagram of electric equipment used for laboratory experiment

From the electric power system using 3fRAT primary ends of transformer T2 were supplied by the line voltage of 400 V. Single-line ground faults in phase C through a range of fault impedance  $R$  values were tested. Electrical parameters as currents  $I_{CA}$ ,  $I_{CB}$ ,  $I_{CC}$ ,  $I_R$  and phase and line voltages  $U_A$ ,  $U_B$ ,  $U_C$ ,  $U_{AB}$ ,  $U_{BC}$  and  $U_{AC}$  were measured. It should also be mentioned that phase voltages at the secondary end of transformer T1 ( $U_a$ ,  $U_b$  and  $U_c$ ) were also monitored in order to verify if the whole state is analogous to the state from 10 kV ungrounded power system that authors analyzed.

The first measuring point was for metal ground fault where fault impedance was  $R = 0 \Omega$ . Thereafter,  $R$  was increased, in the beginning with small increments, and after a few measurements the  $R$  increment was set to the value of 100  $\Omega$  in order to analyze as large as possible range of fault impedance values. Testing was finished for the fault impedance value of  $R = 1500 \Omega$ . Results of this experiment presented as measured electrical parameters diagrams as a function of  $R$  are given in section 4.

#### IV. RESULTS AND DISCUSSION

First of all it should be said that during the whole experiment phase voltages at the secondary end of transformer T1 ( $U_a$ ,  $U_b$  and  $U_c$ ) weren't changing their values what proves the assumption that single-line ground fault at the primary side of transformer connected in  $D_{Yn}$  does not have any influence to electrical parameters at its secondary side.

The following situation that was analyzed was about line voltages at the primary end of T1. During the experiment they

kept their values of  $U_{AB} = 402$  V,  $U_{BC} = 403$  V and  $U_{AC} = 406$  V, that is according with the considered situation from the practice.

Measured currents  $I_{CA}$ ,  $I_{CB}$ ,  $I_{CC}$ ,  $I_R$  and phase voltages  $U_A$ ,  $U_B$  and  $U_C$  change their values with the change of fault impedance what is presented in Figs. 3 and 4.

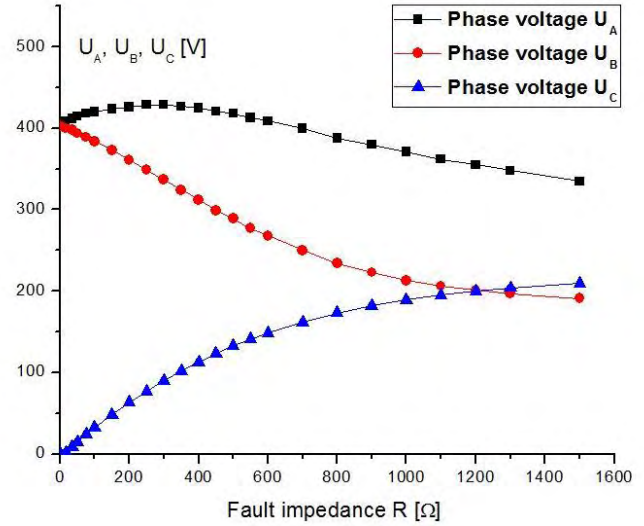


Fig. 3. Diagrams of phase voltages  $U_A$ ,  $U_B$  and  $U_C$  as functions of fault impedance  $R$  obtained by an experiment

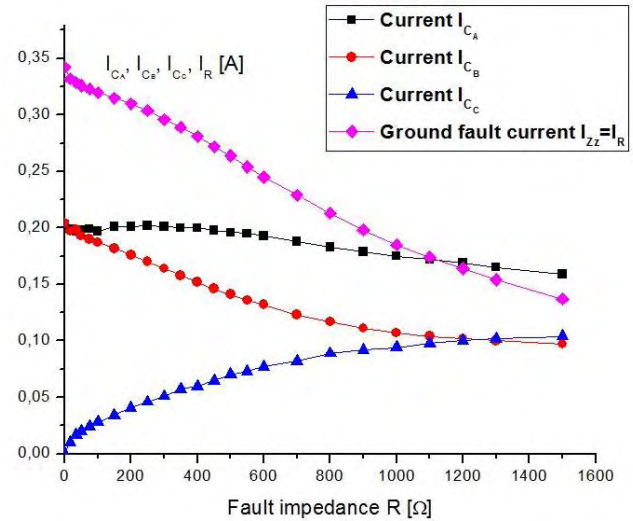


Fig. 4. Diagrams of currents through single-lines to ground capacitances  $I_{CA}$ ,  $I_{CB}$ ,  $I_{CC}$  and ground fault current  $I_R$  as functions of fault impedance  $R$  obtained by an experiment

After the laboratory experiment results were obtained, mathematical model from section 2 was used for their checking. Actually, in Eqs. (8) – (14) measured electrical parameters are calculated in complex domain with their arguments, but if it is known that  $\underline{U}_{AC}$  and  $\underline{U}_{BC}$  can be defined by Eqs. (15) and (16) it is easy to calculate effective values of all electrical parameters obtained by mathematical

model and compare them with values gained from the experiment.

$$\underline{U}_{AC} = U_{AC} \cdot e^{-j \cdot \frac{2\pi}{3}} \quad (15)$$

$$\underline{U}_{BC} = U_{BC} \cdot e^{j \cdot \pi} \quad (16)$$

Diagrams of electrical parameters changes gained by mathematical model for the same range of  $R$  values, as it was in the experimental investigation, are presented in Figs. 5 and 6. It should be noted that in mathematical model  $R$  was used without increments in order to obtain continuous diagrams.

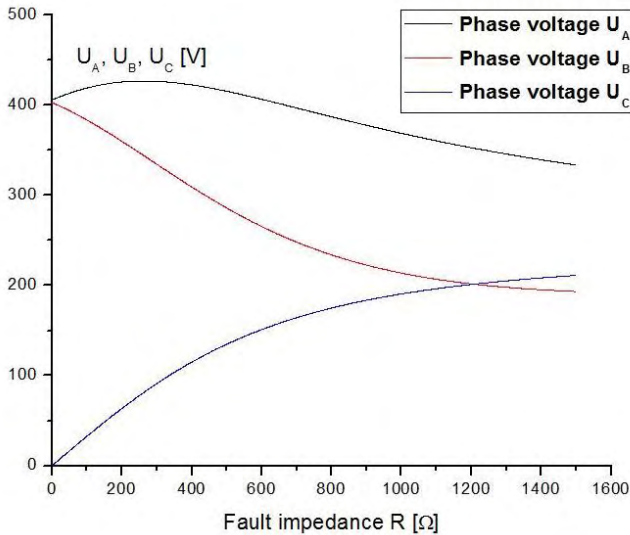


Fig. 5. Diagrams of phase voltages  $U_A$ ,  $U_B$  and  $U_C$  as functions of fault impedance  $R$  obtained by mathematical model

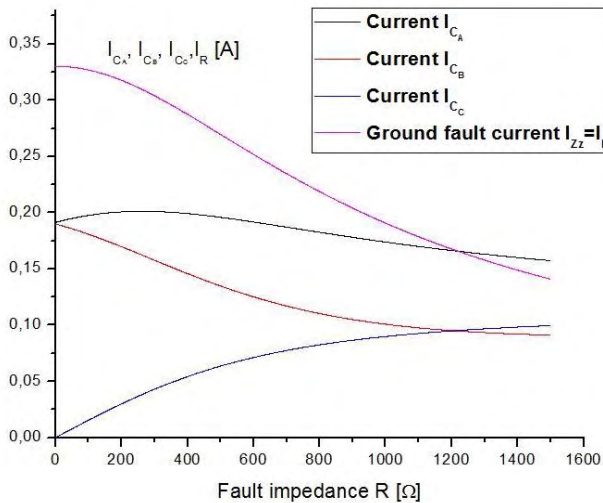


Fig. 6. Diagrams of currents through single-lines to ground capacitances  $I_{CA}$ ,  $I_{CB}$ ,  $I_{CC}$  and ground fault current  $I_R$  as functions of fault impedance  $R$  obtained by mathematical model

If Figs. 3 and 4 are compared with Figs. 5 and 6, it is obvious that phase voltages  $U_A$ ,  $U_B$  and  $U_C$ , currents through capacitance  $C_A$ ,  $C_B$  and  $C_C$ ,  $I_{CA}$ ,  $I_{CB}$  and  $I_{CC}$ , respectively, and ground fault current  $I_R$  have same diagrams obtained by experiment as diagrams obtained by mathematical model. In this way authors prove their assumptions made about problem with single-line ground fault through impedance in 10 kV ungrounded electric power system that they have been analyzing.

## V. CONCLUSION

Single-line ground fault through different values of impedance in ungrounded electric power system was analyzed in this paper. It was shown that phase voltages and ground fault current in distribution systems highly depends on fault impedance value and that its influence must be taken into the account in ground fault analyses.

In the experiment described in this paper active resistance as fault impedance was used. Future research in this area can cover different impedance characters, such as reactive or combination of active and reactive characters. In future investigation impedance of transmission lines and line-to-line capacitances can be taken into account and mathematical model for highest-precision analyses can be created. Finally, this type of analyses done for ungrounded electric power system can be implemented to grounded systems, including individually all grounding methods.

## ACKNOWLEDGEMENT

The paper is a part of the research done within the project III 44006, supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia within the framework of technological development.

## REFERENCES

- [1] D. P. Stojanović, *Prenos električne energije*, Elektronski fakultet, Niš, 2008. (in Serbian)
- [2] N. Rajaković, D. S. Tasić, *Distributivne i industrijske mreže*, Akademska misao, Beograd, 2008. (in Serbian)
- [3] J. Nahman, *Uzemljenje neutralne tačke distributivnih mreža*, Naučna knjiga, Beograd, 1980. (in Serbian)
- [4] N. Fischer, D. Hou, "Methods for detecting ground faults in medium-voltage distribution power systems", *White paper*, Schweitzer Engineering Laboratories, Inc., Pullman WA USA, 2006.
- [5] B. Brown, "Section 6: System grounding", Square D Engineering Services.
- [6] Tehnička preporuka Br .6, "Uzemljenje neutralnih tačaka u elektrodistributivnim mrežama 110 kV, 35 kV, 20 kV, 10 kV i 0.4 kV", V izdanje, JP EPS – Direkcija za distribuciju električne energije, Mart 2004. (in Serbian).

# Influence of Induction Generator's Windings Resistance Variation on Optimal CVT Regulation Curves

Milan Radić<sup>1</sup>, Milica Rasić<sup>2</sup> and Zoran Stajić<sup>3</sup>

**Abstract** – In this paper a simple, parameter-based analytical approach for calculation of continuously variable transmission (CVT) optimal regulation curves in a small wind energy conversion (WEC) system is presented. A CVT is placed between turbine's rotor and an induction generator directly connected to the power system, enabling better utilization of available wind power. Methodology for calculation of optimal CVT regulation curve has been explained in details and has been applied to the small WEC with rated power of 55 kW. Effects of generator's winding resistance variation, caused by temperature change, have been studied further. Obtained results have shown that normally expected changes of winding's resistances have no significant impact on calculated curves.

**Keywords** – Wind turbine, Induction generator, Winding temperature, CVT, Optimal regulation.

## I. INTRODUCTION

The crucial part of any wind energy conversion system (WEC) is a wind turbine itself, being an element whose role is to accept kinetic energy of the airflow and to convert it to the driving mechanical torque on the generator's shaft. Accepted aerodynamic theory defines that there is the absolute maximum of a wind turbine's efficiency that can not be exceeded under any circumstances (0.596), which is also known as the Betz limit ([1]). From the historical point of view, efficiency of energy conversion in older designs of wind turbines (such as Savonius turbine or American farm wind turbine), could not reach neither the half of the mentioned theoretical limit. Even in contemporary engineering solutions, obtained by intensive usage of computer-aided design methods, there are no significant breakthroughs that could be regarded as an efficiency „very close to the Betz limit". For example, modern designs of three-blade horizontal axis wind turbines (HAWT) allow the wind energy conversion with the maximum efficiency of about 0.5.

Described problem is further complicated by the fact that the efficiency of the wind energy conversion is, in general, a highly-nonlinear function, depending on both actual angular speed of the rotor and actual velocity of the wind. In order to

reach the optimal operating point, characterized by the maximum efficiency, it is necessary to achieve exact, optimal ratio between rotor's angular speed and actual wind velocity. The value of this ratio differs from one type of wind turbine to another, but it is obvious that, when wind velocity varies, angular speed of the rotor also has to be varied. Otherwise, the wind turbine will not extract the maximum of available kinetic energy from the wind.

Since the rotor of the driven generator has to rotate at constant, or almost constant speed (depending on if a synchronous or an induction generator is applied), in order to deliver power to the electric power system (EPS) with constant frequency, it is necessary to exploit some additional technical devices for solving of the mentioned problem. Contemporary solutions are mostly based on intensive usage of power electronic devices that enable some type of indirect coupling between the EPS and the generator. This approach enables that, despite of constant frequency in the EPS, angular speed of the generator's rotor can be varied in considerably wide range. However, these solutions also have significant drawbacks, at the first line regarding very complex control of power electronic devices and harmonic distortion of generated power.

Although the commercial use of continuously variable transmission (CVT) was introduced by automotive industry more than a half century ago, in recent decades this concept has also occupied the interest of researchers and producers in the area of WEC systems ([2],[3]). There are several studies indicating that an CVT incorporated with a wind turbine can completely substitute power electronics in performing the task of optimal operating point tracking, while the generator is directly connected to the EPS. Main concerns related to wider usage of CVT in small WEC systems are questions about reliability, durability, and also limited capacity for transfer of mechanical energy. However, in [4] was shown that, with innovative approach and careful design of CVT, these problems could be mitigated. In [5], utilization of cage rotor induction generators in CVT based WEC systems has been suggested and some issues related to synchronization with EPS have been discussed. In [6] method for calculation of optimal CVT regulation curves for a WEC system with an induction generator directly connected to the EPS has been proposed, and also effects of the EPS voltage variations on calculated curves have been studied.

The goal of this paper is to study effects that variations of generator's windings resistance, caused by changes of their temperature, could have on optimal CVT regulation curve. The method used for calculation of optimal CVT regulation curves is similar to the one described in [6], however, analysis is directed to different aspects of the problem.

<sup>1</sup>Milan Radić is with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia  
E-mail: milan.radic@elfak.ni.ac.rs

<sup>2</sup>Milica Rašić is with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia  
E-mail: milicarasa@gmail.com

<sup>3</sup>Zoran Stajić is with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia  
E-mail: zoran.stajic@elfak.ni.ac.rs

## II. IDENTIFICATION OF THE PROBLEM

Layout of the WEC system studied in this paper is presented in Fig. 1.

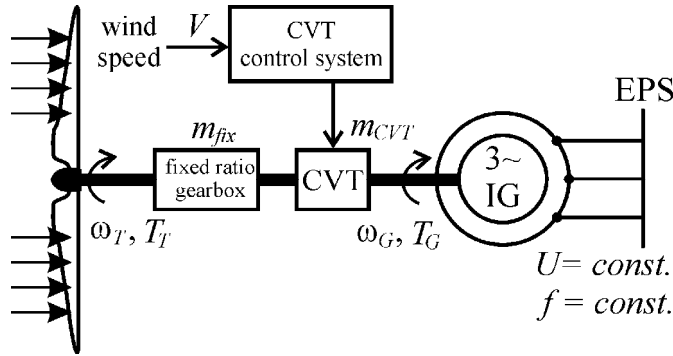


Fig. 1. Layout of the analyzed WEC system

It consists of a HAWT, a fixed-ratio drive train, a CVT and a three-phase cage-rotor induction generator (IG), connected directly to the electric power system.

Induction generator can be represented by its equivalent circuit, shown in Fig. 2. The main difference that can be noticed, compared to the standard equivalent circuit is that stator and rotor resistances,  $R_s$  and  $R_r$  are considered as variable elements, whose values depend on the actual temperature of the winding,  $\theta$ .

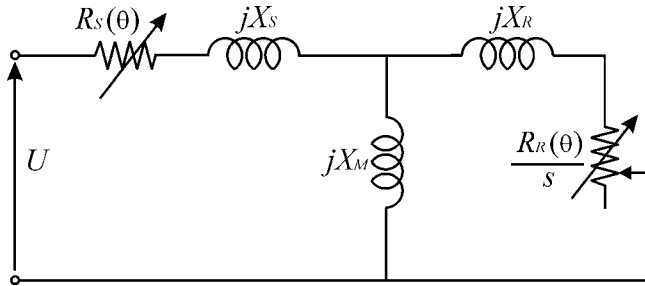


Fig. 2. Equivalent circuit with variable stator and rotor resistances

This approach is justified by the fact that wind is very unpredictable and uncontrollable source of mechanical power, and as a consequence, operation of IG can not be considered as stationary operation with constant load. Instead, it is likely that IG would operate in unpredictable cycles, while winding temperature could vary from ambiental temperature  $\theta_a$ , to the maximum allowed value  $\theta_{hMAX}$  defined by the temperature tolerance class of insulation material.

Such variations in temperature will result in variations of actual winding resistances, according to the equation

$$R_h = R_a \frac{A + \theta_h}{A + \theta_a} \quad (1)$$

where  $R_h [\Omega]$  is resistance of the hot winding, related to the increased temperature  $\theta_h [^\circ C]$ ,  $R_a [\Omega]$  is resistance of the

winding at the ambiental temperature  $\theta_a [^\circ C]$ , and  $A = 235$  (if windings are made of copper), or  $A = 255$  (if windings are made of aluminum).

Mentioned changes of winding resistances will further cause slight degeneration of generator's torque-speed curve, and will affect the optimal CVT regulation curve in some extent. In the next section of the paper, basic mathematical model of the analyzed WEC system is presented.

### A. Mathematical model

Mechanical power extracted from the wind by HAWT is given as:

$$P_T = 0.5 \rho \pi R^2 V^3 C_p(\lambda, \beta) \quad (2)$$

where  $\rho$  is the density of the air [ $kg/m^3$ ],  $R$  is the radius of the turbine's rotor [ $m$ ] and  $V$  is the wind speed in [ $m/s$ ]. In Eq. (2),  $C_p(\lambda, \beta)$  is non-dimensional power coefficient, depending on the actual tip speed ratio:

$$\lambda = \frac{\omega_T R}{V} \quad (3)$$

and on actual rotor's blades pitch angle  $\beta$ . In the normal operating region, where optimal regulation is exploited, angle  $\beta = 0^\circ$  and the power coefficient is a nonlinear function of tip-speed ratio only. Producers of wind turbines usually give function  $C_p(\lambda)$  in the form of look-up table, but there are also different analytical expressions that are exploited for WEC system studies. The curve used in this paper is shown in Fig. 3.

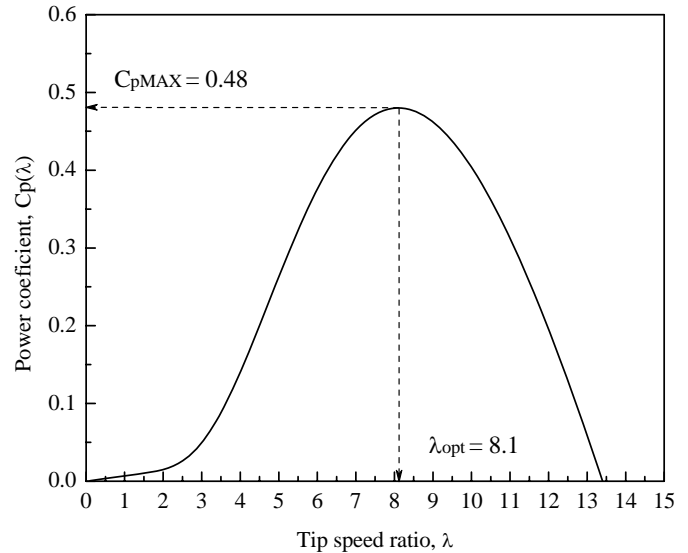


Fig. 3. Power coefficient curve of the analyzed wind turbine

It is very important to understand that the complete shape of the power coefficient curve and its accurate mathematical formulation are not necessary for the analysis presented in this



paper. It is enough to know accurate coordinates of the optimal operating point  $(\lambda_{opt}, C_{pMAX})$ .

Using the general definition of a mechanical torque, and also Eqs. (2) and (3), the actual torque developed by HAWT can be expressed as:

$$T_T = \frac{\rho \pi R^3 V^2}{2\lambda} C_p(\lambda) \quad (4)$$

The electromagnetic torque of the induction generator (which acts as braking torque) is defined by the actual value of voltage  $U$  in the EPS, but also depends on the actual angular speed of the generator's shaft  $\omega_G$  which is not constant, and also on parameters of the machine's equivalent circuit. It is defined as:

$$T_{eG} = \frac{3}{\omega_S} \cdot \frac{R_R}{s} \cdot \frac{U^2}{\left(R_S + \nu \frac{R_R}{s}\right)^2 + (X_S + \nu X_R)^2} \quad (5)$$

where  $R_S$  is stator resistance,  $R_R$  is rotor resistance,  $X_S$  is stator leakage reactance and  $X_R$  is rotor leakage reactance.

Synchronous angular speed of the generator  $\omega_S$  depends on number of the machine's pole pairs  $p$  and, for the constant frequency  $f$  in the EPS, it is defined as  $\omega_S = 60f/p$ .

The value of generator's relative slip  $s$  can be calculated as

$$s = (\omega_S - \omega_G) / \omega_S \quad (6)$$

and finally, stator's leakage coefficient  $\nu$  figuring in Eq. (5) can be determined as  $\nu = 1 + X_S / X_M$ , where  $X_M$  is the magnetizing reactance of the generator.

The wind turbine and the induction generator are mechanically coupled via transmission whose total transfer ratio is expressed as:

$$m_{tot} = m_{fix} \cdot m_{CVT} \quad (7)$$

where  $m_{fix}$  is the transfer ratio of the fixed drive train, and  $m_{CVT}$  is the actual transfer ratio of the applied CVT, which can take arbitrary value between the limits defined by the construction. In order to simplify analysis at this stage, it is considered that transmission is ideal and that all friction losses can be neglected. With such assumptions, static equilibrium of the analyzed WEC system is defined by

$$T_T + m_{tot} T_{eG} = 0 \quad (8)$$

#### B. Method for calculation

Using Eqs. (3) (4), (5), (6), and also the Eq. (9) connecting angular speeds  $\omega_G$  and  $\omega_T$ ,

$$\omega_G = m_{tot} \omega_T \quad (9)$$

Eq. (8) can be rewritten as:

$$0 = \frac{\rho \pi R^3 V^2 C_p(\lambda)}{2\lambda} + \frac{3m_{tot} R_R}{R\omega_S - m_{tot} V \lambda} \cdot \frac{U^2}{\left(R_S + \frac{\nu R_R R \omega_S}{R\omega_S - m_{tot} V \lambda}\right)^2 + (X_S + \nu X_R)^2} \quad (10)$$

Substituting values for  $\lambda = \lambda_{opt}$  and  $C_p(\lambda) = C_{pMAX}$  into the Eq. (10), and considering  $U$ ,  $f$  and all parameters of the equivalent circuit except  $R_S(\theta)$  and  $R_R(\theta)$  are constant, we obtain function

$$g(V, m_{tot}, R_S, R_R) = 0 \quad (11)$$

describing the optimal operation of the analyzed WEC.

Previous relation allows possibility that stator and rotor winding resistances are not constant. If Eq. (11) is solved for a chosen, arbitrary combination of  $R_S$  and  $R_R$  as parameters, and for an arbitrary value of wind speed  $V$ , obtained solution for  $m_{tot} = m_{opt}$  represents the optimal total transfer ratio, desired for reaching the optimal operating point. If chosen combination of parameters  $R_S$  and  $R_R$  is kept, further solving of Eq. (11) for different discrete values of  $V$  results in identification of points that describe functional dependency  $m_{opt} = h(V)$ , valid for the chosen combination of  $R_S$  and  $R_R$ . Finally, it is now possible to calculate another set of optimal CVT transfer ratios  $m_{CVTopt} = f(V)$ , using Eq. (7) and known fixed gear transfer ratio  $m_{fix}$ .

It is not necessary to derive the exact solution of Eq. (11) in a closed form. Instead, it is possible to exploit *fzero* function from the *Matlab Optimization Toolbox*, whose application for chosen  $R_S, R_R$  and  $V$  successfully leads to the identification of the optimal total transfer ratio  $m_{opt}$ .

### III. CALCULATED RESULTS

The explained analytical approach was used to calculate optimal CVT transfer ratios for the small WEC system consisted of 3-blade HAWT whose parameters are  $P_{Tn} = 55 \text{ kW}$ ,  $R = 5.9 \text{ m}$ ,  $V_n = 12 \text{ m/s}$ ,  $C_{pMAX} = 0.48$ ,  $\lambda_{opt} = 8.1$ ,  $m_{fix} = 14$ , and of 4-poles, three phase cage rotor induction machine with nameplate data (given for motoring mode of operation)  $P_n = 55 \text{ kW}$ ,  $U_n = 400 \text{ V}$ ,  $f_n = 50 \text{ Hz}$ ,  $I_n = 101.5 \text{ A}$ ,  $\Delta$  connection, thermal tolerance class F. All reactances from the equivalent circuit shown in Fig. 2 are considered as constant and their values are  $X_S = X_R = 0.754 \Omega$  and  $X_M = 20.83 \Omega$ . Stator and rotor resistances are treated as variable parameters, depending on the actual temperature rise in windings. The first pair of resistances  $(R_{Sa}, R_{Ra})$  that has been used for calculation were resistances valid for the supposed ambiental temperature of

$\theta_a = 20^\circ C$ , and their values have been taken as  $R_{Sa} = 0.14 \Omega$  and  $R_{Ra} = 0.19 \Omega$ . The second pair of resistances used for calculation ( $R_{Sh}, R_{Rh}$ ) has been related to the temperature of hot windings  $\theta_{hMAX} = 155^\circ C$ , which is the maximum allowable operation temperature for the temperature tolerance class F. Using Eq. 1, and knowing that stator winding is made of copper, while rotor cage is made of aluminum, resistances of hot windings have been determined as  $R_{Sh} = 0.214 \Omega$  and  $R_{Rh} = 0.283 \Omega$ . Values of wind speed have been varied in the range between the cut-in speed  $V_{ci} = 5 m/s$  and the nominal wind speed  $V_n = 12 m/s$ , using incremental steps of  $\Delta V = 0.1 m/s$ . Calculated results for optimal regulation of CVT transfer ratio  $m_{CVTopt} = f(V)$  are shown in Fig. 4.

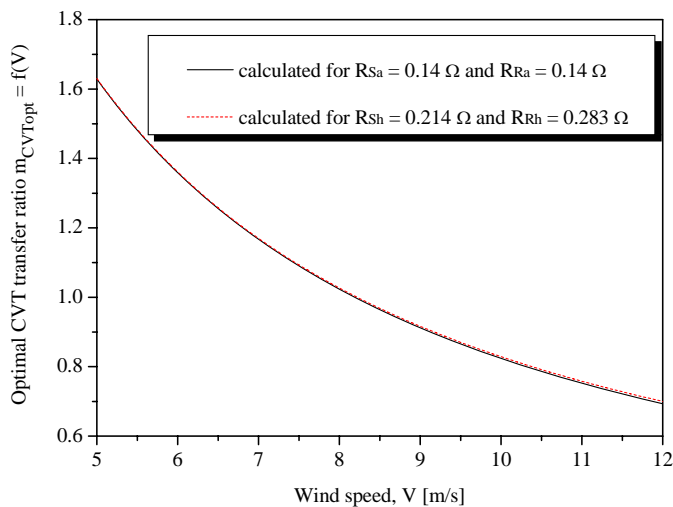


Fig. 4. Calculated optimal CVT regulation curves

At low wind speeds, total transfer ratio  $m_{tot}$  has to be high in order to enable optimal operation of wind turbine. In such regimes, turbine should rotate slowly, extracting maximum of available mechanical power from the wind. As the wind speed increases, turbine's angular speed has also to be increased; otherwise, it will not operate with power coefficient  $C_p = C_{pMAX}$  anymore. Since generator's angular speed can not be increased in a desired amount, due its direct connection to the power system, the total transfer ratio has to take lower values. This action can be performed using a control system for achieving the optimal transfer ratio of the exploited CVT, according to the calculated optimal regulation curve,  $m_{CVTopt} = f(V)$ . In the region of low wind speeds, CVT has to be controlled in the way that its transfer ratio is greater than 1. As the wind speed is increased, transfer ratio of the CVT has to take lower values, keeping the wind turbine in the optimal operating point. Information shown in Fig. 4 can be of great interest during synthesis of any CVT control system.

In the region characterized by low wind speeds (i.e. in the region where  $m_{CVTopt} > 1$ ), it is almost not possible to notice any difference between calculated curves. Small drift occurs

only in the region of wind speeds close to the  $V_n$ , but it is still less than 1%, although the rise in resistances of about 50% has been taken into consideration. It is obvious that if temperature of windings is increased to the highest value acceptable for normal operation, optimal CVT transfer ratio should be just slightly increased. This can be explained by the nature of the induction generator's torque-speed curve, because with higher resistance of windings, machine will develop optimal electromagnetic braking torque at somewhat higher rotational speed, compared to the situation when windings are cold.

#### IV. CONCLUSION

Applying the methodology presented in the paper, for any WEC system with a cage rotor induction generator directly connected to the electric grid, the optimal CVT transfer ratio can be calculated as the function of actual wind speed. As the presented approach does not require complete shape of the turbine's power coefficient curve, it can be very useful tool for synthesis of a control system. Although temperature changes can cause significant variation of resistances in stator and rotor windings, presented results clearly show that such variation has no significant impact on calculated optimal CVT transfer ratio regulation curves. Obtained information is very important, since it enables more comfort during the design process of an appropriate CVT regulator.

#### ACKNOWLEDGEMENT

This paper was realized as a part of the projects TR35005 and III44006, supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia for the period 2011-2016.

#### REFERENCES

- [1] J. Manwell, J. McGowan and A. Rogers, *Wind Energy Explained: Theory, Design and Application*, John Wiley & Sons, New York, 2002, pp. 83-139.
- [2] A. Martens and P. Albers, *Wind Turbine Study: Investigation into CVT application in wind turbines*, Technische Universiteit Eindhoven, Eindhoven, October 2003, pp. 3-27.
- [3] J. Cotrell, "Assessing the potential of a mechanical continuously variable transmission for wind turbines", Conference paper NREL/CP-500-38212, National Renewable Energy Laboratory, August 2005.
- [4] V. Miltenović, M. Velimirović, M. Banić and A. Miltenović, "Design of windturbines drive train based on CVT", *Balkan Journal of Mechanical Transmissions*, vol. 1, iss. 1, pp. 46-56, 2011.
- [5] M. Radić, Z. Stajić, M. Spasić and N. Danković, "Introducing the concept of controlled self-excitation and synchronization of a cage-rotor induction generator driven by CVT-based wind turbine", SAUM 2012, Conference Proceedings, pp. 189-192, Nis, Serbia, 2012.
- [6] M. Radić, Z. Stajić and N. Floranović, "Effects of grid voltage variations on wind turbine's optimal CVT ratio", SAUM 2014, Conference Proceedings, pp. 221-224, Nis, Serbia, 2014.

# Some properties of electrochromic low vacuum evaporated copper(I) oxide thin films

Ratka Neshkovska<sup>1</sup>, Mimoza Ristova<sup>2</sup> and Julijana Velevska<sup>2</sup>

**Abstract** – Electrochromic copper(I) oxide thin films were deposited onto conductive glass substrates using low vacuum evaporation method. An electrochromic test device with aqueous solution as an electrolyte was constructed. The transmittance spectra of the films in their three different states: as-prepared, coloured and bleached in the visible part of spectrum were recorded and the data were used to estimate of the energy band gap.

**Keywords** – Copper(I) oxide, Thin films, Electrochromism, Energy band gap.

## I. INTRODUCTION

Copper oxides are available in two stable forms, copper(I) oxide (or cuprous oxide)  $\text{Cu}_2\text{O}$  and copper(II) oxide,  $\text{CuO}$  (or cupric oxide). These two oxides have two different colours, crystal structures and physical properties.

Copper(I) oxide,  $\text{Cu}_2\text{O}$ , is a semiconductor with a p-type conductivity due to copper vacations [1-3]. The presence of  $\text{Cu}^{2+}$  ions in some places in the lattice instead  $\text{Cu}^+$  ions leads to  $\text{Cu}^+$  vacancies, resulting in a p-type conductivity [1-3]. By using multiple methods,  $\text{Cu}_2\text{O}$  films with n-type conductivity can be produced. One way is by dipping the copper foil in a solution of  $\text{HCl}$  with a  $\text{pH}$  of 3 at  $40^\circ\text{C}$  [1-3], or by potentiostatic electrodeposition.

The optical properties of copper(I) oxide depend on the methods and parameters of deposition. The density of copper (I) oxide ranges from  $5.88$  to  $6.06\text{ g/cm}^3$ . The melting temperature of this oxide is  $1235^\circ\text{C}$ , the boiling point is  $1800^\circ\text{C}$ , the molar mass is  $143.09\text{ g/mol}$ . Its crystal structure is cubic, with a lattice constant of  $4.27 \cdot 10^{-10}\text{ m}$ . The unit cell of this oxide is shown in Fig. 1. The unit cell consists of four copper atoms and two oxygen atoms. The coordinates of the copper and oxygen atoms are as follows [1-3]:

Cu:  $1/4\ 1/4\ 1/4$ ;  $3/4\ 3/4\ 1/4$ ;  $3/4\ 1/4\ 3/4$ ;  $1/4\ 3/4\ 3/4$

O:  $0\ 0\ 0$ ;  $1/2\ 1/2\ 1/2$ .

In this structure each copper atom only has two close oxygen neighbors, while each oxygen atom is surrounded (directly connected) with four copper atoms. The Cu-O

distance is  $1.85 \cdot 10^{-10}\text{ m}$ . Oxygen atoms are at a greater mutual distance ( $3.69 \cdot 10^{-10}\text{ m}$ ) and copper atoms Cu-Cu are at a smaller distance of  $3.01 \cdot 10^{-10}\text{ m}$ . On the other hand, each oxygen atom has six valence electrons, and each copper atom has one valence electron. Thus, any molecule of  $\text{Cu}_2\text{O}$  has eight valence electrons which are necessary in the formation of four homeopolar (covalent) bonds between each oxygen atom and the four surrounding copper atoms. The structure can be envisioned as two inter-penetrating cubic lattices: one face-centered with the lattice points occupied by Cu atoms and the other one body-centered with the lattice points occupied by O atoms. Each O atom is tetrahedrally coordinated by four Cu atoms, while each Cu atom is centered between two O atoms.

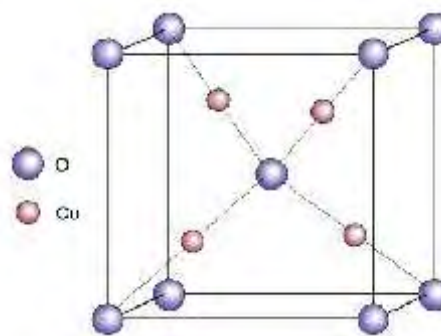


Fig. 1. One unit cell of the crystal structure of  $\text{Cu}_2\text{O}$ .

The energy band gap of  $\text{Cu}_2\text{O}$  is in the range from  $1.9\text{ eV}$  to  $2.40\text{ eV}$  [4]. The difference in the energy band gap is due to the different methods of preparation of films and the substances used. The lowest value of  $1.9\text{ eV}$  was measured in films deposited by the method of electrochemical deposition, and the highest value for the energy band gap was found in chemically deposited films. The acceptor type energy level is situated  $0.4\text{ eV}$  above the edge of the valence band of a  $\text{Cu}_2\text{O}$  single crystal prepared by oxidation of copper in air at a temperature of  $1000^\circ\text{C}$ . Two donor type trapping levels at  $1.3\text{ eV}$ - $1.34\text{ eV}$  and  $1.03\text{ eV}$ - $1.1\text{ eV}$  below the conduction band edge have also been identified [1-3]. The resistivity of copper(I) oxide strongly depends on the heating process of the oxide after its deposition and ranges from several tens of  $\Omega\text{cm}$  to  $10^{14}\ \Omega\text{cm}$ . The resistivity of the films obtained by the method of electrochemical deposition is in the range of  $10^9 - 10^{12}\ \Omega\text{cm}$ , while of the single crystal of copper(I) oxide ranges from  $10^2 - 10^4\ \Omega\text{cm}$ . This high specific resistance of

<sup>1</sup>Ratka Neshkovska is with the Faculty of Technical Sciences at "St. Kliment Ohridski" University of Bitola, Makedonska Falanga 33, Bitola 7000, R. Macedonia, E-mail: ratka\_neshkovska@yahoo.com.

<sup>2</sup>Mimoza Ristova and Julijana Velevska are with the Faculty of Natural Sciences and Mathematics at "Ss. Kiril&Metodij" University of Skopje, Arhimedova 10, Skopje 1000, R. Macedonia.

the film is due to the effect of compensation between the donor's and the acceptor's levels [4] and can be lowered by doping and heating the film.

The methods for deposition of copper(I) oxide often result in combined deposition of copper(II) oxide, CuO, which is also a semiconductor. Cu<sub>2</sub>O is transparent with a yellowish color and copper(II) oxide is black in color. The energy bandgap of CuO is in the range of 1.35eV to 1.7eV. Therefore Cu<sub>2</sub>O absorbs under 630nm and CuO absorbs strongly in the entire visible spectrum. The absorption coefficient of Cu<sub>2</sub>O at 510nm is of the order 10<sup>4</sup> cm<sup>-1</sup> [1-3].

Copper(I) oxide is one of the oldest semiconductor which attracted the attention of physicists. Its photoconductivity was first observed by Pfund in 1916 and since then has been subject to numerous studies. Interest in copper(I) oxide as semiconductor started with the discovery of Cu<sub>2</sub>O rectifiers by Grondhal in 1920 [1]. Properties of Cu<sub>2</sub>O and its characterization were widely investigated in the period from 1930 to 1940. After 1940 the interest in copper(I) oxide decreased, when the interest shifted to selenium, germanium, silicon and other materials. The interest in Cu<sub>2</sub>O returned in the 1970s when researchers started working in the field of photovoltaic conversion in semiconductors. The material was studied as a possible low cost material for solar cells. Cu<sub>2</sub>O remains an attractive alternative to silicon and other semiconductors producing cheaper solar cells. The main advantage of Cu<sub>2</sub>O is its non-toxic nature, abundance of copper as a starting material for the preparation of copper oxide and cheap and easy procedures for obtaining this oxide. The energy band gap of this oxide is acceptable and suitable for manufacturing photovoltaic cells. Also, the theoretical value of the coefficient of efficiency of such cells (13 to 20%) is a great motivation for exploring this material. Beside this application for solar energy conversion, copper-oxide-based materials have found applications in other various areas including heterogeneous catalysis, gas sensing and superconducting. They have been studied for photoelectrochemical hydrogen production from solar-driven water splitting [5] and for the development of a novel electrochemical sensor based on electrodeposition of copper oxide nanoparticles onto a carbon nanoparticle film modified electrode for the analysis of the anti-HIV drug transport [6].

It has been recently found that Cu<sub>2</sub>O thin films exhibit cathode electrochromism [7-11, 13]. The electrochromic materials reversibly and persistently change their optical properties and colour upon application of an external electrical field [7]. Cu<sub>2</sub>O thin films are transparent to visible light in their oxidized state and almost black in their reduced state.

Several methods for obtaining thin copper(I) oxide films have been developed, i.e. thermal oxidation, electrodeposition, sputtering, anodic and chemical oxidation, chemical oxidation, chemical deposition, sol gel technique, laser ablation, etc.

In this work electrochromic Cu<sub>2</sub>O thin films were prepared by low vacuum deposition method onto glass substrates precoated with fluorine-doped tin oxide (FTO) film.

## II. EXPERIMENTAL

### A. Deposition of the conductive layer

Because the purity of the substrates is essential for getting uniform colouring and bleaching of the films, standard microscopic glass slides were cleaned with immersion in a mixture of chromic and sulfuric acid and rinsed with distilled water. The conductive FTO coating was deposited onto glass substrates kept at 400<sup>o</sup> C with the spray pyrolysis technique, performed with a BOSH sprayer, spraying a 0.05 M aqueous solution of SnCl<sub>2</sub>·2H<sub>2</sub>O and crystals of NH<sub>4</sub>F, with a pH of about 7 [12]. The FTO prepared with this method was about 2μm thick, with an 80% transparency for the visible light, and with sheet resistance of 18 – 38Ω/sq .

### B. Deposition of Cu<sub>2</sub>O thin films by low vacuum deposition method

The prepared FTO substrates were subjected to the low vacuum evaporation method for deposition of Cu<sub>2</sub>O thin films [13].

The melting temperatures of Cu, CuO and Cu<sub>2</sub>O are 1083<sup>o</sup> C, 1326<sup>o</sup> C and 1235<sup>o</sup> C respectively. The temperature difference between elemental and oxides melting points makes the evaporation technique attractive for depositing of copper oxide thin films. Stable copper oxides can be deposited over a source temperature region of 1090–1350<sup>o</sup> C. In general, the oxidation rate and the evaporation rate of the source material define the deposition rate of the product grown on the substrate surface. Copper oxide thin films were deposited on glass substrates pre coated with FTO by the low vacuum evaporation technique, using granulated (0.3mm) copper with 99.8 % purity as a starting material. Deposition was performed in a Balzers Automatic BA 510 apparatus. The distance between the source (copper) and the substrates was about 50cm. The evaporation pressure was 0.67 Pa. During evaporation the transparency of the substrates changed by 50 %. The deposition lasted several minutes, until all of the starting material from the tungsten boat was evaporated. If the vacuum in the chamber is very high, pure copper will be deposited on the glass substrates. In our case the vacuum was not so high, so evaporated copper reacted with the gasses in the chamber. As a result of that reaction cuprous oxide thin films were obtained on the substrates. Some films were deposited onto conducting glasses and some were deposited onto glasses without conductive oxide film onto them.

The deposited films were yellowish in appearance. The film thickness, measured by weight different method was about 130nm [13].

The films deposited onto glass without a transparent conducting layer were characterized by X-ray diffractometer [13].

The constructed electrochromic test devices (ECTD) consisted of a glass substrate/FTO/Cu<sub>2</sub>O film as a working



electrode, a 0.1 M NaOH aqueous solution as an electrolyte and an FTO/glass substrate as a counter electrode.

The visible transmission spectra were recorded for the three film states: as-deposited, coloured and bleached, using a Varian CARY 50 Scan UV-Visible spectrophotometer in the wavelength range from 330 to 850 nm. Prior to the recording of the spectrum, the transmission spectrum for the glass/FTO substrate (blank probe) was recorded and it was normalized to 100 % transmission. The data were used to calculate absorption coefficient spectrum and the optical band gap for the films in their coloured and bleached state.

### III. RESULTS AND DISCUSSION

The prepared Cu<sub>2</sub>O films revealed cathode electrochromism. They were coloured at a negative potential of -4 V and were bleached at a positive potential of +4 V.

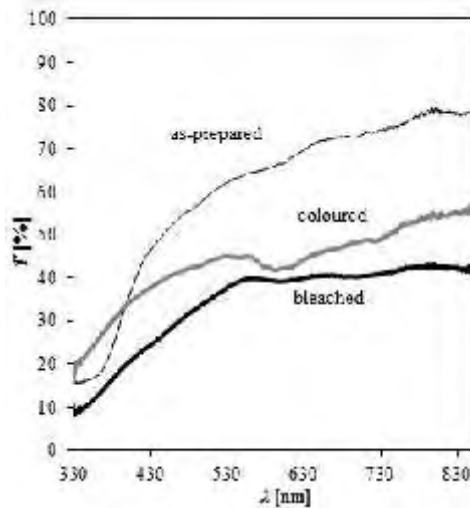


Fig. 2. Visible transmittance spectra of low vacuum evaporated thin Cu<sub>2</sub>O films in: as-prepared, coloured at -4 V and bleached at +4 V states.

The transmission spectra of the films in as-prepared, coloured and bleached state are presented in Fig. 2. The difference in the transmittance in the three states of the films is obvious.

The transmittance data were used to evaluate the absorption coefficient  $r(h\nu)$  and the band gap  $E_g$  for films in coloured and bleached state. The absorption coefficient  $r(h\nu)$  for a given film with thickness  $d$  is related with the photon's energy  $h\nu$  and the optical band gap  $E_g$ :

$$(rh\nu)^{2/n} = A(h\nu - E_g) \quad (1)$$

where  $A$  is a constant, and  $n$  is a number which depends on the type of transition. The absorption coefficient was evaluated from:

$$r = \frac{1}{d} \ln \frac{100}{T(\%)}, \quad (2)$$

where  $T$  are transmittance data values. Calculated values of the coefficient  $r$  versus wavelength of incident radiation are presented on Fig. 3 and Fig. 4.

The value of  $n$  and the type of transition were estimated from the graphical presentation of  $(rh\nu)^{2/n}$  versus photon energy (Fig. 5 and Fig. 6). Linear dependence of  $(rh\nu)^{2/n}$  versus  $h\nu$  is obtained for  $n = 1$ , which corresponds to direct band to band electron transition mechanism.

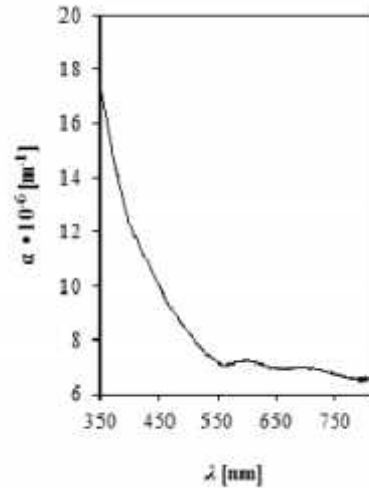


Fig. 3. Absorption coefficient in visible part of spectrum for low vacuum evaporated thin Cu<sub>2</sub>O films in coloured state.

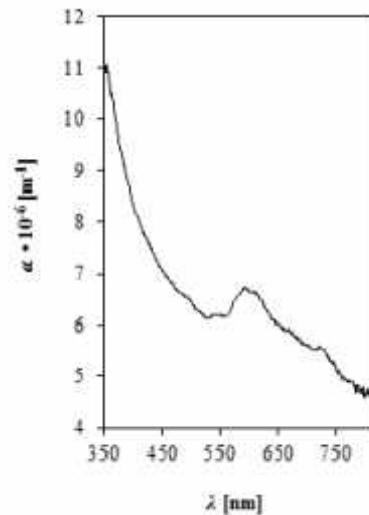


Fig. 4. Absorption coefficient in visible part of spectrum for low vacuum evaporated thin Cu<sub>2</sub>O films in bleached state.

The band gap energy was evaluated from the graphs in Fig. 5 and Fig. 6, as interception of the linear part with the photon energy axis. The evaluated band gap values for the prepared film by method of low vacuum evaporation in coloured state

are 2.83eV and 2.77eV and 0.8eV in bleached state. They are given in Table I. The band gap for the as-prepared films was evaluated elsewhere [13]. The obtained values are higher than reported values for films obtained with other methods of preparation [1-4].

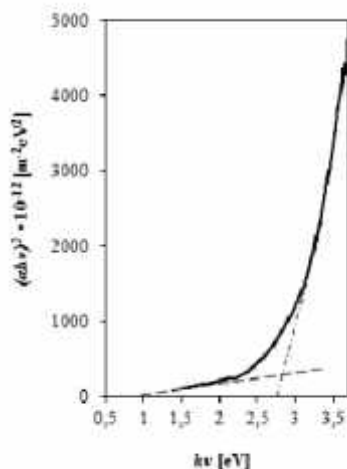


Fig. 5.  $(\alpha h\nu)^2$  versus photon energy in visible part of spectrum for low vacuum evaporated thin  $\text{Cu}_2\text{O}$  films in coloured state. Intercepts of the dashed lines corresponds to the direct band.

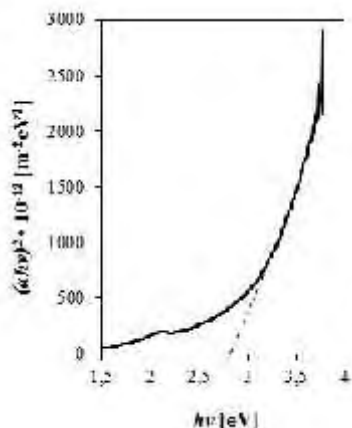


Fig. 6.  $(\alpha h\nu)^2$  versus photon energy in visible part of spectrum for low vacuum evaporated thin  $\text{Cu}_2\text{O}$  films in bleached state. Intercepts of the dashed lines corresponds to the direct band gaps.

#### IV. CONCLUSION

In this paper a method of low vacuum evaporation for depositing of  $\text{Cu}_2\text{O}$  thin films onto conductive and transparent FTO pre-coated glass substrate is given. The deposited films exhibited cathode electrochromism. The difference in the transmittance of films in the visible region in their coloured and bleached state is obvious. The evaluated values of the band gap in the coloured and bleached states of films are different and are higher than those presented in literature. This investigation of energy band gap shift may help mechanism of electrochromism to be understood in future. The thin films

prepared by the presented method could be applied in solar cells or other devices.

#### ACKNOWLEDGEMENT

We would like to express our gratitude to Dr. Metodija Najdoski for the transmittance spectra.

#### REFERENCES

- [1] A.E. Rakhshani, Solid-State Electronics Vol. 29, No. 1 (1986) 7-17.
- [2] C. Jayewardena, K. P. Hewaparakrama, D. L. A. Wijewardena, H. Guruge, Solar Energy Materials & Solar Cells 56 (1998) 29-33.
- [3] L. C. Olsen, F. W. Addis, W. Miller, Solar Cells 7 (1982-1983) 247-279.
- [4] X. Mathew, N. R. Mathews, P. J. Sebastian, Solar Energy Materials & Solar Cells 70 (2001) 277-286.
- [5] Le Chen, S. Shet, H. Tang, H. Wang, T. Deutsch, Y. Yan, J. Turner and M. Al-Jassim, J. Mater. Chem., 2010, 20, 6962-6967.
- [6] S. Shahrokhian, R. Kohansal, M. Ghalkhani, M. K. Amini, Electroanalysis, Wiley Online Library, DOI: 10.1002/elan.201500027.
- [7] H. Demiryont, US Patent 4, 830,471.
- [8] C. G. Granqvist, Handbook of Inorganic Electrochromic Materials, Elsevier, Amsterdam, 1995.
- [9] R. Neškovska, M. Ristova, J. Velevska, M. Ristov, Thin Solid Films 515 (2007) 4717-4721.
- [10] M. Ristova, R. Neškovska, V. Mir eski, Solar Energy Materials & Solar Cells 92 (2007) 1361-1365.
- [11] M. M. Ristova, V. Mirceski, R. Neshkovska, Journal of Solid State Electrochemistry, 2014, DOI 10.1007/s10008-014-2666-x.
- [12] M. Ristov & co. Annuaire Physique 47 p. 113 (1997).
- [13] R. Neshkovska et al., Horizons, Series B, Volume 1 p. 25-32 (2014).

# FTIR spectra of electrochromic copper(I) oxide thin films prepared by different methods

Ratka Neshkovska<sup>1</sup>, Mimoza Ristova<sup>2</sup> and Julijana Velevska<sup>2</sup>

**Abstract** – Electrochromic copper(I) oxide thin films were deposited onto conductive glass substrates using chemical bath deposition method and electrochemical method. An electrochromic test device with aqueous solution as an electrolyte was constructed. The reflectance spectra of the films in their three different states: as-prepared, coloured and bleached in the infrared part of spectrum were recorded.

**Keywords** – Copper(I) oxide, Thin films, Electrochromism, FTIR spectra.

## I. INTRODUCTION

Electrochromism is defined as a change of the optical properties of the materials when a voltage is applied across them. The optical properties should be reversible, i.e., the original state should be recoverable if the polarity of the voltage is changed [1]. The electrochromic materials change their colour, transmittance, absorbance and reflectance. Many inorganic and organic materials exhibit electrochromic properties. Electrochromic materials are incorporated in electrochromic (EC) devices, consisting of two electrodes separated with an ion conductor which can be a liquid electrolyte or a solid inorganic or organic (polymeric) material. The liquid electrolytes are convenient for research work, while solid ones are preferred for practical devices. The electrochromic material is deposited onto a conductive working electrode. Each of the films in the device can have a thickness of less than one micrometer. Electrochromic materials could reveal cathodic or anodic electrochromism, depending on the potential at which they darken. Cathodic electrochromic materials show cathodic coloration state, under the application of a negative potential and are in bleached state on a positive potential. Anodic electrochromic materials coloured at a positive potential and bleached at a negative potential. The coloration/bleaching process results from the insertion/extraction of electrons and charge balancing small ions ( $H^+$ ,  $Li^+$ ,  $Na^+$  and  $K^+$  are the best) from the electrolyte into or out of the material. Cathodic electrochromic materials possess reduced coloured and oxidized bleached state, while anodic materials are coloured in their oxidized state and bleached in their reduced state.

This phenomenon was discovered about 50 years ago and

some materials are widely investigated, their electrochromic properties are improved, new materials are synthesized and EC devices are fabricated. The applications of EC materials has been extended to smart windows, displays, antiglare mirrors and active camouflages [2]. Among them, smart windows represent an important application because they can effectively save energy by regulating solar heat gain, and provide indoor comfort by reversible color changes.

Numerous inorganic and organic electrochromic materials are widely investigated to develop energy efficient windows. If an electrochromic thin film is deposited onto a conductive window, then it is possible to control the transmission or reflection properties of the window, by applying a small voltage of several volts and thus saving energy for heating or cooling of the buildings. Electrochromic materials possess open circuit memory, therefore their coloured/bleached state and optical properties remain stable for a long time, which means that electrical current needs to be applied only during the colouring or bleaching of the film.

## II. EXPERIMENTAL

### A. Deposition of the conductive layer

The conductive layers are necessary for all EC device types. In all EC devices the electrochromic thin films should be deposited on a conductive layer. If the purpose of the EC device is to modulate the incident radiation, then the conductive layer should be transparent and deposited onto a transparent substrate, such as glass substrate or plastic. The quality and cost of transparent conductors is essential for all electrically activated devices. They must have low resistivity (below  $100\Omega/sq$ ), to have good conductivity for ions and electrons, to be transparent and thermally and chemically stable. The conductive fluorine-doped tin oxide,  $SnO_2:F$  (FTO) film meets these requirements. In this work FTO coating was deposited onto cleaned standard microscopic glass substrates kept at temperature of  $400^\circ C$ , by spraying a 0.05 M aqueous solution of  $SnCl_2 \cdot 2H_2O$  and crystals of  $NH_4F$ , with a pH of about 7 [3], with a BOSH sprayer. The FTO prepared with this method was about  $2\mu m$  thick, with an 80% transparency for visible light, and with sheet resistance of  $18 - 38\Omega/sq$ .

### B. Deposition of $Cu_2O$ thin films by chemical bath deposition method

The electrochromic  $Cu_2O$  films were deposited onto glass substrates pre-coated with FTO by successive immersion into

<sup>1</sup>Ratka Neshkovska is with the Faculty of Technical Sciences at "St. Kliment Ohridski" University of Bitola, Makedonska Falanga 33, Bitola 7000, R. Macedonia, E-mail: ratka\_neshkovska@yahoo.com.

<sup>2</sup>Mimoza Ristova and Julijana Velevska are with the Faculty of Natural Sciences and Mathematics at "Ss. Kiril&Metodij" University of Skopje, Arhimedova 10, Skopje 1000, R. Macedonia.

two solutions, a hot one and a cold one [4]. The hot solution was 2 M NaOH at 60–80°C. The cold solution contained a colourless complex mix of 1 M CuSO<sub>4</sub> and 1 M Na<sub>2</sub>S<sub>2</sub>O<sub>3</sub>. The number of immersions determines the thickness of the films, with each immersion yielding 0.1 µm in the film thickness. This procedure lasted until the desired thickness of the film (200 nm) was obtained.

#### C. Deposition of Cu<sub>2</sub>O thin films by electrochemical deposition method

Electrodeposition is a simple and low cost method for deposition of oxides at low temperatures. Such deposited films have high quality. In our case the deposition was performed using a two electrode classical system [5]. A copper clad for printed circuit board, with a thickness of 50 µm with the same dimensions as a classical microscopic glass was used as an anode, while the glass substrate coated with FTO was a cathode. The electrolyte was prepared from 0.4 M anhydrous copper sulfate CuSO<sub>4</sub>, 2.7 M lactic acid C<sub>3</sub>H<sub>6</sub>O<sub>3</sub> and 3.1 M sodium hydroxide NaOH. The temperature of the electrolyte was maintained at 60°C, the voltage between the electrodes was in the 0.5 to 0.6 V range, with a current density of 0.8 to 1 mA/cm<sup>2</sup>.

#### D. Recording of the FTIR spectra

All spectroscopic techniques can be used to identify and study chemicals. Infrared (IR) spectroscopy exploits the fact that molecules absorb specific frequencies that are characteristic of their structure. A common laboratory instrument that uses infrared light is Fourier transform infrared spectrometer (FTIR). FTIR is an interferometer for IR light. The incoming IR light beam is divided into two separate beams in a beam splitter where half of the beam is reflected to a fixed mirror and half is transmitted through a splitter and is directed to a movable mirror. After the reflection from the mirrors, the two light beams are directed to the beam splitter and pass through the sample. The detector records the transmittance or reflectance as a function of wavelength, or wavenumber 1/λ.

The reflectance FTIR spectra for the films prepared by two different techniques, chemical bath deposition and electrochemical deposition were taken using PERKIN ELMER System 2000 FT-IR in the wavenumber range of 400 to 4000 cm<sup>-1</sup>. Spectra were taken for the films in three different states: as deposited, coloured and bleached. In order to perform colouring and bleaching of the films, an electrochromic test device (ECTD) was constructed, consisting of a glass substrate/FTO/Cu<sub>2</sub>O film as a working electrode, a 0.1 M NaNO<sub>3</sub> aqueous solution as an electrolyte and an FTO/glass substrate as a counter electrode.

The films were coloured at negative potential of -4.5 V and bleached at a positive potential of +4 V. Prior to the recording of the spectra, the reflectance spectrum for the glass/FTO substrate was recorded.

### III. RESULTS AND DISCUSSION

The prepared Cu<sub>2</sub>O films revealed cathode electrochromism [6-8]. They were coloured at a negative potential of -4 V and were bleached at a positive potential of +4 V. The spectra were taken ex-situ.

The reflectance FTIR spectrum for the glass/FTO is given in Fig. 1. Fig. 2 represents the reflectance FTIR spectra for the film prepared by chemical bath deposition method in its as deposited, coloured and bleached states. The same spectra for the electrochromic Cu<sub>2</sub>O film prepared by electrodeposition method are presented in Fig. 3.

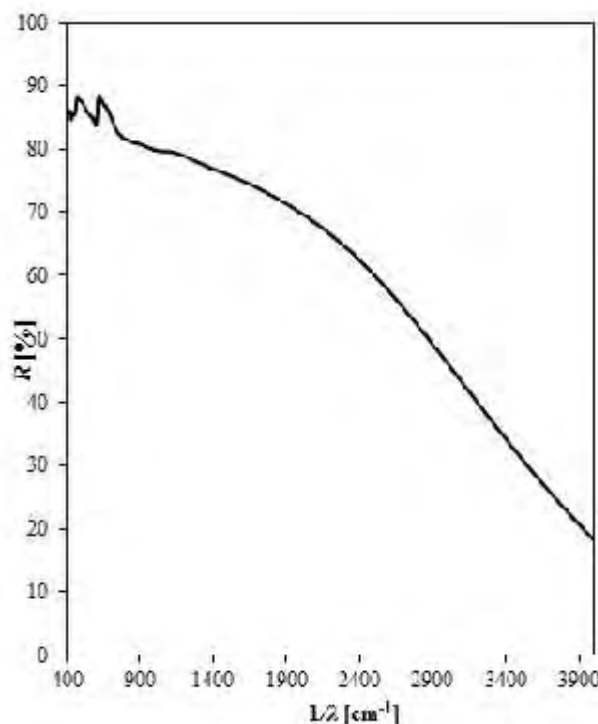


Fig. 1. FTIR reflectance spectra of glass/FTO.

From the given spectra one can notice that they do not exhibit peaks characteristic for Cu<sub>2</sub>O, so they can't be used for chemical identification of the Cu<sub>2</sub>O films or for the possible ion intercalation/deintercalation processes. The reason for this may be a small thickness of the films. However, they are useful for determining the reflective properties of the films in the infrared part of the spectrum. The difference in the reflection coefficients of the films prepared by two different methods in coloured and bleached states is evident for a wavenumber value over about 2400 cm<sup>-1</sup>.

The electrochromic smart windows are the most desired application of the electrochromic materials. With the application of a small current or voltage the windows can dynamically darken and lighten, so daylight, solar heat gain and internal heat loss through the windows of buildings and vehicles could be easily controlled and energy could be saved. The calculations showed that 4.5% of the annual energy use in the USA could be saved with highly insulating dynamic windows applied in both the commercial and residential sector



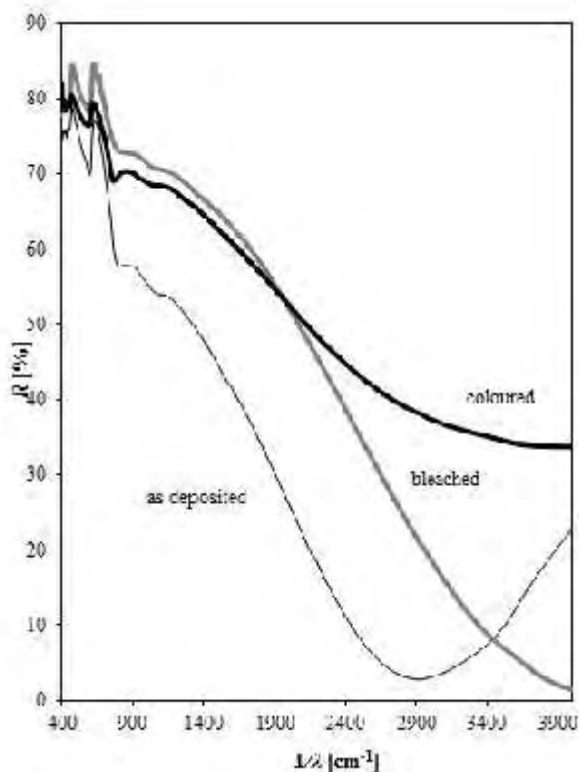


Fig. 2. FTIR reflection coefficients in the infrared part of the spectrum for the chemically deposited thin  $\text{Cu}_2\text{O}$  film in its as deposited, coloured and bleached states.

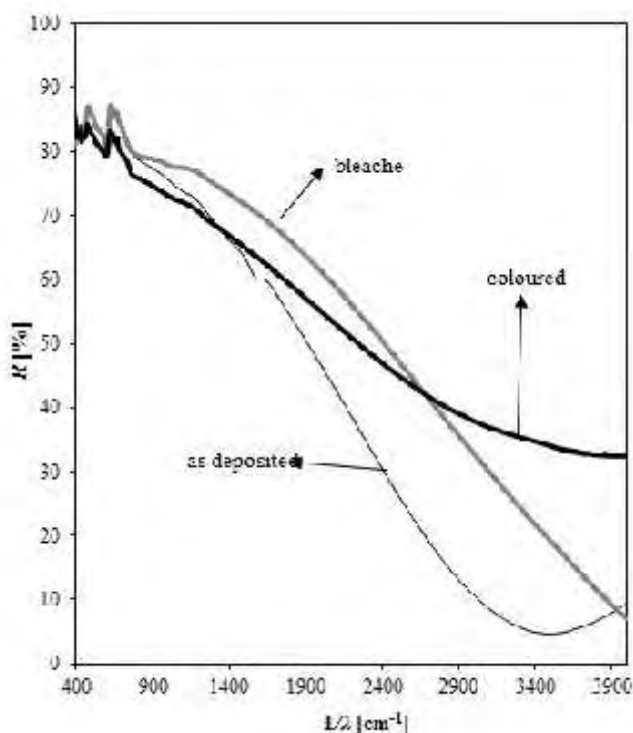


Fig. 3. FTIR reflection coefficients in the infrared part of the spectrum for the electrochemically deposited thin  $\text{Cu}_2\text{O}$  film in its as deposited, coloured and bleached states.

[9]. The colour of the electrochromic film is essential, because windows should be transparent enough to provide good illumination of the working and living area, but they should also reduce the amount of heat that enters the room (in summers) or coming out of it (in winters). From the reflectance FTIR spectra the modulation of the incident solar spectrum in the infrared part of spectrum could be calculated. The films with higher reflectance are preferred for application in smart windows.

#### IV. CONCLUSION

$\text{Cu}_2\text{O}$  thin films were deposited onto conductive and transparent FTO pre-coated glass substrate by two different methods. The deposited films exhibited cathode electrochromism. The FTIR spectra of the films in three different states: as prepared, coloured and bleached were taken ex situ. The difference in the reflection coefficient in the infrared part of the spectrum is obvious for values of the wavenumber over  $2000 - 2400 \text{ cm}^{-1}$ . These films may find their own application in electrochromic devices, such as smart windows, as an alternative of other thin films (for example the most explored  $\text{WO}_3$  thin films) because of their cost, simple methods of preparation, non toxic nature and abundance of the starting material.

#### ACKNOWLEDGEMENT

We would like to express our gratitude to Dr. Vladimir Petrushevski for the transmittance spectra and Verka Georgieva for electrodeposition.

#### REFERENCES

- [1] C. G. Granqvist, Handbook of Inorganic Electrochromic Materials, Elsevier, Amsterdam, 1995.
- [2] J. Wang, X. W. Sun, Z. Jiao, *Materials* **2010**, 3, 5029-5053.
- [3] M. Ristov & co. *Annuare Physique* 47 p. 113 (1997).
- [4] M. Ristov, G.J. Sinadinovski, I. Grozdanov, *Thin Solid Films* 123 (1985) 63.
- [5] Georgieva, M. Ristov, *Solar Energy Materials & Solar Cells* 73 (2002) 67-73.
- [6] R. Neškovska, M. Ristova, J. Velevska, M. Ristov, *Thin Solid Films* 515 (2007) 4717-4721.
- [7] M. Ristova, R. Neškovska, V. Mir eski, *Solar Energy Materials & Solar Cells* 92 (2007) 1361-1365.
- [8] M. M. Ristova, V. Mirceski, R. Neskovska, *Journal of Solid State Electrochemistry*, 2014, DOI 10.1007/s10008-014-2666-x.
- [9] T. Gillaspie, R. C. Tenent, A. C. Dillon, *J. Mater. Chem.*, 2010, 20, 9585-9592.

This Page Intentionally Left Blank

# Control Systems





# Multicomponent Analysis of Basic Physico-chemical Parameters of Bulgarian Yoghurt

Mariyana Sestrimska<sup>1</sup>, Tanya Titova<sup>2</sup>, Veselin Nachev<sup>3</sup> and Chavdar Damyanov<sup>4</sup>

**Abstract** - In this paper are considered basic methods for research the properties of multicomponent mixtures based multifactor statistical regression models.

**Keywords** - regression models, mixture design, simplex centroid design, experiments with mixtures, principal component analysis

## I. INTRODUCTION

Modern food production is characterized with multiparametric composition on the technology of the product. Requirements for quality and safety of products require the compilation of reliable multifactor models. This requires the use of methods for assessing the parameters of the research object, optimization of specify qualitative and quantitative characteristics and study of their relationship with different instrumental and experimental factors.

The task of optimization of products focuses on modeling mixture through a finite number of experiments.

The purpose of modeling is to clarify the influence of individual components on the quality of food to provide for the nutritional properties of the finished product, and create a product with optimal quality characteristics [1,2].

In the present paper is committed multi-parameter study of mixtures in a specific task to investigate the effect of adding fruit mixture on basic physicochemical parameters of Bulgarian yoghurt.

## II. MATERIALS AND METHODS

### A. Problem definition

Statistical optimization of products in the food industry, uses the analogy of the problem of the synthesis of multicomponent mixtures in many other areas - metallurgy, medicine, chemical industry and others. The research product

is presented as the sum of its constituent components. The total quantity of mixture is taken to be equal to one (or 100%):

$$\sum_{i=1}^q x_i = 1, \quad 0 \leq x_i \leq 1, \quad i = 1, 2, \dots, q, \quad (1)$$

where:  $x_i$  is  $i$ - th component in the mixture,  $q$ - total of the components constituting the mixture.

These conditions determining factor space, require special coordinate system for presenting multicomponent mixture. It is called *barycentric coordinate system* or *simplex diagram* of the composition of the mixture. It consists of  $(q-1)$  - dimension simplex defined area of change components (factors). For a three-component mixture barycentric coordinate system is an equilateral triangle presented in Figure 1 [1, 6, 7].

Experimental-statistical, simplex centroid plan for a three-component mixture is presented in Table 1.

TABLE I  
A SIMPLEX CENTROID DESIGN FOR A MIXTURE SYSTEM  
WITH THREE COMPONENTS

№	Components			Coordinates	Response
	$x_1$	$x_2$	$x_3$		$y$
1	1	0	0	(1, 0, 0)	$y_1$
2	0	1	0	(0, 1, 0)	$y_2$
3	0	0	1	(0, 0, 1)	$y_3$
4	0.5	0.5	0	(0.5, 0.5, 0)	$y_{12}$
5	0.5	0	0.5	(0.5, 0, 0.5)	$y_{13}$
6	0	0.5	0.5	(0, 0.5, 0.5)	$y_{23}$
7	0.33	0.33	0.34	(0.33, 0.33, 0.33)	$y_{123}$

Simplex-centroid plans are saturated plans, i.e. the number of coefficients of the model is equal to the number of experiments in the plan. On the basis of symmetrical coordinates of the model in barycentric coordinate system (Fig.1) are derived formulas for calculating the coefficients of the polynomial model.

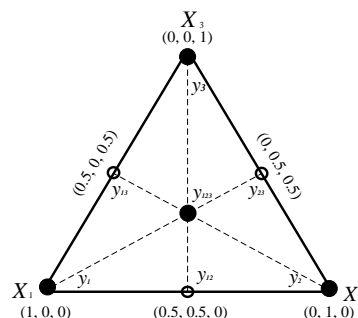


Fig.1 Simplex - centroid baricentric diagram of a three-component mixture

<sup>1</sup>Mariyana Sestrimska, Department Automation and Control Systems University of Food Technologies, 26 Maritza Blvd, Plovdiv, 4002 Bulgaria, e-mail: mariana\_brusova@abv.bg

<sup>2</sup>Tanya Titova, Department Automation and Control Systems University of Food Technologies, 26 Maritza Blvd, Plovdiv, 4002 Bulgaria, e-mail: t\_titova@abv.bg

<sup>3</sup>Veselin Nachev, Department Automation and Control Systems University of Food Technologies, 26 Maritza Blvd, Plovdiv, 4002 Bulgaria, e-mail: v\_nachevbg@yahoo.com

<sup>4</sup>Chavdar Damyanov, Department Automation and Control Systems University of Food Technologies, 26 Maritza Blvd, Plovdiv, 4002 Bulgaria, e-mail: chavdam@yahoo.com

The used model in our study is aligned cubic polynomial: i.e incomplete polynomial of third degree

$$\hat{y} = \sum_{i=1}^q b_i x_i + \sum_{1 \leq i < j \leq q} b_{ij} x_i x_j + \sum_{1 \leq i < j < k \leq q} b_{ijk} x_i x_j x_k, \quad (2)$$

where:  $\hat{y}$  a predicted value for research property,  $b_i$ ,  $b_{ij}$  and  $b_{ijk}$  are regression coefficients.

Coefficients of a mathematical model are obtained by replacing in equation (2) the coordinates of a point of the plan and the respective experimental value of the index  $y$  [4,5].

After obtaining regression equation is necessary to verification of the mathematical model. For this purpose they performed additional  $L$  (control) and the experience is determined experimentally properties of the mixture therein  $y_u$  ( $u=1,2,\dots,L$ ). The number of control points  $L$  is usually chosen by  $4 \div 6$ . The residual variance is determined by the expression:

$$S_{ad}^2 = \frac{1}{L} \sum_{u=1}^L (y_u - \hat{y}_u)^2, \quad (3)$$

where:  $y_u$  and  $\hat{y}_u$  are respectively experience and modal value of the property of the mixture  $y$  in control points  $u$ .

For determining the variation of the experiment is carried out  $n$  parallel experience in one of the points of the plan or any of the additional points. Reproducibility variance is:

$$S_e^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2; \quad (4)$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i. \quad (5)$$

Verification of adequacy is performed by Fisher criterion:

$$F = \frac{S_{ad}^2}{S_e^2}. \quad (6)$$

### B. Preparation of the samples.

Samples to be analyzed are structured in three separate groups. The first group is composed of the experimental samples, which consist of yoghurt (80%) and fruit pulp (20%) of the strawberries -  $x_1$ , raspberries -  $x_2$  и blueberries -  $x_3$ . Yogurt and frozen fruits are bought from commercial supermarket in Bulgaria. The third group of 35 samples of different brands of fruit yogurts bought retail chain. The fruits are finely pureed and pasteurized.

For obtaining experimental data, according to the statistical plan are prepared seven types of mixtures. First three mixtures were prepared by adding each of the types of fruit puree

separately. The next three mixtures satisfy the two-component mixture, i.e fruit ratio is 50:50% and the final mixture is a three-component, i.e. fruit puree were mixed in equal proportions 33: 33: 34%. The second group consists of 35 samples of different retail chain brands of yogurt that painted coloring food E122 (azorubine, carmoisine) which is authorized in Bulgaria. The third group of 35 samples of different brands of fruit yogurts bought from retail chain.

### C. Measurement on soluble solids and pH.

Dry matter is related to the content of protein, fat, carbohydrates and minerals in the product. By definition Brix degrees represent percentages by weight extract identified as equivalent to the sucrose content in the solutions or mixtures, i.e.:

$$1^\circ \text{Brix} = 1\text{g saccharose} / 100\text{g solution} = 1\% \text{ sugar solution.}$$

In this study the total soluble solids, expressed as the equivalent of sucrose is determined by a digital refractometer HANNA, HI96-801, at a temperature of 20°C.

To determine the pH of the analyzed samples used digital pH-meter PH-201.

### D. Determining the color characteristics

The color is the main organoleptic characteristics associated with quality and perception of food. To determine the color characteristics of the fruit yoghurts ( $L^*$ ,  $a^*$ ,  $b^*$ ,  $c^*$ ,  $h^*$ ) is used fiber-optic spectrophotometer USB4000 company Ocean Optics, equipped with a suitable sample holder, and the necessary optics.

The spectrophotometer is connected to a personal computer and specialized software (Spectra Suite), by which the data obtained are recorded in digital and graphic form. Software may change the number of repetitions waive the features, the number of averaged to obtain a spectrum step of scanning and frequency range.

## III. RESULTS AND DISCUSSION

Table 2 presents the values of the studied physico-chemical indicators of the fruit yogurt obtained according to accepted experimental and statistical plan (Table 1).

TABLE II  
BASIC PHYSICO-CHEMICAL PARAMETERS OF FRUIT MIXTURE

Group 1	N <sub>o</sub> on sample	$x_1$	$x_2$	$x_3$	pH	•Brix	$L^*$	$a^*$	$b^*$	$c^*$	$h^*$
	1	1	0	0	4.24	7.94	37.96	1.56	23.85	23.9	1.51
	2	0	1	0	4.17	7.40	48.71	2.33	18.13	18.28	1.44
	3	0	0	1	4.22	7.40	46.33	0.18	16.72	16.72	1.56
	4	0.5	0.5	0	4.20	7.76	43.04	4.59	15.98	16.63	1.29
	5	0.5	0	0.5	4.28	7.72	33.74	4.58	10.18	11.16	1.15
	6	0	0.5	0.5	4.21	7.52	36.29	5.03	12.68	13.64	1.19
	7	0.33	0.33	0.34	4.25	7.56	43.04	3.64	10.73	11.33	1.24

$x_1$  – strawberries puree;  $x_2$  – raspberry puree;  $x_3$  – puree of blueberries;  $a^*$  - color coordinates on the axis, corresponding to the red-green color, according to standard CIELAB;  $b^*$  - color coordinates on the axis corresponding to the yellow-blue;  $L^*$  - relative brightness (brightness),  $c^*$  - saturation,  $h^*$  - hue.

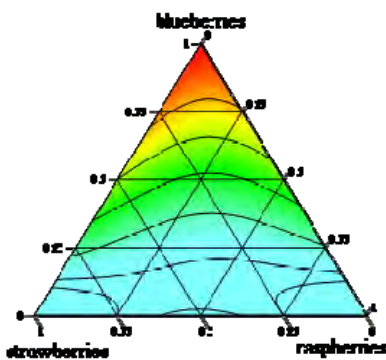


Fig.2 Mixture design for pH

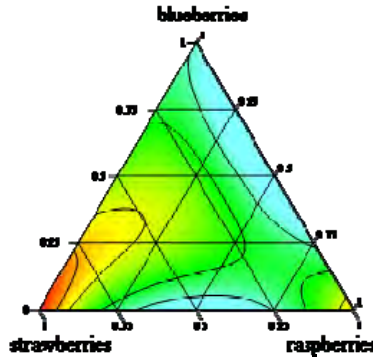


Fig.3 Mixture design for soluble solids

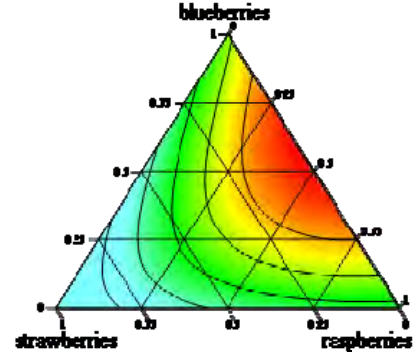


Fig.4 Mixture design for color descriptor L\*

Fruit purees used to make fruit yogurts Group I, affect the pH and color characteristics. In Fig. 2-4 show simplex isolines for the characteristics of pH,  $a^*$ ,  $b^*$ ,  $L^*$  and soluble solids.

Physico-chemical parameters of other two groups also differ samples (Table 2). Total sucrose content trademarks fruit yogurts is the highest that could be explained by further adding sugar production conditions or a different sort the fruit. The acidity of the tested samples ranged in very small ranges.

On the basis of experimental-statistical plan and the experiments are derived mathematical regression models for the content of soluble solids, pH and color characteristics.

TABLE III  
REGRESSION MODELS

$\hat{y}_1(pH) = 4.24x_1 + 4.17x_2 + 4.22x_3 + 0.02x_1x_2 + 0.20x_1x_3 + 0.06x_2x_3 + 0.36x_1x_2x_3$
$\hat{y}_2(^{\circ}Brix) = 7.94x_1 + 7.40x_2 + 7.40x_3 + 0.36x_1x_2 + 0.20x_1x_3 + 0.48x_2x_3 - 3.66x_1x_2x_3$
$\hat{y}_3(L^*) = 37.96x_1 + 48.71x_2 + 46.33x_3 - 1.18x_1x_2 - 33.62x_1x_3 - 44.92x_2x_3 + 204.24x_1x_2x_3$
$\hat{y}_4(a^*) = 1.56x_1 + 2.33x_2 + 0.18x_3 + 10.58x_1x_2 + 14.84x_1x_3 + 15.10x_2x_3 - 59.91x_1x_2x_3$
$\hat{y}_5(b^*) = 23.85x_1 + 18.13x_2 + 16.72x_3 - 20.04x_1x_2 - 40.42x_1x_3 - 18.98x_2x_3 - 0.27x_1x_2x_3$

The obtained coefficient of determination  $R^2$  for the models from Table 3 is not less 0.97. These coefficients of determination shows that the selected factors significantly affect the resulted sign, which confirms the correctness of their inclusion in the experimental - statistical model. For the assessment of practical suitability of regression equations are been made additional attempts at random points experimentally-statistical plan. On the basis of conducted an additional experiment were obtained dispersion and dispersion adequacy of the experience as follows:

$$S_{ad}^2(pH) = 3,25 \cdot 10^{-4} \quad \text{and}$$

$$S_{\varepsilon}^2(pH) = 2,2 \cdot 10^{-4}; S_{ad}^2(^{\circ}Brix) = 0,00141 \quad \text{and} \quad S_{\varepsilon}^2(^{\circ}Brix) = 0,012$$

The calculated values Fisher's criteria are respectively

$F(pH) = 1,49$  and  $F(^{\circ}Brix) = 0,118$ . Both values for Fisher's criteria are smaller than the critical ( $F_{critical} = 6,39$  for degrees of freedom  $v_{ad} = 4$  и  $v_e = 4$  level of significance  $\alpha = 0,05$ ). Therefore deduced based on experimental-statistical plan regression models were adequate and could be used to predict the values of soluble solids and pH at fruit mixtures.

Results from the analysis of samples by the Principal Components Analysis (PCA) are shown in Figure 6 and 7.

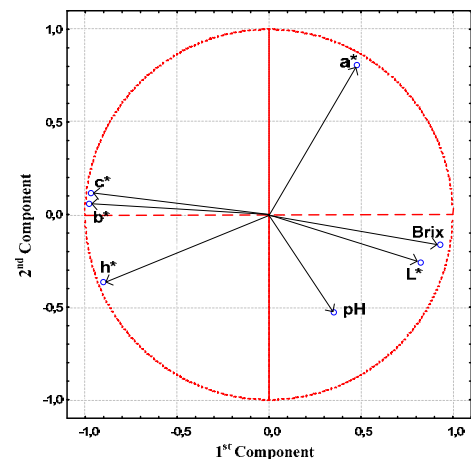


Fig.6

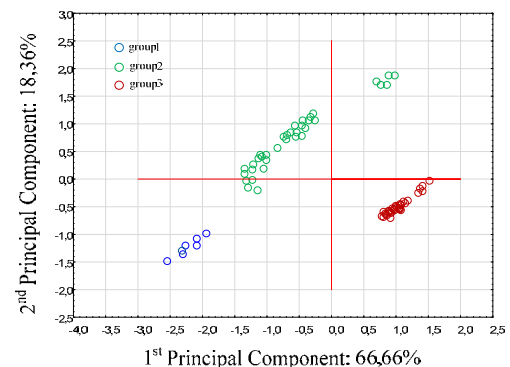


Fig.7

#### IV. CONCLUSION

In the present article discussed the influence of three types of fruit purees on basic physical-chemical parameters of Bulgarian yoghurt. Based on experimental and statistical simplex centroid plan for analyzing mixtures are derived adequate regression equations that could be applied to predict and obtain fruit yogurts with desired physico-chemical characteristics (pH, soluble solids, color).

#### ACKNOWLEDGEMENT

This research was supported in part by the framework of the 13/15-H project "Complex assessment of quality, safety, authenticity and healthiness of food by data fusion on from an electronic nose and electronic tongue and other sensors", financed by the Science Fund of the University of Food Technologies- Plovdiv.

#### REFERENCES

- [1] Хр. Йончев, И. Вучков, Планиране и анализ на експеримента при изследване на свойствата на смеси и сплави, Държавно издателство „Техника“, София, 1979
- [2] K. André, I., Response surface methodology and related topics, University of Florida, USA, 2005
- [3] J. Cornell, A., A primer on experiments with mixtures, John Wiley and Sons, Inc., New York, 2011
- [4] J. Cornell, A., Experiments with mixtures designs, models and the analysis of mixture data, John Wiley and Sons, Inc., 3ed., New York, 2002
- [5] S.Ferreira, R. Bruns, E. Da Silva, W. Lopes dos Santos, Statistical design and response surface techniques for the optimization of chromatographic systems, Journal of Chromatography, 2007
- [6] K.Mutai, J. Koske, J. Mutiso, G. Kerich, Optimal design for mixture experiments and their application in agricultural research, Journal of Biostatistics, 2012
- [7] P. Prescott, N. Draper, Balanced, nearly optimal mixture experiments for models with interactions with process variables, Quality Technology and quantitative management, 2009



# Comparative analysis of more commonly used recursive methods for parameter estimation in adaptive systems

Ivan V. Grigorov<sup>1</sup>, Nasko R. Atanasov<sup>2</sup>

**Abstract** – The adaptive control covers a set of methods that provide a systematic approach for automatic control in real-time. Recursive methods for parameter estimation have to meet the requirements for identification algorithms in real time. This is determined from the fact that the adjustment of the model after the submission of new data from monitoring, and the development of new control action should be made in a single cycle of discretization. In this article will be made an analysis of some of more commonly used recursive methods for parameter estimation in adaptive systems and their variations.

**Keywords** – adaptive system, instrumental variable method, least squares method, recursive methods for parameter estimation

Mandatory condition in the control of one system is understanding the dynamic features of the object.[3] Recursive methods for parameter estimation are used in the identification in real time. During the work on system together with the management can continue the specification of the model based on new data received. This is particularly important when the characteristics of the object change over time. These changes are recorded in the elaboration of the control signal from the so-called adaptive systems. Often used in the transmission and processing of signals.

Recursive methods are independent control tool in real time.[1]

## 1. Recursive versions of least squares method for parameter estimation.

### 1.1. Recursive weighted least squares

To obtain a recursive estimates on the weighted least squares is used

$$\hat{\theta}_{N+1} = \hat{\theta}_N + \frac{C_N f_{N+1}}{\frac{1}{w_{N+1}} - f_{N+1}^T C_N f_{N+1}} (y_{N+1} - f_{N+1}^T \hat{\theta}_N)$$

According to (1) the old value  $\hat{\theta}_N$  is adjusted proportionally to the difference  $y_{N+1} - \hat{y}_{N+1}$  with a vector coefficient of proportionality

$$\Gamma_{N+1} = \frac{C_N f_{N+1}}{\frac{1}{w_{N+1}} - f_{N+1}^T C_N f_{N+1}} \quad (2)$$

The value  $\hat{y}_{N+1} = f_{N+1}^T \hat{\theta}_N$  is the predicted value for  $y_{N+1}$ , calculated on the basis of the coefficients of the previous iteration  $\hat{\theta}_N$ .

Using remainder instead of prediction errors

In formula (2) is involved by the prediction error

$$e_{N+1} = y_{N+1} - \hat{y}_{N+1}(\hat{\theta}_N) = y_{N+1} - f_{N+1}^T \hat{\theta}_N \quad (3)$$

For improving the accuracy of prediction instead of  $e_{N+1}$  is used the remainder

$$r_{N+1} = y_{N+1} - f_{N+1}^T \hat{\theta}_{N+1} \quad (4)$$

From (2) and (3) follows

$$r_{N+1} - e_{N+1} = -f_{N+1}^T (\hat{\theta}_{N+1} - \hat{\theta}_N) \quad (5)$$

If from (2) is determined the difference  $(\hat{\theta}_{N+1} - \hat{\theta}_N)$  and substituted in (5), follows

$$r_{N+1} = \frac{\frac{1}{w_{N+1}} e_{N+1}}{\frac{1}{w_{N+1}} - f_{N+1}^T C_N f_{N+1}} \quad (6)$$

Starting the calculation procedure

To obtain the estimates by the method of weighted least squares are needed  $\hat{\theta}_N$  and  $C_N$  from previous calculations.

$N \geq k$  observations are collected and with non-recursive method of weighted least-squares and the initial estimates  $\hat{\theta}_N$  and  $C_N$  are. Then continue the calculations in (1) and (7)

$$C_{N+1} = C_N - \frac{C_N f_{N+1} f_{N+1}^T C_N}{\frac{1}{w_{N+1}} - f_{N+1}^T C_N f_{N+1}} \quad (7)$$

The algorithm of the recursive weighted least squares underlies a number of other recursive procedures. It is as follows:

- $N \geq k$  observations are collected and with non-recursive method of weighted least-squares and the initial estimates  $\hat{\theta}_N$  and  $C_N$  are.

- The new estimates are calculated by the formula

$$\hat{\theta}_{N+1} = \hat{\theta}_N + \frac{C_N f_{N+1}}{\frac{1}{w_{N+1}} - f_{N+1}^T C_N f_{N+1}} (y_{N+1} - f_{N+1}^T \hat{\theta}_N)$$

- Recalculating matrix  $C_{N+1}$  to prepare for the next iteration

$$C_{N+1} = C_N - \frac{C_N f_{N+1}^T f_{N+1}^T C_N + 1}{w_{N+1} - f_{N+1}^T C_N f_{N+1}} \quad (7)$$

- With The new estimates  $\hat{\theta}_{N+1}$  and  $C_{N+1}$  starts the next iteration from point 2 of the algorithm.

### 1.2. Recursive ordinary least squares.

Recursive least squares can be seen as a special case of RWLS at  $W = I$ . This means all weights to adopt equal unit, which is possible only if  $\rho = 1$ . Therefore, the recursive least squares (RLS) can be obtained, as in the procedure of paragraph 1.1. to lay everywhere  $\rho = 1$ . [1,3]

## 2. Recursive method of instrumental variables.

Finding estimates by the method of instrumental variables much like the least squares

The Instrumental vector  $v_{N+1}$  is chosen from (8), or (8')

$$v_t^T = [-\hat{y}_{t-1} \quad -\hat{y}_{t-2} \quad \dots \quad -\hat{y}_{t-n} \quad u_{t-1} \quad u_{t-2} \quad \dots \quad u_{t-m}], \quad (8)$$

$$v_t^T = [u_{t-1} \quad u_{t-2} \quad \dots \quad u_{t-m}], \quad (8')$$

Recursive algorithm method of instrumental variables follows:

- The evaluation starts with recursive weighted LS. This increases the resistance to initial estimates. After accumulating a certain number  $N > k$  observations we pass to the recursive method of instrumental variable. N is typically chosen to be equal to 3 to 5 times the number of coefficients.
- The estimated coefficients of  $N + 1$ st iteration are calculated by the formula

$$\hat{\theta}_{N+1} = \hat{\theta}_N + \frac{C_N v_{N+1}}{\rho + f_{N+1}^T C_N v_{N+1}} (y_{N+1} - f_{N+1}^T \hat{\theta}_N)$$

- The next iteration is prepared by recalculating  $C_{N+1}$  by the formula

$$C_{N+1} = \frac{1}{\rho} \left( C_N - \frac{C_N v_{N+1} f_{N+1}^T C_N}{\rho + f_{N+1}^T C_N v_{N+1}} \right) \quad (10)$$

- Calculations start from item 2 with the new estimates. If it is not necessary to weigh the observations in (9) and (10) is applied  $\rho = 1$ . [1]

## 3. Experimental research and results

Studies here relate to the performance, capabilities and reliability of already described recursive methods for parameter estimation in adaptive control systems.

For this purpose, the study used a random input signal dispersion unit and zero mathematical expectation.

Studies are made by using System Identification Toolbox in Matlab\Simulink. For this purpose are made blocks for estimation in real time for the recursive least squares(RLS), recursive least squares with using the remainder instead of the prediction error(RLSr) and

recursive instrumental variables (IVu) using S-functions.

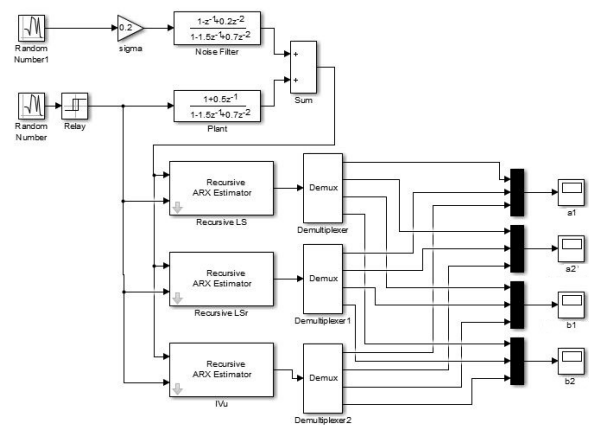


Fig. 1 - Block-diagram for recursive estimation with recursive estimator (RLS,RLSr,IVu)

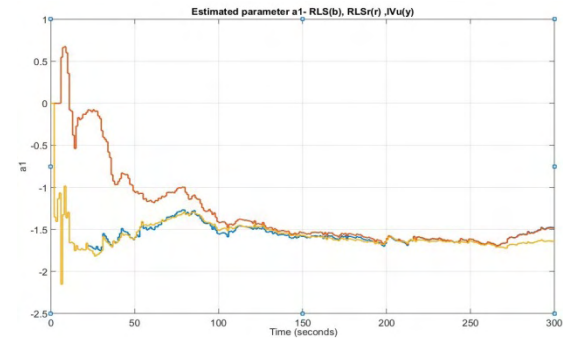


Fig. 2 – Estimated parameter a1- RLS(b),RLSr(r),IVu(y)

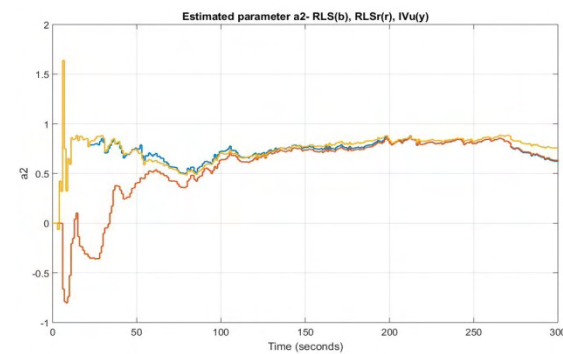


Fig. 3 – Estimated parameter a2- RLS(b),RLSr(r),IVu(y)

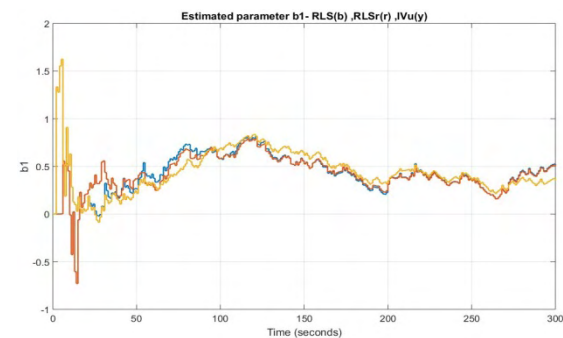
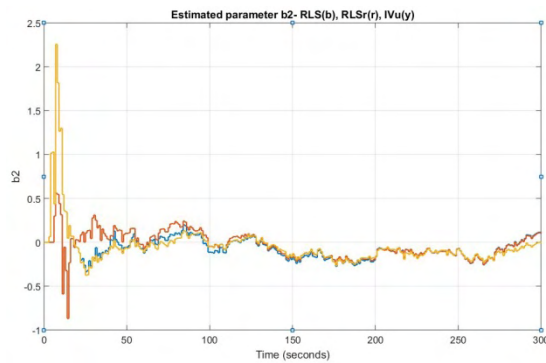


Fig. 4 – Estimated parameter b1- RLS(b),RLSr(r),IVu(y)



**Fig. 5** – Estimated parameter  $b_2$ - RLS(b),RLSr(r),IVu(y)

#### 4. Analysis and conclusions

- Recursive least squares method (blue line), recursive least squares with using the remainder instead of the prediction error (red line) and recursive instrumental variables (yellow line) are ones of the most reliable methods for parameter estimation and they can be modified for better performance and added to the adaptive control systems in real time
- When selecting a different weighting can be obtained robust variety of assessments that can be noise resistant.
- Recursive methods for parameter estimation in adaptive systems shown in this paper can be used in the field of reliability and diagnostics for fault detection and correction.

#### REFERENCES:

- [1]. Вучков. И.: Идентификация, ИК Юрамел, 1996
- [2]. Генов Г.: Идентификация на системи –теоретични основи, Варна, 2007
- [3]. Велев К.: Адаптивни системи, София, 1995
- [4]. Велев К., Хаджийски М.: Автоматизация на технологични процеси. Методи и алгоритми за управление, София, 1992
- [5] Allan Aasbjerg Nielsen, Least Squares Adjustment: Linear and Nonlinear Weighted Regression Analysis, 2013
- [6] Joshua D. Angrist and Alan B. Krueger: Instrumental Variables and the Search for Identification, 2001
- [7] Landau, I.D., Lozano R., M'Saad H., Kariirii A.: Adaptive Control Algorithms, Analysis and Applications, 2011
- [8] Naira Hovakimyan, Chengyu Cao: Adaptive Control Theory *Guaranteed Robustness with Fast Adaptation*, 2010

#### Contacts:

Ivan V. Grigorov,  
[ivan.grigorov.md@gmail.com](mailto:ivan.grigorov.md@gmail.com), phone: 0887189383;  
 Nasko R. Atanasov,  
[nratanasov@tu-varna.bg](mailto:nratanasov@tu-varna.bg), phone: 0899905045;

#### Acknowledgments:

The studies presented in this article were conducted in connection with the project NP6 at the Technical University of Varna, financed from the state budget for 2016

This Page Intentionally Left Blank



# Mobile Robot Guidance in Presence of Obstacles, using the Potential Fields Method

Vesna Antoska-Knights<sup>1</sup>, Zoran Gacovski<sup>2</sup>, Stojce Deskovski<sup>3</sup>

**Abstract** – In this paper we have researched and simulated the mobile robot guidance and control in the environment full of obstacles, by using the potential fields method. We have considered a known environment where fixed potentials were assigned to the goal and the obstacles. We have applied a potential field's method with one attraction potential assigned to the goal point and fixed repulsion points assigned to the obstacles. It moves successfully within different obstacle configurations (closely spaced obstacles), and it solves the problem with a local minimum occurrence.

**Keywords** – Mobile Robot Guidance, Obstacle avoidance, Potential Field Method.

## I. INTRODUCTION

The autonomous robot navigation problem consists of the determining the possible path between two points, an initial and a final (goal) point. The local navigation method should result in optimal (possibly shortest) path, avoiding the obstacles present in the working environment. Usually, the obstacles and the target could be fixed or dynamic. The goal of the path planning method is to determine the robot's moves but to avoid collisions while reaching the objective. Potential fields are elegant solutions to the path finding problem [5]. From different authors, different approaches are taken to calculate appropriate field configurations.

The robot uses functions defining potential fields at its position to calculate component vector [6]. The entire field doesn't have to be computed - only the portion of field affecting robot will be computed for each behavior of the potential field; the sum of the vectors at the robot's position will get resultant output vector. If the sum of the vectors is zero, that is the local minima; if the robot reaches local minima, it will just sit still. We can have issues with combining potential fields; the impact of update rates is that lower update rates can lead to "jagged" paths. If the robot is treated as a mass object, it cannot be expected to change velocity and direction instantaneously (cannot happen). We should find a solution for local minimum problem; if the global minimum is not guaranteed, we'll need to apply something else than gradient descent. The functions should be chosen in such a way that global minimum can be guaranteed and the robot will escape the local minima. "Avoid-past" behaviors can be included - remembering where robot has been and bringing the robot to other places. Numerical

techniques, Random walk methods and Navigation functions [7], i.e. "navigation templates" have been combined to get the resulting vector that will avoid local minima. They give "avoid" behaviour a preferred direction and insert tangential fields around obstacles.

This paper is organized as follows: in Section 2 we present the kinematic and dynamic model of the 4-wheeled mobile robot; in Section 3 - the Guidance and control method is presented, and in Section 4 - the trajectory planning algorithm using the potential field method. Finally we present the simulation results, our summary conclusions and directions for further work.

## II. MATHEMATICAL MODEL OF THE ROBOT

In this section we will elaborate the robot movement in horizontal plane. We have considered a mobile robot with 4 wheels (Ackerman drive), fixed rear wheels, and steering front wheels. Fig.2.1 shows the robot geometry while moving in 2D space. The kinematic and dynamic equations for the movement of the mobile robot can be found in [1,2,10,11], and here we'll use these equations without derivation.

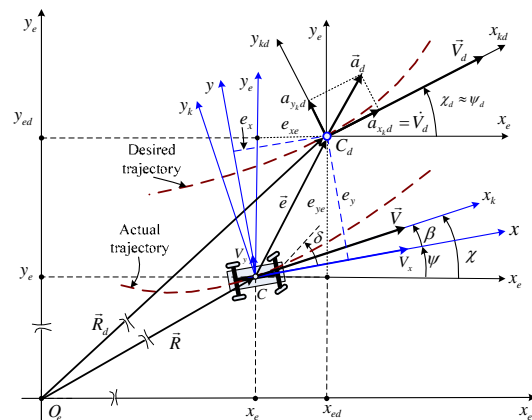


Fig.2.1 Kinematics of lateral vehicle motion: desired and actual trajectories of the mobile robot.

In many cases, when the kinematics and dynamic model of a car-like robot is required - a bicycle model is used, where front and rear vehicles are presented with a single wheel, as in Fig.2.2 [1].

The movement of the mobile robot in horizontal plane can be described as a system with 3-degrees of freedom- two translations of the mass center (C) along  $x$  and  $y$  axes, and one rotation around  $z$  axis. Three coordinate systems are used: inertial coordinate system - fixed to the ground (global)  $E(O_e; x_e, y_e)$ , coordinate system fixed to the robot body (local)  $B(C; x, y)$  and coordinate system attached to the

<sup>1</sup>Vesna Antoska-Knights is with Faculty of Technol. and Technic. Sciences – Veles, MK, E-mail: vesna\_antoska@gmail.com.

<sup>2</sup>Zoran Gacovski is with the European University-Skopje, MK, E-mail: zgacovski@yahoo.com.

<sup>3</sup>Stojce Deskovski is with Technical Faculty-Bitola, MK, E-mail: stojce.deskovski@gmail.com.

robot trajectory  $K(C; x_k, y_k)$ , where  $x_k$  is axis - tangent, and  $y_k$  is axis normal to the trajectory.

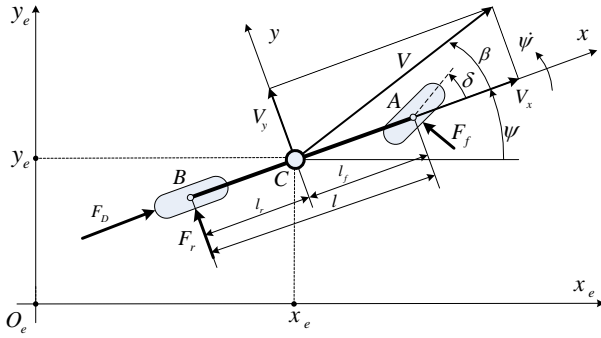


Fig.2.2 Bicycle model of lateral mobile robot dynamics

The kinematic equations of the mobile robots are [1]:

$$\begin{aligned}\dot{x}_e &= V \cos(\psi + \beta) \\ \dot{y}_e &= V \sin(\psi + \beta) \\ \dot{\psi} &= \frac{V \cos \beta}{l_f + l_r} \tan \delta\end{aligned}\quad (2.1)$$

where:  $V$  is velocity of center of gravity (c.g.) of the vehicle which is at point C,  $\psi$  is yaw angle (orientation angle with respect to global coordinate  $x_e$ ),  $\beta$  is vehicle slip angle,  $\chi = \psi + \beta$  is  $l = l_f + l_r$ , the angle of turn of the vehicle,  $\delta$  is the steering angle of front wheels,  $l_f$  and  $l_r$  are distances of points A and B from c.g. of the vehicle, respectively.

Dynamic equations can be derived from Fig. 2.2 and they are:

$$\begin{aligned}m(\dot{V}_x - V_y \dot{\psi}) &= -F_f \sin \delta + F_D \\ m(\dot{V}_y + V_x \dot{\psi}) &= F_f \cos \delta + F_r \\ J\ddot{\psi} &= l_f F_f \cos \delta - l_r F\end{aligned}\quad (2.2)$$

where  $m$  and  $J$  are mass and inertial moment of the vehicle around the mass center,  $F_D$  is driving force on the rear axis - along  $x$  axis,  $F_f$  and  $F_r$  are resultant side forces on the front and the rear wheel. The first two equations in (2.2) define the forces balance along the system axes  $B(C; x, y)$ , while the third equation in (2.2), defines the moments' balances around the  $z$  axis, normal to the plane  $Cxy$ .

Complete model of mobile robot includes kinematic (2.1) and dynamic (2.2) equations, equations of non-holonomic constraints, model of the steering wheels actuator, and model of DC motor for driving force  $F_D$  generation.

The mobile robot model obtained by the upper equations is displayed in Fig.3.2. Input in the model are the control signals  $u_1$  - for front wheels control,  $u_2$  - for velocity control. The output of the system is the complete state vector:

$\mathbf{x} = [x_e, y_e, \psi, V_x, F_D, \delta]^T$  and other variables that are dependent on the robot state.

### III. GUIDANCE AND CONTROL SYSTEM

The robot guidance and control system should provide accurate trajectory following which can be known in advance, or can be computed in real time.

Fig.3.1 shows the geometry during the referent trajectory following by the mobile robot. This figure can help us in determining the kinematic equations and the errors in following. The point  $C_d$  of the referent trajectory defines the required robot position given with the global coordinates  $x_{ed}, y_{ed}$ . The real position is determined by the point C (c.g. - center of gravity of mobile robot) whose coordinates are  $x_e, y_e$ . The error in following can be represented in the projections in the global  $\mathbf{e}_E = [e_{x_e}, e_{y_e}]^T$  and in local coordinates  $\mathbf{e}_B = [e_x, e_y]^T$ , and the transformation is:

$$\begin{aligned}\mathbf{e}_B &= \mathbf{T}_{BE} \mathbf{e}_E, \text{ or } [e_x, e_y]^T = \mathbf{T}_{BE} [e_{x_e}, e_{y_e}]^T, \\ \mathbf{T}_{BE} &= \mathbf{T}_\psi = \begin{bmatrix} \cos \psi & \sin \psi \\ -\sin \psi & \cos \psi \end{bmatrix}\end{aligned}\quad (3.1)$$

where  $\mathbf{T}_{BE}$  is transformation matrix from global  $E$  to local  $B$  coordinate system.

One way to generate the referent trajectory, given the accelerations along the trajectory tangent and normal respectively-  $a_{x_k d}, a_{y_k d}$ , is to apply the following equations:

$$\dot{V}_d = a_{x_k d}, \quad V_d(0) = V_{d0}, \quad \dot{\psi}_d = \frac{a_{y_k d}}{V_d}, \quad \psi_d(0) = \psi_{d0} \quad (3.2)$$

$$\dot{x}_{ed} = V_d \cos \psi, \quad x_{ed}(0) = x_{ed0}, \quad \dot{y}_{ed} = V_d \sin \psi, \quad y_{ed}(0) = y_{ed0} \quad (3.3)$$

The diagram of the control system is given in Fig.3.1. The „Mobile robot“ block computes the equations, the block „Steering wheel“ generates steering angle of front wheels, the block „Driving force“ generates the driving force  $F_D$ , the block „Transformation to robot coordinates“ computes the errors:

$$e_{x_e} = x_{ed} - x_e, \quad e_{y_e} = y_{ed} - y_e, \quad e_\psi = \psi_d - \psi, \quad (3.4)$$

transforming them in local coordinate system with (3.1), the „Reference trajectory“ block computes the referent trajectory. To obtain better guidance accuracy, especially when the trajectory needs fast maneuvers, together with the feedback controller - we must apply a feedforward controller based on the inverse robot model. The controller (its algorithm) generates control signal  $u_1$  to control the steering vehicles. To generate the signal  $u_1$  we have used a PID controller. The signal  $u_2$  defines the given velocity of the robot movement along the given trajectory.

Based on the above model and the block diagram shown on Fig.3.1, we have developed simulation model in Matlab/Simulink which can be used for testing of the guidance algorithm. The generation of the referent trajectory in an obstacle environment is described in the following section.

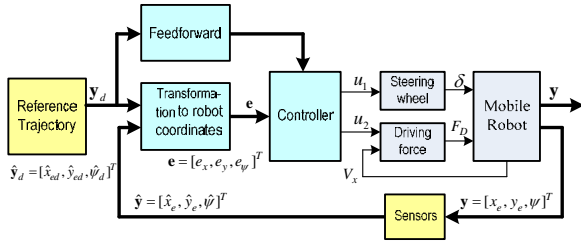


Fig.3.1 Blok diagram of mobile robot guidance and control system

#### IV. TRAJECTORY GENERATION BY POTENTIAL FIELDS METHOD

Programming a single potential field - can be done by a repulsive field with linear drop-off:

$$V_{direction} = 180^\circ$$

$$V_{magnitude} = \begin{cases} \frac{D-d}{D} & \text{for } d \leq D, \text{ where } D \text{ is the max range} \\ 0 & \text{for } d > D \end{cases}$$

of the field's effect. (4.1)

To generate potential field function  $U(q)$  for the robot - we will start with the force acting on a robot at point  $q$ :  $F(q) = -\nabla U(q)$ , and environment is represented by potential function:  $U(x, y)$ . Force is proportional to the gradient of the potential function:

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = -\nabla U(x, y) \quad (4.2)$$

Environment is sum of the potential of fields:

$$U(q) = U_{att}(q) + U_{rep}(q) \quad (4.3)$$

where,  $U_{att}(q)$  is attracting (goal) and  $U_{rep}(q)$  is repulsing (obstacle) fields.

Gradient of the potential function must be differentiable:

$$\nabla U = \begin{bmatrix} \frac{\partial U}{\partial x} \\ \frac{\partial U}{\partial y} \end{bmatrix} \quad (4.4)$$

From above equations, artificial force field acting on a robot  $F(q)$ , as the gradient of the potential field is given by:

$$F(q) = -\nabla U(q)$$

$$F(q) = F_{att}(q) + F_{rep}(q) = -\nabla U_{att}(q) - \nabla U_{rep}(q) \quad (4.5)$$

Converting to robot control, we set the robot velocity  $(v_x, v_y)$ , proportional to the force  $F(q)$  generated by the field, the force field drives the robot to the goal and the robot model is derived in section 3.

Functions of attractive potential field will be: linear function of distance, quadratic function of distance and combination of the both. Linear function of distance is:

$$U_{att}(q) = \xi_{att} \|q - q_{goal}\|,$$

$$F_{att}(q) = -\nabla U_{att}(q) = -\xi_{att} \frac{(q - q_{goal})}{\|q - q_{goal}\|} \quad (4.6)$$

where  $\xi_{att}$  is a positive scaling factor, and  $q - q_{goal}$  is the distance.

- Quadratic function of distance is given by:

$$U_{att}(q) = \xi_{att} \frac{1}{2} \|q - q_{goal}\|^2 \quad (4.7)$$

Attracting force converges linearly towards 0 (goal):

$$F_{att}(q) = -\nabla U_{att}(q) = -\xi_{att} \|q - q_{goal}\| \frac{(q - q_{goal})}{\|q - q_{goal}\|} = -\xi_{att} (q - q_{goal}) \quad (4.8)$$

Repulsive Potential Field should generate a barrier around all the obstacles: strong, if close to the obstacle, or no influence - if far from the obstacle:

$$U_{rep}(q) = \begin{cases} \frac{1}{2} \xi_{rep} \left( \frac{1}{\rho(q)} - \frac{1}{\rho_0} \right) & \text{if } \rho(q) \leq \rho_0 \\ 0 & \text{if } \rho(q) \geq \rho_0 \end{cases} \quad (4.9)$$

where  $\rho(q)$  is the minimal distance to the obstacle from  $q$ ;  $\rho_0$  is distance of influence of obstacle.

Field is positive or zero and tends to infinity, as  $q$  gets closer to the obstacle:

$$F_{rep}(q) = -\nabla U_{rep}(q) = \begin{cases} \xi_{rep} \left( \frac{1}{\rho(q)} - \frac{1}{\rho_0} \right) \frac{1}{\rho^2(q)} \frac{q - q_{obstacle}}{\rho(q)} & \text{if } \rho(q) \leq \rho_0 \\ 0 & \text{if } \rho(q) \geq \rho_0 \end{cases} \quad (4.10)$$

For potential field path planning we have used harmonic potentials, since the robot is moving in an environment with fixed obstacles, ensuring that there are no local minima.

#### V. SIMULATION RESULTS

In accordance with block diagram in Fig.3.1 and equations that describe each module we developed the simulation model of the mobile robot in Matlab/Simulink.

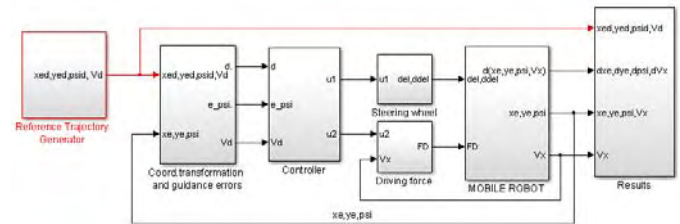


Fig. 5.1. Mobile robot control system model - developed in Simulink

Referent trajectory in an obstacle environment is generated in the block Reference Trajectory Generator using potential field method. Fig. 5.2 shows how the mobile robot follows this referent trajectory. Obstacles in Fig.5.2 are presented by circles.

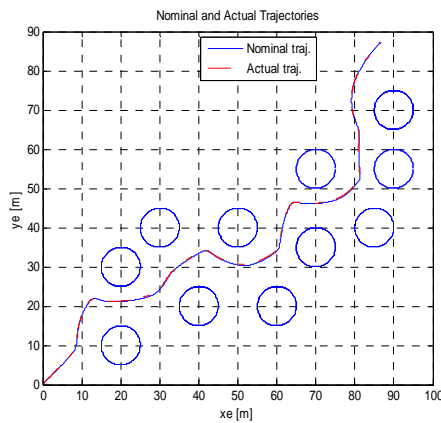
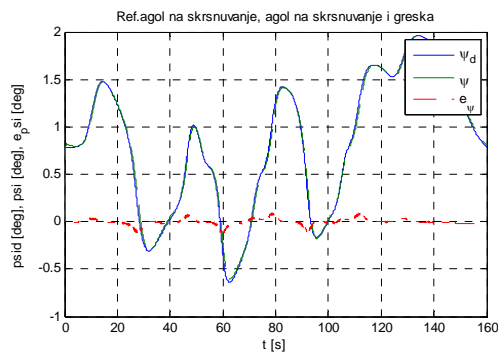


Fig. 5.2. Referent trajectory following



$\psi_d$  - nominal angle,  $\psi$  - real angle,  $e_{\psi}$  tracking error

Fig. 5.3. Robot course angle following during the movement along referent trajectory

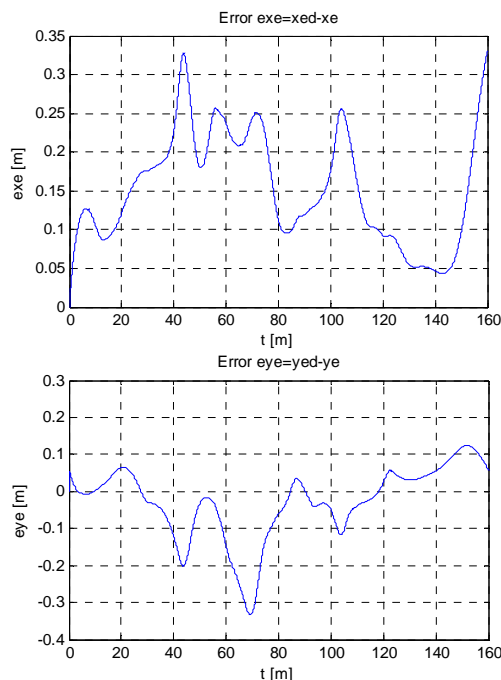


Fig. 5.4. a) Tracking error along x axis, b) Tracking error along y ax.

## VI. CONCLUSION

In this paper we have researched and simulated the mobile robot guidance and control in the environment full of obstacles, by using the potential field's method. The mobile

robot has 4-wheels configuration, electric drive on the rear vehicles, and is directed from the front wheels (Ackerman control algorithm). We have simulated a movement in a horizontal (2D) plane and the robot is modeled as a 3-DOF system (three degrees of freedom).

Our method uses functions defining potential fields at its position to calculate component vector. Only the portion of field affecting robot was computed for each behavior of the potential field; the sum of the vectors at the robot's position gave the resultant output vector. We did not have issues with combining potential fields; the impact of update rates is that lower update rates can lead to "jagged" paths. The robot was treated as a mass object, it could not to change velocity and direction instantaneously (cannot happen). We found a solution for local minimum problem; if the global minimum is not guaranteed, we have chosen the functions in such a way that global minimum can be guaranteed and the robot will escape the local minima. "Avoid-past" behaviors were included – remembering where robot has been and bringing the robot to other places. They gave "avoid" behaviour a preferred direction and inserted tangential fields around obstacles.

## REFERENCES

- [1] Rajesh Rajamani, Vehicle Dynamics and Control, Second Edition, *Springer - New York*, Dordrecht, Heidelberg, London, 2011.
- [2] Spyros G. Tzafestas, Introduction to Mobile Robot Control, *Elsevier*, 32 Jamestown Road, London NW1 7BY, 2014.
- [3] Eric N Moret, Dynamic Modeling and Control of a Car-Like Robot, *Master Thesis, Virginia Polytechnic Institute and State University*, 2003.
- [4] K. Kozłowski, D. Pazderski, I. Rudas, J. Tar, Modeling and control of a 4-wheel skid-steering mobile robot: From theory to practice, *Polish-Hungarian Bilateral Technology Co-operation Project and statutory grant No. DS 93/121/04 of Poznan University of Technology*.
- [5] Shimoda S., Kuroda Y., Iagnemma K., (2005), Potential field navigation of high speed unmanned ground vehicles on uneven terrain, *IEEE Int. Conf. on Robotics and Automation*, Barcelona, Spain, pp.2828-2833.
- [6] Khatib O. (1990), Real-time obstacle avoidance for manipulators and mobile robots, In *Autonomous Robot Vehicles*, (Edited by I.J. Cox and G.T. Wilfong), pp. 396-404, Springer-Verlag.
- [7] Rimon Elon, Koditschek Daniel (1992). Exact Robot Navigation using Artificial Potential Functions. *IEEE Trans. on Robotics and Automation*, pp. 501-517, October (1992).
- [8] Kun-Hsiang Wu, Chin-Hsing Chen, and Juing-Ming Ko, Path planning and prototype design of an AGV. *Mathematical and Computer Modelling*, 30, 147-167 (1999).
- [9] Borenstein J. and Koren Y. (1991), The vector field histogram-fast obstacle avoidance for mobile robots, *IEEE Transactions on Robotics and Automation* 7, 278-288.



# Method and device for registration and analysis of car crashes at smart traffic lights. Hardware model.

Rosen Vitanov<sup>1</sup> and assoc. prof. Georgi Stanchev<sup>2</sup>

**Abstract** – In this paper hardware model of a smart traffic light system that prevents and registers car incidents at traffic intersections. In addition it may ease any car crash investigation by providing data on how and when it happened, revealing the one holding responsibility for it.

**Keywords** – *Smart traffic light, Car crash, Traffic intersection, Car crash investigation.*

## I. INTRODUCTION

Undeniable is the contribution of cars in the history of mankind. From the 1885/1886 year, when Karl Friedrich Benz invented the first automobile [1], development of this industry is unsettling, closing the roads and making the world lesser and smaller. Individual becomes able to travel great distances, more often and cheaper.

In the city of Ohio, in 1891, the first car crash was registered. James Lambert crashed into a tree [2]. Since then the number of road accidents is growing rapidly. According to statistics for 2007, Bulgaria is ranked 12 in number of deaths per 1,000 people - 131.7 [3]. For 2013 became 6759 total accidents with 577 killed and 8491 wounded, 557 were accidents at traffic light intersection. For 2014 these statistics area 6736 road accidents, 615 deaths, 8358 wounded, 581 at road junctions [4].

## II. SMART TRAFFIC LIGHT SYSTEM

### A. Overview

The proposed system is an additional method for preventing and resolving disputes arising at road intersection due to car accidents. Its purpose is to provide information about the configuration of traffic light signals emitted before, during and after the crash. Thus easing and speeding investigating authorities to resolve a traffic accident. Modularity is the basis of the design. Proposed system may be used with other traffic monitoring systems or can be mounted on traffic lights as a standalone device.

Smart traffic lights emit a pilot signal with the current state of lights. This information is received by all cars at the intersection. Based on previous coordinates and current

status, the vehicle detects its direction of movement and filters required data. Upon collision of two vehicles, the system generates a distress call for an accident and the smart traffic light stores the state of light signals at the current time and up to 30 seconds back. This is sent to all cars, involved in the crash. The other participants are signaled to pass with caution.

### B. Block-Diagram

A block-diagram of the proposed system is shown on Fig.1. It is composed of three stand-alone devices that may exchange data using wireless radio communication – a smart traffic light control system, a reporter and a middleman.

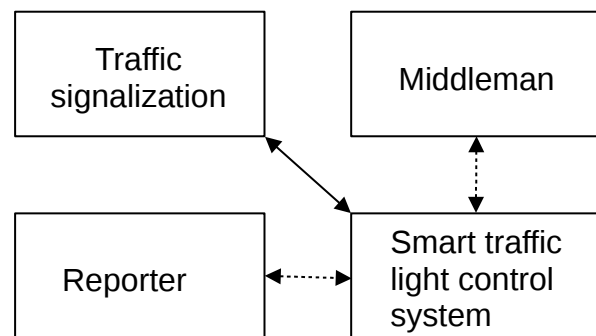


Fig. 1. Block diagram of traffic light accident control and resolve system.

### C. Smart traffic light control system

Smart traffic light control system is a system that supervises and stores in its memory information about signals that were shown on the traffic signalization. Each intersection has its own supervising system that is constantly active to retain consistent data. It sends on a pilot channel data, related to the current state of the intersection to all traffic participants. Supervisor may not detect a incident by itself. It relays to a different device called *Reporter* to acknowledge an emergency. In case of a collision, a record will be stored on internal non-volatile memory with the current state, up to 30 seconds of signal history, up to 5 seconds of proceeding signals, including day and time of the accident, with additional it may report to all vehicles at the intersection to be very cautious when passing. This new data will be sent to all participants in the accident. Records may be recalled by authorities at any time when needed.

As shown in Fig.2, smart traffic light control system consists a total of 7 modules – power supply, real-time clock, non-volatile memory, wireless communication module, traffic light signalization interface, communication interface and microprocessor.

<sup>1</sup>Rosen Vitanov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: rosenvitanov@gmail.com

<sup>2</sup>assoc. prof. Georgi Stanchev is with the Faculty of Mechanical Engineering at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria. E-mail: gstanchev@tu-sofia.bg

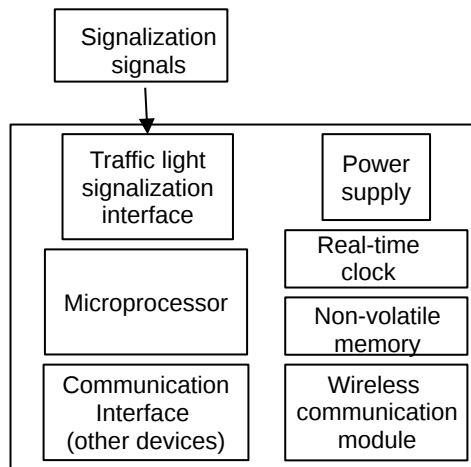


Fig. 2. Block diagram of Supervisor

*Wireless communication module* provides connection between all three modules. In this case, is used MRF24J40 by Microchip, working ant 2,4GHz. *Real time clock* is used to keep up date/time. *Traffic light signalization interface* is connection between the traffic light itself and the control unit. It parses received signals from the traffic signalization to usable information. *Non-volatile memory* is used to store records for every car crash, including systems settings and metadata. A generic SecureDigital card may be used as a mass-storage device.

*Microprocessor* is hearth of the system. AtMega328p is used. With current hardware implementation and selected components can be tracked up to six three-segment traffic lights or 9 two-segment. When selecting a suitable controller such as MEGA2560, the functionality and traffic light count may be increased several times.

#### D. Reporter

*Reporter* is a extension to HeERO project, that reports an accident to the traffic light control system. A block diagram is presented on fig. 3.

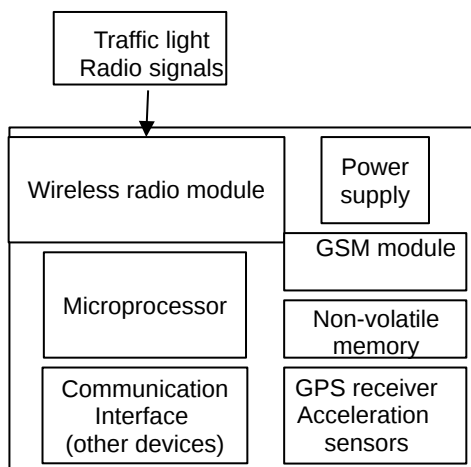


Fig. 3. Block diagram of Reporter

This device is situated in vehicles and constantly monitors for any pilot signals from traffic intersections. If any detected, it will be stored into memory.

It consists of seven different modules. Power supply is self explanatory, but should be kept in mind, that the source of power is limited to a vehicle's battery, thus it should be designed with emphasis to power saving.

- A Wireless radio module is to setup and a connection with traffic light control system.
- A GSM module is responsible to retain a connection to a cellular network, and in case of an emergency to make an e-call to the headquarters that are responsible to take care of the accident.
- Non volatile memory could be a traditional microSD card or a flash memory chip. It is used to store the collected information from sensors (from the HeERO project), the traffic light signals and records of any abnormal situations.

#### E. Middleman

Middleman, also known as mediator, is a device that connects wirelessly to "Supervisor" and "Reporter". Its purpose is to read the information for the intersection - name of intersection, traffic lights status, car crash records archive, date and time and condition the signal system at each received signal. It may be a separate device, computer terminal or phone with Android based OS . When selecting a traffic accident on the device's screen to visualize the status of the traffic lights in the respective colors, the exact time of the accident. It also may be used to control and operate the traffic light – enter new settings, date and time, delete data, format memory, etc.

In addition, if *Reporter* is still online after the car crash, it may connect to and download from it all available data that was stored and is unreachable for the control system. This data may be stored trajectory, differential time between vehicle goes and incident.

#### F. Prototype model

On the picture shown (fig. 4a & 4b ) is a top and bottom view of a prototype model of presented model of a smart traffic light system for monitoring, prevention and reporting vehicle collisions.

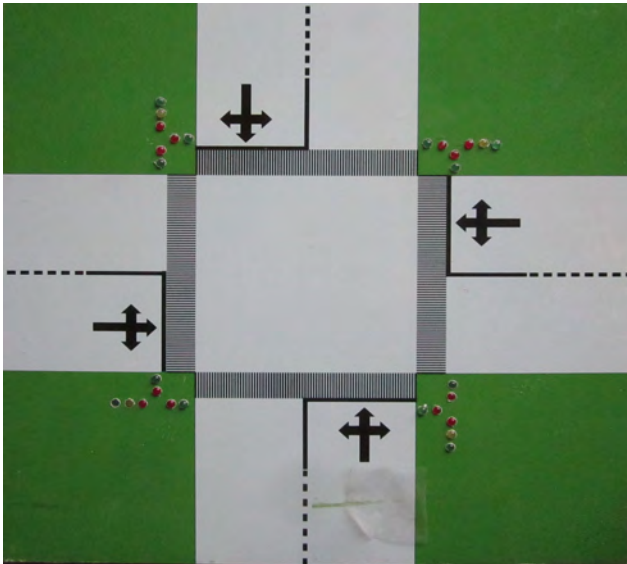


Fig. 4a. Prototype model – top view.

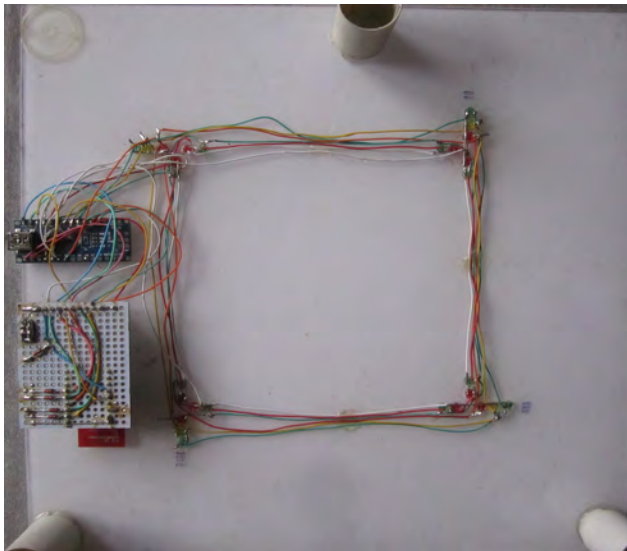


Fig. 4b. Prototype model – bottom view.

### III. CONCLUSION

In this paper presentation of hardware model of a smart traffic light collision detection and prevention system, that may help in investigation of a car crash.

### REFERENCES

- [1] Library of Congress, "Everyday Mysteries", <https://www.loc.gov/rr/scitech/mysteries/auto.html>.
- [2] Ohio History Central, "World's First Automobile Accident". 2006. <http://www.ohiohistorycentral.org/entry.php?rec=2596>.
- [3] State-Public Consultative Commission on the Problems of traffic safety on the roads, "World Statistics", 2015.
- [4] State-Public Consultative Commission on the Problems of traffic safety on the roads, "Statistics driver's fault", 2015.
- [5] Ordinance № 17 of 07.23.2001 "Regulation of traffic signals with lights", 2001.

Presented is an early alpha version of the prototype with many of the features still in development or may not provide reliability. For testing purposes in-vehicle device is has only touch-sensitive switch and is not fully functional.

As a result the reporter can detect when it is in the premises of a road junction. In case of a crash it may send adistress call and receive a response from the traffic control system.

This Page Intentionally Left Blank



# Method and device for registration and analysis of car crashes at smart traffic lights. Software model.

Rosen Vitanov<sup>1</sup> and prof. Rumen Arnaudov<sup>2</sup>

**Abstract** – In this paper software and communication model of a smart traffic light system that prevents and registers car incidents at traffic intersections.

**Keywords** – Prevention, smart traffic light, car crash, traffic control system.

## I. INTRODUCTION

Following is described a software model for a collision detection and prevention smart traffic light. Proposed system is based on currently developing eCall system, such as “Harmonized eCall European” (HeERO) and is an add on to provided services.

As stated an ecall “Using timely signalization and precise GPS positioning the system allows for faster and more adequate rescue force reaction, thus increasing the chance to save more people” [1].

The main purpose of such device is to serve as a black box for a vehicle. As such, it records a minimum data set (MDS) throughout the travel. This data is set, but not limited, to vehicle ID, vehicle energy storage type, time stamp, vehicle location, vehicle direction, number of passengers and optional data [3]. In case of an incident an ecall will be automatically setup and MDS chunk will be sent to emergency center.

When and how to recognize a transport accident is defined by Veronica 2 standard (EN 15722). It uses sets of parameters that are constantly monitored and analyzed to recognize an abnormal situation. According to Veronica II incident is any unwanted or unintentional event or a specific sequence of such events which have harmful consequences. That is, any event that results in damage is “unacceptable” and a record for such an event should be stored in “black boxes”. [4]. Before rising a trigger for an unwanted event, statistics of two or more parameters are in consideration – longitudinal acceleration, lateral acceleration, vehicle speed, vehicle roll angle, engine throttle, engine speed, ABS activity, steering wheel angle, brake status, frontal airbag status, safety belt status, etc [2].

Thus, using the existing HeERO device as a platform, a new functionality is added, relaying on already proven concepts. In this way proposed system contributes additional help to clarification process of an accident.

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

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

## II. SOFTWARE PRINCIPLES

### A. Control unit – record structure

On fig. 1 is shown what a data structure to store information, related to a traffic light on an intersection looks like.

#### Smart Junction structure

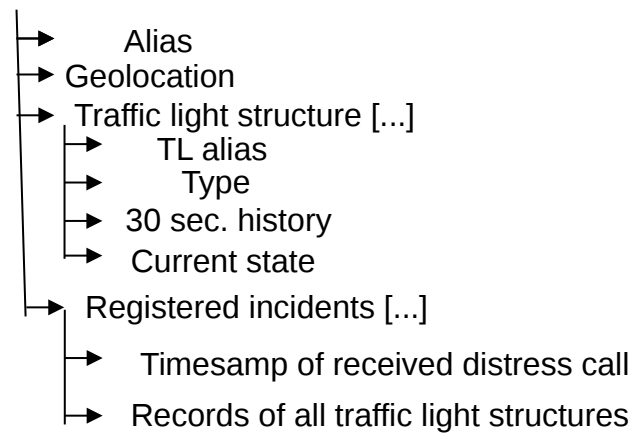


Fig. 1. Showing data structure of a smart traffic light system.

A road junction may have a name, that is stored in “Alias”. It is used to ease human operators. “Geolocation” stores coordinates of a traffic light, so the eCall operator knows the exact place of a car crash. These values, once set, are constant through the whole lifecycle of the system.

As known, an intersection consists of more than one traffic signalization devices (traffic lights). “Traffic lights tructure” is holds individual information for each and every traffic light. As many are the traffic signaling as many copies are hold. “TL alias” represents name of particular signaling device (eg, “North/South, East/West). “Current state “ stores current state of the traffic light signals. The heart beat of a traffic light is one 1 Hz. That is why its current state is sampled once every second. Before that, the oldest record in history is deleted, to make room for the already “aged” current state, before resampling.

Detection and providing information about signals is the prime aim of the proposed device. That is why it should keep track of all abnormal situations that occurred. When such situation occurs, a copy of the tracked history is stored along with a time stamp of the incident. Memory that is spared would be 20 bytes. For every second a roughly 3 bit of data are spared (1).

$$Mem = 3 \text{ bits/s} * 31 \text{ secs} + meta = 93 + 67 = 160 \text{ bits}$$

### B. Control unit – control software

Main purpose of the software is to keep track of the traffic light signals and look for distress signals. Block diagram is shown on figure 2.

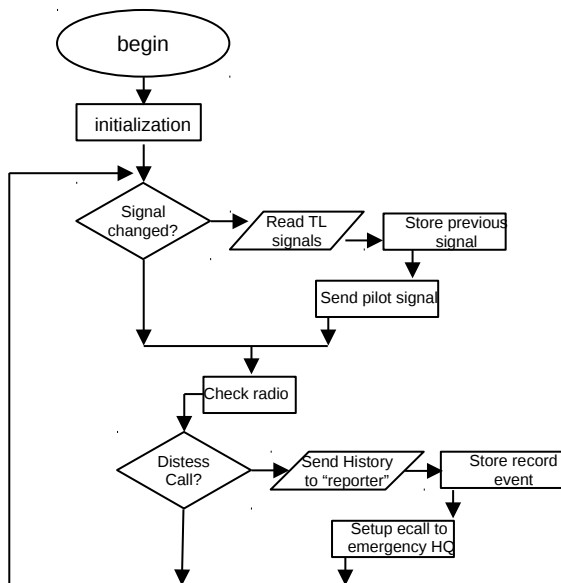


Fig. 2. Block diagram of smart traffic light software.

In the beginning an initialization procedure is hold. When a control unit is started alias and location data should be filled along with associating it to all traffic lights at the road junction. After that the system will start to monitor all traffic signalizations and keep 30 seconds history. In the meantime the system would broadcast a pilot signal, including timestamps and colors of the traffic lights – announcing all vehicles that the are in the borders of a street junction. Once per second colors are reread and stored in controller's memory. The old ones are shifted to make place for the newly read one thus deleting the oldest. After that the pilot signal is resend.

Crucial moment is to constantly monitor for any distress calls send from vehicles in case of an emergency. When such event is detected, a new record with time stamp of the received distress call, current and history states of all traffic signalizations are stored into non-volatile memory.

After a that a the same record is sent to the reporter that notified the system for the car crash. In the meantime an ecall is setup. This is needed in case the on-board system is unable to notify emergency centers for the incident. It contains all data related to the traffic intersection – name, geolocational data, timestamps, etc. These records are retrievable at any time remotely from the eCall headquarter, police or investigators.

An additional feature is the ability to “replay” the situation on the junctions signalization devices themselves, thus providing better visual knowleadge and further understanding of the occurred situation.

### C. Reporter – control software

On figure 3 is shown a functional block diagram of the software of device Reporter. It's main aim is to acknowledge when a vehicle is in the limits of a traffic intersection by receiving a pilot signal from the traffic light control system.

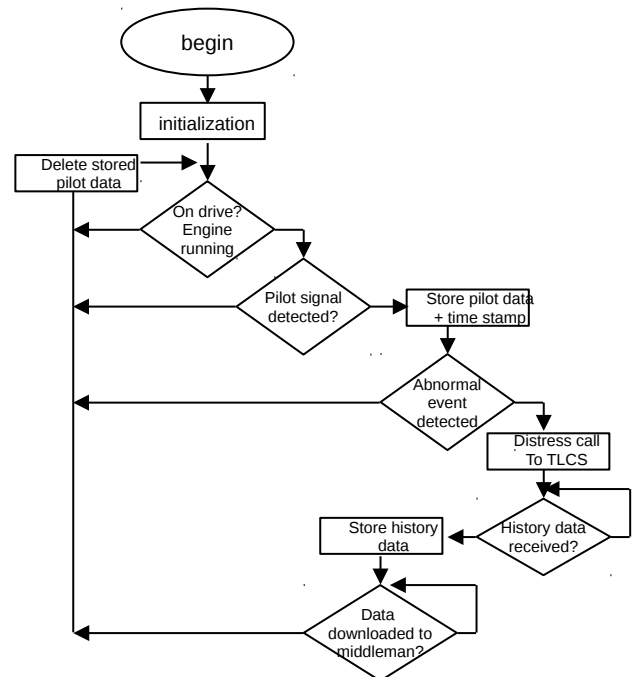


Fig. 3. Block diagram of Reporter device

Device only works when the vehicle is in use - by checking number of parameters – one of which is if the engine is working.

If the vehicle is in use (on drive), device constantly monitor for any pilot signals sent by a junction. Pilot signal carries information about the traffic signals and timestamps that are crucial for synchronization with internal monitored parameters by the HeERO device.

When a pilot signal is detected, *reporter* understands that it is within the limits of a junction and starts to record all incoming traffic light data. When the vehicle is no longer in junction's limit (junction is passed without any trouble), this information becomes useless and so deleted.

If an abnormal condition is detected by the HeERO device *reporter* is acknowledged and a distress call is sent to smart traffic light control system and waits for a response, containing all traffic light history. After it is received, it is stored into non-volatile memory along with all received pilot signals and timestamps.

Meanwhile and ecall is setup by the onboard HeERO device. *Reporter* will wait for a connection and stored data to be downloaded by a *mediator* device, in possession of reinforcement law.

#### D. Communication protocol

On figure 4 is shown a timeline of a typical communication between *traffic light control system* and *reporter*.

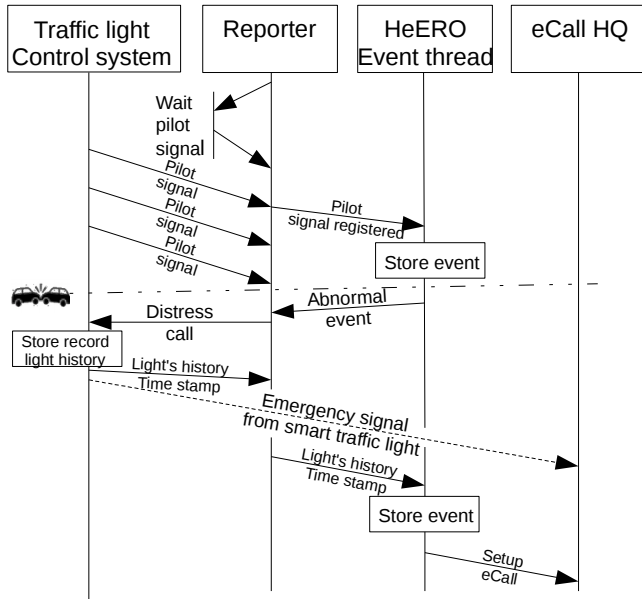


Fig. 4. Communication process between *traffic light control system* and *reporter* between *traffic light control system* and *reporter*.

As the figure is self-explanatory some remarks should be made. On the figure above, HeEro black box is responsible to store a record for the event occurred., in contrast explained in the upper chapter. Presented software model presented is aimed as to extend the functionality of a intelligent in-vehicle transport system. Thus, “who” willstore the event is disputable.

#### III. ADDITIONAL REMARKS

Source code of the project could be find at GitHub:

<https://github.com/rosenvitanov/smfTLeCall.git>

#### IV. CONCLUSION

Software and prototype model for a smart traffic light system, that enriches the data collected from an e-call black box, and clarifies the circumstances around a occurred road accident. This new method is in help for the investigation committee to have a deeper and richer overview of the collision.

As a future development new features shall be added that are in help to the driver itself - informing if it is safe to enter the road junction or will a fault be made if done so.

#### REFERENCES

- [1] M. Aldimirov, G. Stanchev, Rumen Arnaudov, “Integradet system for car park management and Ecall road accident signalization”, IJEIT International Journal of Engineering and Innovative Technology vol. 4, issue 10, 2015.
- [2] Continental AG. Veronica 2 project. European
- [3] Commission website. [Online] Oct 2009. [ec.europa.eu/transport/road\\_safety/pdf/projects/veronicaii.pdf](http://ec.europa.eu/transport/road_safety/pdf/projects/veronicaii.pdf)
- [4] Technical committee CEN/TC 278.EN15722:
- [5] Intelligent transport systems - eSafety - eCall minimum set of data (MSD). 2011. ISBN 978 0 580 69997 9
- [6] M. Aldimirov, G. Stanchev, “Analysis andproposal for improvements of the devices described in the IEEE1616 standart and ecall system”, Bulgarian journal for engineering design, issue 10, 2014.

This Page Intentionally Left Blank



# Control-Command and Signalling Subsystems - Testing Framework within the APIS Process

Denitsa Kireva-Mihova<sup>1</sup>

**Abstract** – This article focuses on the testing framework of the CCS SS that are part of the Railway System, within the APIS process. Testing is part of verification and validation to support product development, verification and validation of technical compatibility and safe integration. Based on a study implementation the highlights of the testing within the APIS legal framework are pointed out and issues are discussed and analysed.

**Keywords** – Railway system, CCS, TSI, Authorisation for Placing in Service

## I. INTRODUCTION

The Interoperability Directive 2008/57/EC [1] sets the legal framework for the authorisation for placing in service (APIS) of the railway structural subsystems and vehicles. It also introduces the technical specifications for interoperability (TSI). The TSIs specify the essential requirements for each subsystem and the functional and technical specifications to be met by these subsystems and their interfaces. According to Directive 2008/57/EC article 15, “Member States (MSs) shall take all appropriate steps to ensure that these subsystems may be placed in service only if they are designed, constructed and installed in such a way as to meet the essential requirements concerning them when integrated into the rail system”. In particular, the MS has to check technical compatibility and safe integration before subsystems may be placed in service. The Commission Implementing Regulation (EU) No 402/2013 [8] on the common safety method for risk evaluation and assessment (CSM-RA) describes the risk management process to be implemented in case of any change to the railway system. The Commission Recommendation 2014/897/EU [5] clarifies the procedure for APIS of structural subsystems and vehicles as set out in [1].

APIS is one of the activities supporting the development of an integrated, safe and interoperable railway system in the European Union (EU), particularly in the field of cross-acceptance of railway subsystems, i.e. mutual recognition of authorisations for the placing in service of railway vehicles.

Testing is part of verification and validation to support product development, subsystem integration, subsystem verification, and validation of technical compatibility and safe integration.

Currently many issues arise with the different CCS subsystem (CCS SS) definitions and the relevant scope of the final testing phase under the legal framework in force.

## II. CONTROL-COMMAND AND SIGNALLING SUBSYSTEM DEFINITION

Within the European legal framework for interoperability the Control-command and Signalling subsystem is one of the structural subsystems specified in [1] and is divided into track-side (TS) and on-board (OB) parts. In Annex II to [1] the CCS SS is defined as ‘all the equipment required to ensure safety and to command and control movements of trains authorised to travel on the network’. The features of the CCS subsystems are [3]: the functions that are essential for the safe control of railway traffic, and that are essential for its operation, including those required for degraded modes; the interfaces; the level of performance required to meet the essential requirements. The CCS TSI specifies only those requirements which are necessary to assure the interoperability of the EU rail system and compliance with the essential requirements.

The CCS TSI [3] specifies the following TS-CCS parts: *trackside train protection (class A)*; *trackside radio communication (class A)*, *train detection (as interface requirements, to ensure compatibility with rolling stock)*, and OB-CCS parts: *on-board train protection (class A)* and *on-board radio communication (class A)*. The Class A train protection system is ERTMS/ETCS whilst the Class A radio system is GSM-R. For Class A train detection this TSI specifies only the requirements for the interfaces with the other subsystems [3]. The scope of TSI CCS subsystem parts is shown on Figure 1 below.

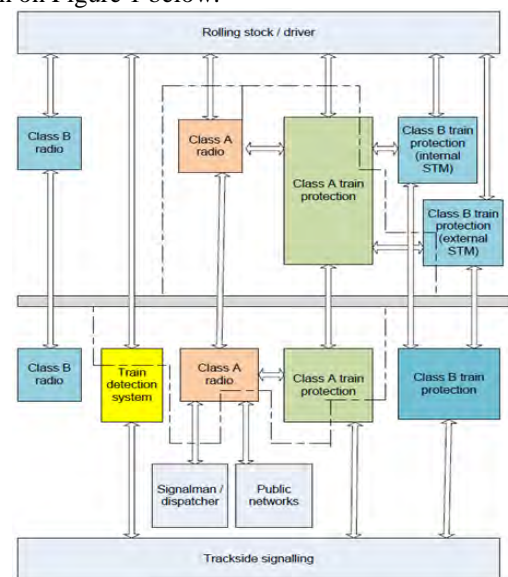


Fig. 1. Scope of TSI Control-Command and Signalling [4]

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

### III. APIS PROCESS AND TESTING

#### Authorisation for placing in service process

Commission Recommendation 2014/897/EU [5] defines the APIS of a subsystem as ‘the recognition by the Member State that the applicant for this subsystem has demonstrated that it meets, in its design operating state, all the essential requirements of Directive 2008/57/EC when integrated into the rail system’.

In principal Directive 2008/57/EC [1] regulates the technical characteristics of the railway subsystems and vehicles and the process of their authorisation for placing in service and Directive 2004/49/EC [2] regulates the processes of their use, operation and maintenance after the authorisation. After a subsystem is placed in service, care should be taken to ensure that it is operated and maintained in accordance with the essential requirements relating to it. Under [2], responsibility for meeting these requirements lies, for their respective subsystems, with the infrastructure manager (IM) or the railway undertaking (RU). The National Safety Authority (NSA) acts, on behalf of the MS to grant the APIS. Some NSAs also grant a temporary authorisation or authorisation for operation on specific lines, with restrictions.

#### Testing process

The goal of the testing process within an APIS could be distinguished in between interoperability requirements for obtaining evidences for operation under full operational conditions and the evidences that ensure the technical compatibility and the safe integration of the CCS subsystems.

According to the CCS TSI the main target is “A certified train that can run on certified lines with only a few additional checks and tests (corresponding class B systems)”. To achieve such target situation the following parameters shall be checked in accordance with [1]: the technical compatibility of these subsystems with the system into which they are to be integrated, and the safe integration of these subsystems in accordance with Articles 4(3) and 6(3) of [2].

Testing process of the CCS SS under the different definitions scope is analysed below and presented on Figure 2.

#### Testing process within an APIS for CCS SS

In general, according to [5], the only tests that may be required for authorization, which have to be performed before the authorization for placing in service and which requires the involvement of an assessment body, should be the tests which are [5]:

- explicitly specified in the TSIs, modules, and, where relevant, in national rules;
- defined by the applicant for demonstrating the compliance with the requirements of the TSIs and/or national rules;
- defined in other EU legislation, or
- defined by the applicant, in accordance with the application of CSM RA (i.e. safe integration

between the elements composing a subsystem and/or safe integration between subsystems that constitute a vehicle or a network project).

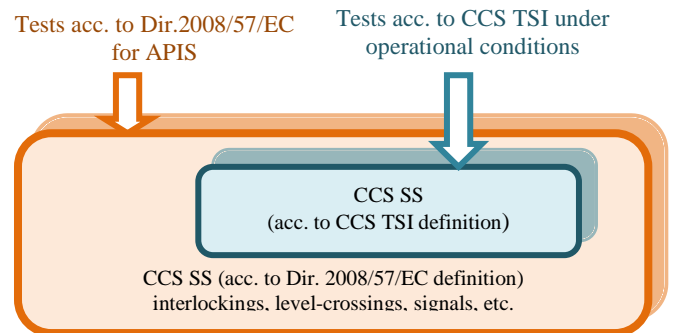


Fig. 2. Scope of the CCS subsystems tests

Table I summarises the participants and their roles within the testing process, according to the Commission Recommendation 2014/897/EU [5].

TABLE I  
ROLES AND TESTING

Roles (acc. to [5])		Applicant	NoBo	DeBo	AsBo
Explicitly specified in the TSI		X	X	NA	NA
Modules		X	X	NA	NA
National rules (where relevant)		X	NA	X	X
Defined by the applicant	Demonstrating compliance with TSI/NR	X	X	X	NA
	CSM-RA application	X	X see Table 6.2 and 6.3 [3]	NA	X
Defined in other EU legislation		X	X see [1]	NA	NA

#### Testing and Notified Bodies

Depending on the modules used in accordance with [6], testing responsibility rests with the different parties involved in the EC verification and assessment procedure. Table II below presents the testing execution responsibilities.

TABLE II  
TESTING AND NOTIFIED BODIES

Module	Responsibility		
	Applicant	NoBo	Other Body
SB+SD	NO	X	NO
SB+SF	NO	X	NO
SG	X (if stated in the TSI but not valid for the CCS TSI)	X (as overall responsibility)	X
SH1	X	NO	X (if on behalf of the applicant)

Using SB+SF or SB+SD modules the tests fall under the Notified Body (NoBo) responsibilities with the related execution. Using SG module the tests results fall under the NoBo responsibilities but the execution is also possible by other bodies (different from the NoBo) under comparable conditions (previously recognised by the NoBo assessing their accreditations/recognitions). Using SH1 module the responsibilities of the test results as well as the execution fall under the applicant. Tests can be also carried out by other appointed bodies on behalf of the applicant. The NoBo does not need to perform any special recognition. A NoBo assesses the compliance with the relevant TSI requirements using appropriate testing through the QMS assessment.

#### *Some inputs*

- ✓The selection of the certification module is crucial in order to define roles and responsibilities of testing;
- ✓When a module including testing is selected it is beneficial to agree a test plan between the applicant and the NoBo including the definition of roles and responsibilities;
- ✓A major involvement of the Applicant in a testing phase (who knows really what to test);
- ✓The amount of test cases could be increased but the operational scenarios should be the real one;
- ✓Sometimes a NoBo has not in-house testing capabilities and external bodies have to be hired causing a time and cost extension and such entity(ies) need to be recognised by the NoBo in accordance with its internal procedure or the accreditation standards.
- ✓The separation of tests in and out the scope of an APIS should be clear (e.g. tests needed by RU to establish train-route compatibility before using a vehicle type or new subsystem on a particular route).

#### **IV. RESULTS FROM THE ANALYSIS**

The main findings made out of the analysis of the testing framework within the APIS process could be summarised, as follows:

- ✓The scope of the CCS SS definition is different within the legal interoperability documents that specify both trackside and on-board CCS subsystems, namely:
  - The definition of CCS subsystems given in the CCS TSI covers the **basic parameters** in relation with such TSI (i.e. class A systems and their interfaces);
  - The definition of CCS subsystems given in Annex II of the [1] covers **an extended scope** (including in principle class B systems and other safety-related systems – interlockings, level-crossing systems, block systems, signals, etc.);
- ✓In general APIS covers a subsystem and therefore the definition of Annex II of the Directive 2008/57/EC

should be the reference definition. Furthermore, an APIS requires **only the essential requirements** in [1] to be met through the use of TSIs.

- ✓In practice when the CCS TSI refers to final tests under full operation conditions it addresses the **basic parameters** within the TSI.

#### **Critical issues:**

- An APIS is granted to a structural subsystem having a wider scope (including all safety-related equipment) as intended in [1];
- CCS TSI does not cover the whole scope of CCS SS but the interoperability aspects only;
- A question arises: Should the parts not covered by the CCS TSI be tested within the APIS process?

The Commission Recommendation 2014/897/EU [5] says “NO” but this is not the current practice in the Member States.

- ✓The use of conditional statement in [5] like *may be* for “*the only tests that may be required for authorization*” provides a certain degree of freedom;
- ✓The conformity assessment modules [6] addressed only the responsibility for making required tests but what kind of tests is not specified;
- ✓The requirement for tests [5] to be executed under national rules, where relevant, is too generic and leaves ambiguities (i.e. national technical rules or national safety rules);
- ✓Railway specific tests defined in other EU legislation and the responsible party for them are difficult to be controlled due to the wide-range scope;
- ✓The requirement for tests defined by the Applicant also provides a certain degree of freedom depending on the real aim of the Applicant - to verify only the TSI requirements, the national requirements where relevant or something else;

#### **Critical issues:**

- Some actors are missing in the testing framework within the APIS: An Assessment Body (e.g. the NoBo according to the duties and responsibilities given through the selected conformity assessment module), IM or RU, NSA. However, in general, such entities might be part of a testing process for APIS.
- In case of an APIS process includes also other actors how could the CCS parts not subject to interoperability be tested? Commission Recommendation 2014/897/EU excludes them but these parts can be considered in the CSM–RA application.

#### **V. EVOLUTION OF THE TESTING FRAMEWORK WITHIN THE APIS PROCESS**

In our days, evidence showing that the relevant operational situations are functional can only be given for the respective network operational situations tested. Such system integration operational tests are required due to the following issues: a) Specifications not fully validated; b) Test cases and test environment not fully validated; c) Certificates of the

subsystems with restrictions and conditions; d) Track-train system integration not fully validated; e) Different principles for CCS implementation caused by freedom of engineering applying the CCS specifications; f) Insufficient experience with harmonised transitions from one level to the other (e.g. ETCS L1 to L2)

The applicant for an APIS of a CCS SS has to prove integration of his subsystem within the railway system and for each network where it is intended to operate. To ease this process the technical development should allow transferring more tests into laboratories. However, execution of exhaustive on-site tests will be only a transitory situation. Having the growth of experience, stability and validation of the specifications and products, the amount of validation tests for system integration could be stepwise reduced to a minimum, as shown on Figure 3.

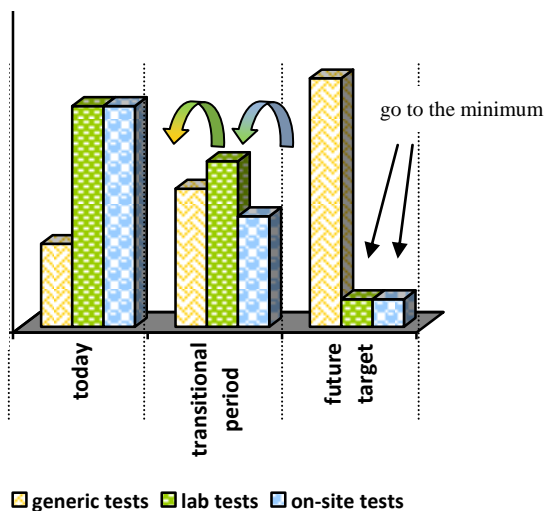


Fig. 3. Reduction of the CCS SS project specific lab and on-site testing

Some options are given here for the applicant to demonstrate that some tests have become no longer necessary.

- Validation test for system integration applied to an on-board SS CCS is proven to be fully covered by mandatory conformity tests (Subset-076);
- Lab tests can be taken into account if they have been performed in a way that ensures the same system behaviour on-site;
- CCS on-site tests can re-use the manufacturer's field tests during their CCS subsystem integration tests, if those cover the system integration operational test cases;
- Track-train system verification tests of a vehicle related to a specific route or network can be reduced by those tests successfully passed on other routes or networks if the conditions of the other routes or networks (engineering rules, operational scenarios) ensure the same system behaviour (equivalent test conditions).

The full scope of system integration tests will be tested only in the first projects within one network project. In the target

situation, only conformity tests and a small set of final route or network specific tests will be necessary to re-authorise a vehicle or CCS SS for a network. Track-train system verification testing could finally be reduced to a few site tests for verification of TSI open points and some daily applied operational scenarios.

## VI. CONCLUSION

In our days, the CCS SS testing framework call forth fuzzy and incomplete system integration and validation process. This is due to variety of subsystem definitions, certain degree of freedom for the required tests for APIS, not fully validated specifications, lack of national rules, etc. Deep analysis with interesting highlights of the legal framework, processes and participants within the APIS process is made in this paper to produce a proposal for improvement of the testing framework. The need for transitional period is discussed to consolidate the experience gained during the CCS SS testing with the purpose to reach a stable target situation in the next years.

## ACKNOWLEDGEMENT

The analysis is part of the "ERTMS Harmonised and International Procedures for Placing into Operation of Products and Subsystems" project, number 2014-EU-TM-0128-S, which is co-funded under CEF 2014-2020 and is contributed by Rina Services.

## REFERENCES

- [1] Directive 2008/57/EC and further amendments (the "Railway Interoperability Directive").
- [2] Directive 2004/49/EC and further amendments (the "Railway Safety Directive").
- [3] Commission Decision of 25 January 2012 on the technical specification for interoperability relating to the control-command and signalling subsystems of the trans-European rail system (2012/88/EU) and further amendments (2012/696 and 2015/14).
- [4] European Railway Agency, *Guide for the application of the TSI for the Subsystems Control-Command and Signalling Track-side and On-board*, ver.2.0, 2013.
- [5] Commission Recommendation of 5 December 2014 on matters related to the placing in service and use of structural subsystems and vehicles under Directive 2008/57/EC and 2004/49/EC of the European Parliament and of the Council (2014/897/EU).
- [6] Commission Decision of 9 November 2010 on modules for the procedures for assessment of conformity, suitability for use and EC verification to be used in the TSI adopted under Directive 2008/57/EC of the European Parliament and of the Council (2010/713/EU).
- [7] RFC1 NSA working group, *Guideline for CCS authorization on rail freight corridor 1*, ver.1.0, 2013.
- [8] Commission Implementing Regulation (EU) No 402/2013 of 30 April 2013 on the common safety method for risk evaluation and assessment and repealing Regulation (EC) No 352/2009.



# Application of Automatic Landing using GPS-GBAS

Olivera Petrovska<sup>1</sup>, Vesna Antoska Knights<sup>2</sup> and Stojche Deskovski<sup>3</sup>

**Abstract** – In this paper a mathematical model of the aircraft and the system for automatic guidance for landing using GPS – GBAS, is presented. Simulations are made for landing guidance system when using the ILS and GPS – GBAS. For simulations the programming package MATLAB/SIMULINK is used. In this paper, comparative analysis of the results of the simulations and the discussion the advantages of GBAS in terms of ILS are represented.

**Keywords** – Automatic landing, Mathematical model, Aircraft, autopilot, GPS, GBAS.

## I. INTRODUCTION

Aircraft landing is one of the most important phases of flight in terms of the technical complexity and security. The airports and the aircraft are equipped with appropriate equipment for landing but it is not enough sufficiently response for the landing problem especially when landing is under low visibility conditions.

This is a constant challenge to finding new solutions to improve the process of landing in terms of overcoming the problems of the low visibility conditions and better safety. The classic instrument landing systems (ILS, MLS) does not provide landing of aircraft in all weather conditions, unlike of those systems the new technology which is based on GPS (DGPS, WAAS, LAAS) provides an opportunity for a new approach to solving the problem of landing offering flexible solutions and application

Satellite based landing system GBAS has many advantages in terms of ILS [1]. From the perspective of this paper, can appoint the fact that GBAS provides guidance of aircraft during approach and landing in different curvilinear trajectories defined by control points (Way Points) unlike from ILS which provides guiding of the one fixed path determined by antennas of localizer and provider of the line of landing. Also GBAS, a one GBAS station on the airport can simultaneously serve all runways, while an ILS is only tied to a one runway. Using GPS and GBAS for landing requires installing the proper equipment at airports, equipping aircraft with additional devices and modifying new algorithms for automatic landing implemented in the autopilot.

This paper is organized as follows. In Section II, is

described a system for automatic landing of aircraft using the GPS. In Section III is given a short description of mathematical models of aircraft. In Section IV is described a concept of guidance system for automatic landing of aircraft using the GBAS. In Section V are presents the results of a simulation for the system for automatic landing using MATLAB/SIMULINK. Conclusion is given in Section VI.

## II. AUTOMATIC LANDING SYSTEMS USING GPS

Ground-Based Augmentation System (GBAS) is a system that provides differential corrections and integrity monitoring of Global Navigation Satellite Systems (GNSS). GBAS provides navigation and precision approach service in the vicinity of the host airport (approximately a 23 nautical mile radius), broadcasting its differential correction message via a very high frequency (VHF) radio data link from a ground-based transmitter. GBAS yields the extremely high accuracy, availability, and integrity necessary for Category I, and eventually Category II, and III precision approaches [1],[6]. GBAS demonstrated accuracy is less than one meter in both the horizontal and vertical axis. The GBAS architecture is shown on Figure 1.

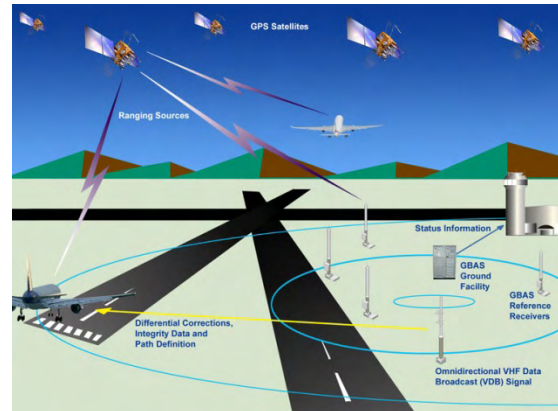


Fig. 1. GBAS architecture

Ground-Based Augmentation System (GBAS) is constituted of three subsystems: satellite, ground and avionics. The Satellite subsystem is formed by satellites GNSS (GPS, etc.). This subsystem provides signals for GPS receivers set of aircraft and signals for ground GBAS station.

The ground subsystem is constitute from three or four referent receivers, which collect pseudo ranges for all the primary GNSS satellites in view and computes and broadcasts differential corrections and integrity-related information for them based on its own surveyed position. These differential corrections are transmitted from the ground system via a Very High Frequency (VHF) Data Broadcast (VDB). GBAS also

<sup>1</sup> Olivera Petrovska is with the Faculty of Technical Science, Bitola, Ivo Lola Ribar bb, 7000 Bitola, Macedonia  
E-mail: [petrovskaoli@gmail.com](mailto:petrovskaoli@gmail.com)

<sup>2</sup> Vesna Antoska Knights is with the Faculty of Technology and Technical Sciences Veles, Dimitar Vlahov bb, 1400 Veles, Macedonia  
E-mail: [vesna.knights@uklo.edu.mk](mailto:vesna.knights@uklo.edu.mk), [vesna.antoska@gmail.com](mailto:vesna.antoska@gmail.com)

<sup>3</sup> Stojche Deskovski is with the Faculty of Technical Science, Bitola, Ivo Lola Ribar bb, 7000 Bitola, Macedonia  
E-mail: [stojce.deskovskii@gmail.com](mailto:stojce.deskovskii@gmail.com)

transmit information, which is used to define the reference trajectory.

GBAS Avionic subsystem allows users to receive and decode VDB data and integrate with navigation signals which coming from the satellite constellation.

Existing systems for landing, ILS, have disadvantage and limitations where to the new technology for landing based on GPS, GBAS, should in due time to fully replace them. Some of the benefits from the implementation of GBAS are: reduction of critical and sensitive areas; curved approach; positioning Service; provision of service in several runways in the same airport; provision of several approach glide angles and displaced threshold; guided missed approach; and adjacent airports use.

### III. AIRCRAFT MATHEMATICAL MODELS

With a view to test the process of landing with simulations is used the complete 6-DOF model of aircraft and linearized models for longitudinal and lateral movements [6], [7]. The complete 6-DOF model is presented in the following system of differential equations written in vector-matrix form:

$$\dot{\mathbf{V}}_B = -\tilde{\boldsymbol{\omega}}_B \mathbf{V}_B + m^{-1}(\mathbf{R}_B^A \mathbf{F}_B) + \mathbf{g}_B \quad (1)$$

$$\dot{\boldsymbol{\omega}}_B = -\mathbf{I}_B^{-1} \tilde{\boldsymbol{\omega}}_B \mathbf{I}_B \boldsymbol{\omega}_B + \mathbf{I}_B^{-1}(\mathbf{M}_B^A + \mathbf{M}_B^F) \quad (2)$$

$$\dot{\mathbf{q}} = \mathbf{f}(\boldsymbol{\omega}_B, \mathbf{q}) \quad (3)$$

$$\dot{\mathbf{R}}_G = \mathbf{T}_{BG}^T \mathbf{V}_B \quad (4)$$

where:  $\mathbf{V}_B = [u, v, w]^T$  is the velocity vector of the center of mass, CM, in the body fixed coordinate system  $B(P; x, y, z)$ ,  $\boldsymbol{\omega}_B = [p, q, r]^T$  is the angular rate vector,  $p$ ,  $q$  and  $r$  are the angular rates of rolling, pitching and yawing of the body,  $\mathbf{R}_G = [x_0, y_0, z_0]^T$  is a position vector of the center of mass,  $\mathbf{g}_B = [g_x, g_y, g_z]^T$  is the gravitational acceleration;  $\tilde{\boldsymbol{\omega}}_B$  is skew symmetric matrix which is composed of components of the vector  $\boldsymbol{\omega}_B$ ;  $\mathbf{T}_{BG}$  is transformation matrix from  $G$  to  $B$  coordinate systems;  $m$  is the mass and  $\mathbf{I}_B$  is the inertia matrix;  $\mathbf{R}_B^A = [X, Y, Z]^T$  and  $\mathbf{F}_B = [F_x, F_y, F_z]^T$  are aerodynamic and reactive forces,  $\mathbf{M}_B^A = [L, M, N]^T$  and  $\mathbf{M}_B^F = [M_x, M_y, M_z]^T$  are vectors of aerodynamic and propulsive moments, respectively. Subscripts B and G denote that vectors are presented in body (B), or in inertial coordinate system (G). The matrix  $\mathbf{T}_{BG}$  is the product of the elementary matrices of rotation;  $\mathbf{q}$  is a vector with four components-quaternions which uniquely determines attitude (angular orientation) of the body.

For calculation of the matrix of transition  $\mathbf{T}_{BG}$  and Euler angles  $\mathbf{S} = [\phi, \theta, \psi]^T$  are used quaternion:

$$\mathbf{T}_{BG} = \mathbf{T}_{BG}(\mathbf{q}) \quad (5)$$

$$\mathbf{S} = \mathbf{S}(\mathbf{q}) \quad (6)$$

The vector of state of the model with quaternions is composed of 13 components:

$$\mathbf{x} = [\mathbf{V}_B^T, \boldsymbol{\omega}_B^T, \mathbf{q}^T, \mathbf{R}_G^T]^T = [V_x, V_y, V_z, p, q, r, q_0, q_1, q_2, q_3, x_0, y_0, z_0]^T \quad (7)$$

The linearized models of flight are derived from general 6-DOF model and models are described in the matrix-vector form. For longitudinal movement linear model has form:

$$\begin{bmatrix} \dot{u} \\ \dot{\alpha} \\ \dot{q} \\ \dot{\theta} \\ \dot{h} \end{bmatrix} = \begin{bmatrix} X_u & X_\alpha & 0 & -g \cos \theta^0 & 0 \\ Z_u & Z_\alpha & 1 & -\frac{g \sin \theta^0}{U^0} & 0 \\ U^0 & U^0 & 0 & 0 & 0 \\ M_u & M_\alpha & M_q & -\frac{M_\alpha g \sin \theta^0}{U^0} & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & U^0 & 0 \end{bmatrix} \begin{bmatrix} u \\ \alpha \\ q \\ \theta \\ h \end{bmatrix} + \begin{bmatrix} X_{\delta_m} & X_{\delta_r} \\ Z_{\delta_m} & Z_{\delta_r} \\ M_{\delta_m} & M_{\delta_r} \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \delta_m \\ \delta_r \end{bmatrix} \quad (8)$$

In equations (8) the state variables  $u, \alpha, q, \theta, h$  are the small declinations from their nominal values. The elements of the matrix depend on geometry, mass and aerodynamics features of the specific aircraft and can be calculated by the formulas given in [6], [7]. In the simulation are used data for aircraft Boeing 747-100 taken from [7].

### IV. AUTOMATIC AIRCRAFT LANDING SYSTEM USING GBAS

The concept of the guidance system for landing based on GBAS is presented on Fig.2. In short, DGPS station at the airport, through the GPS antennas (4 antennas) receiving signals from satellites on which determines its own position. Since, on the other hand, the exact position of the DGPS station is exactly known, calculates the error in position.

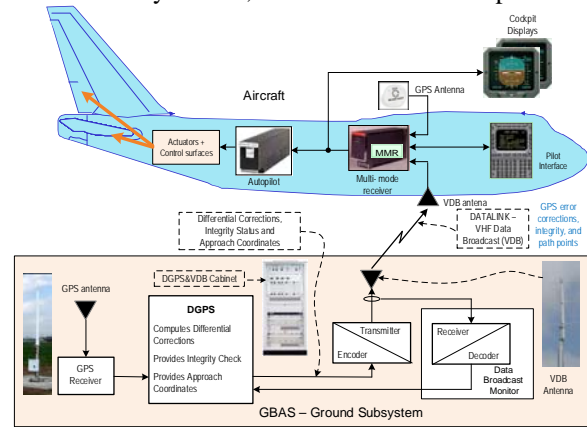


Fig. 2 GPS based landing system using GBAS

The error in position is the same error that has a position of the aircraft defined on the GPS signal received in the receiver in the aircraft. Correction of this error is made on the basis of the calculated differential correction in DGPS station which is transmitted through VDB link multi - mode receiver of the aircraft. Multi - Mode Receiver - MMR receives signals from the GPS antenna on the aircraft, the pilot through the appropriate interface, the VDB antenna, and the ILS in the case when ILS is used on landing. The aircraft through VDB link gets information about the coordinates of the reference points for generating the reference trajectory for approach and landing.

The geometry of the landing in space is shown on Fig.3. In the case of ILS landing the receiver of aircraft measured errors, The guidance of aircraft using GPS, or GBAS, is used current coordinates of CG measured with GPS aircraft receiver which are corrected with error corrections obtained by VDB line. Corrected current coordinates are compared with the reference coordinates (trajectory) obtained on the basis of the reference points obtained from GBAS.

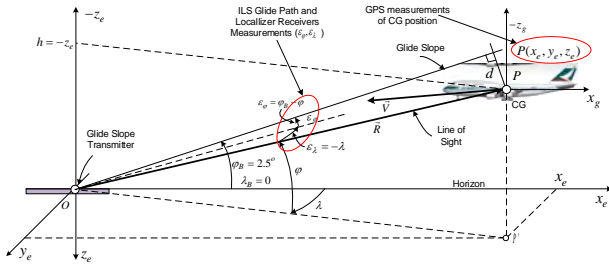


Fig. 3: Geometry of landing in the space

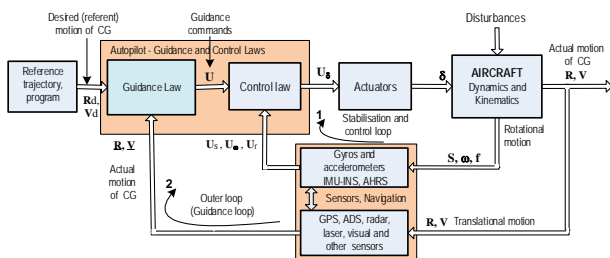


Fig. 4 Block diagram of the guidance and control system of the aircraft

Fig.4 shows block diagram of the flight guidance and control system of the aircraft which is used for the development of the simulation model in MATLAB/ SIMULINK. The inner loop of stabilization and control mainly includes loops of stabilization and control with rotation in relation to the three axes - roll, pitch and yaw, and the loop for speed control of aircraft. The outer loop (guidance loop) provides guidance of aircraft along reference trajectory using information about the current position from the appropriate sensors.

When is used ILS we have a Line of Sight - LOS guidance, while using GBAS we have tracking curvilinear reference trajectory. Guidance laws in both cases significantly are different. In case of GBAS is required generator of reference trajectory, and to provide higher accuracy for tracking the reference trajectory, besides the feedback controller, often is introduced feedforward controller which can use inverse model of the dynamics of the aircraft.

## V. SIMULATION OF THE AUTOMATIC AIRCRAFT LANDING SYSTEM USING GBAS

To appraise the performance of the guidance system for landing simulation models are developed for the cases when are used the ILS and GBAS. Fig.5 shows the basic block diagram of the SIMULINK model of the aircraft guidance and control system, Fig.4, in 3D space. The model allows simulations of different modes of flight. The presented model

uses data of aircraft Boeing 747-100 concerning on the approach phase and landing given in [13].

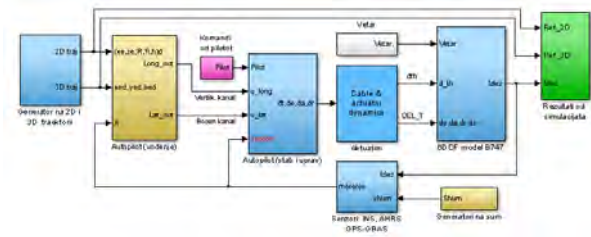


Fig.5. Basic SIMULINK block diagram of the guidance with 6DOF model of aircraft

Selected results of the simulation are shown in the figures below. The Fig.6 shows the reference and actual trajectory of landing and Fig.7 shows a vertical landing velocity. Trajectory of landing (Fig.6) ends with "flare" maneuver by exponential curve that starts at a height of 15 m above the threshold of runway. Vertical landing velocity has a value of about  $-3.3$  m/s, which is reduced during the "flare" maneuver to value of about  $-0.5$  m/s at the point of contact is required for reliable contact and movement on the runway. Trajectory of landing in Fig. 6 has an angle of  $2.5^\circ$  and correspond in the case when ILS is used.

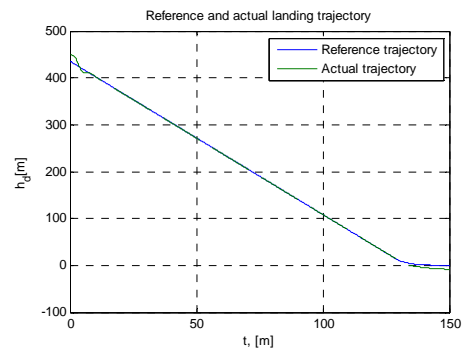


Fig. 6. Following the line of landing

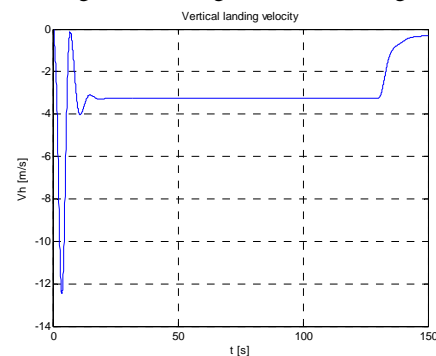


Fig. 7. Vertical speed of landing

The next experiment concerns the guidance landing on curvilinear reference trajectory which at the beginning has a horizontal part, after a certain time, a parabolic curve is transformed into direction of angle of  $2.5^\circ$ , and ends with a "flare" maneuver. Landing in this case is achieved using GPS and GBAS. Fig.8 shows tracking of curvilinear reference trajectory, and Fig.9 shows the vertical speed of the aircraft

landing. Both diagrams are made in case of the effect of rear wind with average speed of 10 m/s and variance 2.

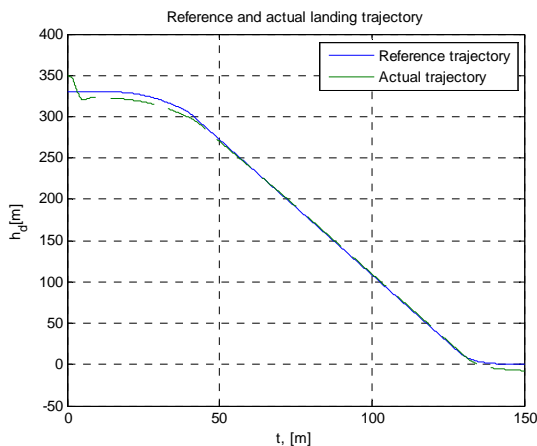


Fig. 10. Tracking curvilinear reference trajectory

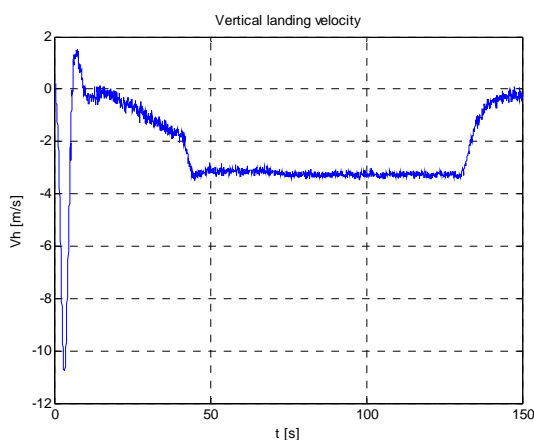


Fig. 9. Vertical landing velocity in condition of wind

## VI CONCLUSION

In this paper the system of automatic aircraft landing using GPS based technology is presented. The system GBAS (Ground Based Augmentation System) is composed of ground and aircraft subsystems. It uses signals from GPS for determining the position in 3D space, and has more advantages than existing ILS landing systems which in the near future should be fully replaced. These advantages are outlined in the paper. What especially is highlighted is the possibility of aircraft guidance in the phases of approach and landing along flexible curvilinear trajectories. This fact is suggesting the possibility of landing under larger angles and avoiding barriers in the area around the airport, reducing noise over populated areas etc. In terms of traffic, the GBAS is increasing the capacity of airports and supports a larger number of aircraft landings per unit time. With the development of GPS technology and increasing the accuracy of measuring the coordinates, and using differential GPS, which is at the base of GBAS, achieved accuracies that meet the prescribed requirements for landing. In this paper, aircraft models and the guidance system models of landing which are used for Matlab/SIMULINK simulation of the landing process

are presented. The modelling approach and simulation is a good methodology for developing landing algorithms, evaluation of system performance and the effects of the action of various disorders such as wind and errors of measurement of sensors.

## REFERENCES

- [1] International civil aviation organization (ICAO), **Guide for Ground Based Augmentation System Implementation**, May 2013.  
<http://www.icao.int/SAM/eDocuments/GBASGuide.pdf>
- [2] M. Steen, T. Feuerle, M. Stanisak, T. Yoshihara, P. Hecker, GBAS Curved Approach Procedures: Advantages, Challenges and Applicability, ICAS 2012- 28<sup>th</sup> International Congress of the Aeronautical Sciences
- [3] Ambrožová, S. Pleninger: **Implementation of GBAS System at the Václav Havel Airport**, Number 2, Volume VIII, July 2013.  
[http://pernerscontacts.upce.cz/30\\_2013/Ambrozova.pdf](http://pernerscontacts.upce.cz/30_2013/Ambrozova.pdf)
- [4] S. Agarwal, H.B. Hablani: **Automatic Aircraft Landing over Parabolic Trajectory using Precise GPS Measurements**, 2nd International Conference and workshop on Emerging Trends in Technology, Proceedings published by International Journal of Computer Applications, pages 38-45, 2011.
- [5] E. De Lellis, F. Corrado, U. Ciniglio, Pa. Canzolino, L. Garbarino, S. Gaglione, V. Nastro: **An EGNOS Based Navigation System for Highly Reliable Aircraft Automatic Landing**,  
[http://dsa.uniparthenope.it/dsa\\_web/LinkClick.aspx?fileticket=E7ExYS7S4Wc%3D&tabid=199&mid=908&language=en-US](http://dsa.uniparthenope.it/dsa_web/LinkClick.aspx?fileticket=E7ExYS7S4Wc%3D&tabid=199&mid=908&language=en-US)
- [6] О. Петровска, **Развој на алгоритам за слетување на воздухоплов со примена на GPS во услови на надворешни нарушувања**, докторска дисертација, Технички факултет, Битола, 2015.
- [7] J. Roskam: **Airplane Flight Dynamics and Automatic Flight Controls, part I**, DAR corporation, 120 East Ninth Street, Suite 2, Lawrence, KS 66044, USA, 2001.
- [8] Underwood, SmartPath<sup>®</sup> SLS-4000 - Ground Based Augmentation System (GBAS), Honeywell ppt presentation ([PPT]GBAS - Airport Show)
- [9] D. Delahaye, S. Puechmorel, P. Tsiotras and E. Feron, **Mathematical Models for Aircraft Trajectory Design: A Survey**, Author manuscript, published in "EIWAC 2013, 3rd ENRI International Workshop on ATM/CNS, Tokyo : Japan (2013)
- [10] J. Bell, Basic GPS Navigation, A practical guide for GPS navigation, Edited November 16, 2010



# **Measurement Science and Technology**



# Evaluation of Wines Based on Mathematical and Statistical Analysis

Krastena Nikolova<sup>1</sup>, Ivan Maslinkov<sup>2</sup>, Irena Ivanova<sup>3</sup>

**Abstract** –A total 17 samples of white and red wines have been investigated and characterized with regard to the following optical properties – color coordinates, lightness, Hue angle, chroma and transmission coefficient in visible area. Based on the examined parameters the wines are evaluated in their complexity by hierarchical cluster analysis. The samples are grouped into two main clusters presented graphically in the dendrogram. The first one contains all samples of white wines, the second one combines red wines and separates them by kind of grape, production method and region of growing the grapevine. The made classification allows increasing of the objectivity of the evaluation and uses optical properties in various selection programs. A factor analysis using PCA was applied on the basis of the found correlation matrix. Thus the number of the studied indices was reduced to 2 factors, which explained 93.32% from the entire variation. The first one was related mainly to color parameters a and b, lightness L, Hue angle H in SIELab and transmission coefficient at four different wavelength, which are important for optical properties of wine. The second one unified the color coordinates x, y in XYZ colorimetric system and chroma Cab in SIELab colorimetric system.

**Keywords** –wine, optical properties , cluster analysis, factor analysis, PCA, dendrogram

## I. INTRODUCTION

Wine is a fermented alcoholic beverage containing various compounds of different type, with polyphenols which are one of the most important components influencing their quality parameters [1]. Polyphenols contribute to the organoleptic characteristics, such as color, astringency and bitterness and also exert antimicrobial and antioxidant properties [2].

Authenticity and commercial value of wines is often linked to its geographical origin and certain countries or regions are known for producing excellent wines of high commercial value [3]. Due to its composition and worldwide availability, the controlled denomination of origin (CDO) is usually required to demonstrate the provenance of wines. In this work the optical properties of wines are investigated and the following goals are presented: using mathematical and statistical methods to group different types white and red wines on the base of investigated physico- chemical parameters and to identify those affecting most heavily on the

clustering.

The same technics are used for estimation the European kinds of wine by factor and cluster analysis [4-5].

## II. MATERIALS AND METHODS

The study involves 17 types of wines commercially available or obtained from a winery in the region of Starosel village. The specification of the tested wines is presented in Table 1.

TABLE I  
INVESTIGATED SAMPLES

	Place	Type
1	Wine-cellar	Cabernet Sauvignon
2	Wine-cellar	Rkaciteli
3	Wine-cellar	Misket
4	Wine-cellar	Merlot
5	Wine-cellar	White wormwood wine
6	Wine-cellar	Red wormwood wine
7	Wine-cellar	Rosé
8	Wine-cellar	Cabernet Sauvignon
9	supermarket	Gamza
11	Supermarket	Red wine
12	Supermarket	Vechernij zvon
13	Supermarket	Mehandjisko
14	Supermarket	Mavrud
15	Supermarket	Trakia
16	Supermarket	Manastirsko shushukane
17	Supermarket	Assenitza

The following 14 physical parameters have been investigated: color coordinates (1) X , (2) Y, (3) Z, (4) x, (5) y, lightness (6) L, color parameters (7) a, (8) b, (9) Cab, Hue angle (10) H, transmission coefficients at 445 nm (11), 495 nm (12), 550 nm (13), 625 nm (14).

Using a software package VISIONlite ColorCalc for spectrophotometer Helios Omega with a cuvette of a 10 mm length, the color parameters in CIELab colorimetric system have been obtained. All measurements have been carried out at room temperature. Color coordinates, color parameters a, b and brightness L of tested samples have been measured. Parameters such as chroma Cab and hue angle H were defined as follows:

$$C_{ab} = \sqrt{a^2 + b^2}, \quad h_{ab} = \arctg(b/a) \quad (1)$$

To compare and group the types of wines on the base of all investigated physical parameters a hierarchical cluster

<sup>1</sup>Krastena Nikolova is with the Faculty of Economics at University of Food technologies, Plovdiv, Bulgaria, E-mail: kr.nikolova@abv.bg

<sup>2</sup>Ivan Maslinkov is with the Technical Faculty at University of Food technologies, Plovdiv, Bulgaria

<sup>3</sup>Irena Ivanova is with the Faculty of Economics at Agricultural university, Plovdiv 4000, Bulgaria. irenaiv@au-plovdiv.bg

analysis with a measure of similarity Euclidean distance [6, 7] have been used. The data have been preliminary standardized in order to avoid the influence of different dimensions.

A factor analysis was conducted. Thus the input set of correlating data was transformed into a new set with a smaller number of uncorrelated artificial variables, so-called „factors” or „principal components”, which explain the greatest part of the data variation. The reduction in the number of original variables was achieved by grouping the correlated variables in a common factor and the separation of uncorrelated ones in different factors. A stronger distinction of the variables to one or another factor is performed by further factor rotation using Varimax method.

All data analyses have been performed using the SPSS software program.

### III. RESULTS AND DISCUSSION

17 samples of wine have been analyzed in regard to 14 physical parameters. Using cluster analysis, the studied wines have been divided according to their similarity and distance depending on the studied physical parameters. The grouping of the samples in clusters is presented with the dendrogram on Fig. 1 and the values of the distances between groups is shown in Table II.

TABLE II  
COMBINING OF CLUSTERS AND DISTANCES BETWEEN GROUPS

Steps №	Combined clusters		Coefficients
	Cluster 1	Cluster 2	
1	14	15	0,117
2	1	4	0,227
3	10	16	0,284
4	10	13	0,356
5	2	3	0,637
6	10	12	0,783
7	6	9	0,929
8	11	17	1,191
9	1	7	1,236
10	2	5	1,249
11	6	14	2,568
12	6	10	3,490
13	6	11	6,948
14	2	8	7,877
15	1	6	9,875
16	1	2	58,220

The results from the investigation show that the wines are divided into two main clusters. The first of them includes 13 samples and is divided into some sub-clusters. The closest are the samples 14 and 15- red wormwood wine and gamza. Samples 10, 16, 13 and 12 form another sub-cluster. They are combined from the fact, that all are produced from the same variety of grape – Cabernet Sauvignon. The samples 1,4 and

7- red and pink wines, united in other sub cluster obtained from the wine cellar. The samples 11 and 17 are semi-dry wines and this is the reason of their similarity.

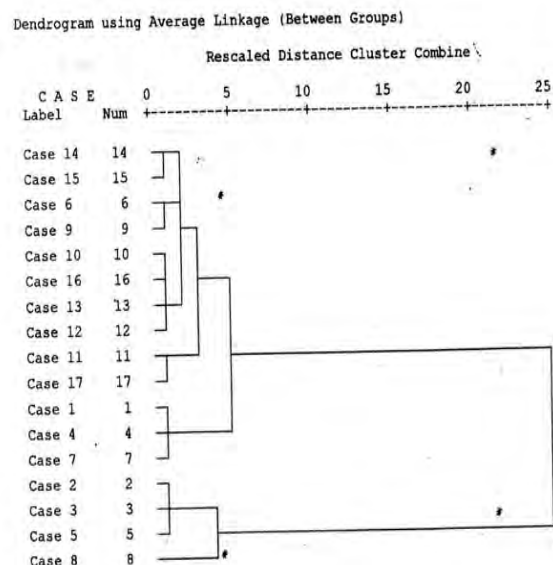


Fig.1. Dendrogram based on the hierarchical cluster analysis

The second cluster combines samples of the white wines 2, 3, 5 and 8, which have been obtained directly from the wine-cellar. The distance between 2 and 8 is the greatest in this cluster - 7,88.

The percentage distribution of the wine in the distinct sub-cluster is presented in table III.

TABLE III  
DISTRIBUTION OF THE WINE SAMPLES IN CLUSTERS

№	Samples in clusters	Number of samples	Distribution, %
1.1	14, 15	2	11,8
1.2	6, 9	2	11,8
1.3	10,16, 13,12	4	23,5
1.4	11, 17	2	11,8
1.5	1, 4, 7	3	17,6
2	2, 3, 5, 8	4	23,5
total		17	100 %

The samples from white wine and the second sub-cluster of the first one have the highest percentage of the samples - 23.5%. The wines from the first group are characterized with high lightness between 93-98 relative units, Hue angle is in the range between 80.0 and 87.0, these samples have high transmission coefficient at wavelengths 445 nm and 625 nm, which are 66 % - 87 % and 83% - 93% respectively. The yellow color predominates over the blue, but it is very weakly expressed – the color parameter “b” is between 5 and 12 . The wines which are from obtained from the wine-cellar possess the smallest Hue angle – between 16.0 and 27.0, the lowest lightness 5-13 relative units and their transmission coefficient is less than 1% at 445 nm, it is from 1 % to 8 % at 625 nm. The red component prevails over the yellow (a>b). The red Wormwood wine and the Gamza are



characterized with almost the same lightness L and chroma Cab. The values of the red and yellow components are similar and they are almost equal a~b. The sub-cluster with the wines obtained from the supermarket has lightness in the range from 15 to 23, Hue angle from 41.0 to 49.0. The lightness for the Mavrud and Trakia wines is close to this of the red wormwood wine and Gamza - at about 30 a.u., Hue angle is between 53.0 and 54.0. However, the yellow component dominates the red (b>a) for them.

It is known that the investigated parameters are involved with different weight. To determine the influence of parameters in the split into clusters factor analysis with the method of principal components (PCA) has been performed. In principle the component analysis helps to identify variables and their relative importance in explaining the observed variation and types of wine. Data was found suitable for PCA and the factor analysis gives the correct results in our investigation, because the value of KMO-test is 0,705 (> 0, 5) and Bartlett's test of sphericity has a significance level lower than 0.001. The results of PCA are presented at Figure 2. Two principle components out of the fourteen have eigenvalued greater than one and explained 93,32 % of total variation. That is why we choice two factors. The first factor (principle component one - PC 1) explains 81, 57 % of total variation and strongly correlated with color parameters in color system CIELab and transmission coefficients in visible area. The color coordinates x, y in XYZ colorimetric system and chroma Cab in SIELab colorimetric system mainly contributed towards the second factor (principle component two - PC2), which explains 11,75 % of total variation. This factor integrates color coordinates in the color space for a small color differences SIELab and spectral characteristics in the visible area PC 2 gives the relation between some parameters from different color systems XYZ and SIELab. Color coordinates x, y determines one point in the plane which is connected to the characteristics of the color-brightness, dominant wavelength and purity. These characteristics correlated with the Chroma - parameter in SIELab color system. The rotated component matrix, obtained with varimax transformation of the main components is presented in table 4 . The results of the factor analysis are shown on Figure 2.

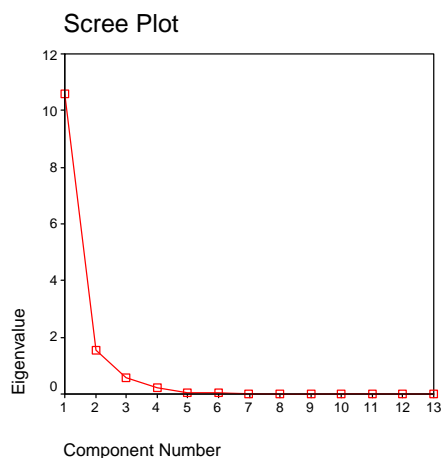


Fig.2 Factor analysis

TABLE 4  
THE ROTATED COMPONENT MATRIX WITH VARIMAX  
TRANSFORMATION OF PRINCIPAL COMPONENTS

№	Parameter s	Principal Components	
		1	2
1	X <sub>1</sub>	0.711	0.699
2	X <sub>2</sub>	0.759	0.640
3	X <sub>3</sub>	-0,657	-0,729
4	X <sub>4</sub>	0,044	0,828
5	X <sub>5</sub>	0,549	0,823
6	X <sub>6</sub>	-0,887	-0,418
7	X <sub>7</sub>	-0,970	0,053
8	X <sub>8</sub>	-0,967	-0,185
9	X <sub>9</sub>	0,158	0,876
10	X <sub>10</sub>	0.765	0.623
11	X <sub>11</sub>	0.760	0.633
12	X <sub>12</sub>	0,761	0,632
13	X <sub>13</sub>	0,760	0,590

## IV CONCLUSSIONS

The research allows the following conclusions:

Knowing the spectral and color characteristics of the wines allow the separation of the samples according to the predominant grape variety or the place of purchase - from a winery or from a supermarket.

The most distinguishing samples according to the angle of Hue - with the smallest values are samples obtained from the wine cellars, followed by those obtained from the supermarkets and which are not semi-dry wines.

## REFERENCES

- [1] M. S. Fernández-Pachyn, D. Villaco, M. C. García-Parrilla, & A. M. Troncoso, "Antioxidant activity of wines and relation with their polyphenolic composition", *Analytica Chimica Acta*, vol. 513, p. 113-118, 2004
- [2] J. Godoy-Navajas, M. P. Aguilar-Caballos, & A. Gymez-Hens, "Automatic determination of polyphenols in wines using laccase and terbium oxide nanoparticles", *Food Chemistry*, vol. 166, p. 29-34, 2015
- [3] V. S. Šelih, M. Šala, & V. Drgan, "Multi-element analysis of wines by ICP-MS and ICP-OES and their classification according to geographical origin in Slovenia", *Food Chemistry*, vol. 153, p. 414-423, 2014
- [4] Pges Escofier, "Multiply factor analysis", *Computational Statistics and data analysis*, vol. 18, p. 121-140, 1994
- [5] Katsota Arvanitoyannis, "Aplication of quality control methods for assessing wine authenticity", *Food science technology*, vol. 10, p. 321-336, 1999
- [6] B. Duran, P. Odelle, *Cluster analysis*, Moscow, Statistics, 1977
- [7] JH. Ward, "Hierarchical grouping to optimize all objective functions", *Journal of the American Statistical Association*, vol. 58, pp. 236-244, 1963

This Page Intentionally Left Blank

# Tablet Computers and Extremely Low-Frequency Magnetic Field Emission

Darko Brodić<sup>1</sup>, Milena Jevtić<sup>1</sup>, Jordan Radosavljević<sup>2</sup> and Ivo R. Draganov<sup>3</sup>

**Abstract** – In this paper, the tablets' emission of the magnetic field in the extremely low frequency region has been explored. The experiment is conducted on five Android tablets. For the testing purposes, we used the measurement geometry which is previously proposed for the measurement of laptop's magnetic field. The measured values are compared to the TCO proposed safe standard. At the end, the list of the proposed suggestions in order to avoid the exposure to the high magnetic field from the tablets is proposed. The obtained measurement values can be invaluable for users as well as for designers of the tablets.

**Keywords** – Magnetic field, Measurement, Statistical analysis, Tablet computers.

## I. INTRODUCTION

Magnetic fields can be created naturally or artificially. If it has a natural origin, then it is created by Earth. In contrast, the artificial magnetic field is the product of electric current which is flowing usually through electrical and electronic device. Hence, the magnetic fields are present everywhere in our environment. One of the main characteristics that define a magnetic field is its frequency. The magnetic field of different frequencies can interact with the human's body in different ways. Although, it is not destructive as ionizing radiation, such as gamma rays, cosmic rays, and X-rays, it can have some potential negative effects to the humans.

The magnetic field is the strongest if it is close to their origin. However, it decreases rapidly if we are getting away from the source in the case of low frequencies. It should be noted that the magnetic field is resistant to the walls of the building, but some materials can block its spreading.

The time varying magnetic fields produced by electrical appliances is understood as the extremely low frequency (ELF) magnetic field. The ELF magnetic field contains the frequencies from 50 Hz to 300 Hz. Hence, it is everywhere around us.

In this paper, we have special attention to the extremely low frequency magnetic field. Its source can be: transmission

lines, substations, kitchen appliances, computers, office equipment, etc. At low frequencies, the levels of induced currents inside the body are too small to produce obvious effects. In spite of that, these fields induce currents within the human body, which can be sufficient to produce a various effect to the humans during a long work. Also, there is no doubt that short-term exposure to very high levels of magnetic fields can be harmful to the human health. The current public concern focuses on possible long-term health effects caused by exposure to magnetic fields at levels below those required to trigger acute biological responses. Up to date, the extensive research in many studies which have been conducted in an ELF magnetic field area did not give a clear answer how it is harmful to the human health. With more and more research expertise available, it has become increasingly likely that exposure to electromagnetic fields constitutes a serious health hazard. Nevertheless, some uncertainties remain. The original scientific discussion about the interpretation of controversial results has shifted to become a societal as well as the economic issue.

International guidelines and national safety standards for magnetic fields are developed on the basis of the current scientific knowledge to ensure that the fields humans encounter are not harmful to health. ICNIRP issues guidelines on the basis of the current scientific knowledge [2]. Most countries draw on these international guidelines for their own national standards. Standards for low-frequency magnetic fields ensure that induced electric currents are below the normal level of background currents within the body. The TCO standard proposed the prescribed geometry of the measurement and test procedure with the following reference values of magnetic flux density [3]:

- For a band I (5 Hz to 2 kHz),  $\leq 200$  nT measured at 0.30 m in front of and around the laptop computer,
- For band II (2 kHz to 400 kHz),  $\leq 25$  nT measured at 0.30 m in front of and around the laptop computer.

A tablet computer, commonly shortened tablet, is a mobile computer with a touchscreen display, circuitry, and battery in a single device. Tablets come equipped with sensors, including cameras, a microphone, an accelerometer and the touchscreen display. The touchscreen display uses the recognition of finger or stylus gestures replacing the mouse and keyboard. They usually feature on-screen, pop-up virtual keyboards for typing. Tablets may have physical buttons for basic features such as speaker volume and power, and ports for network communications and battery charging. They are characterized and grouped according to the size of the screen. It is supposed that the smallest tablet incorporates the screen of at least 7".

Up to now, many researchers [4], [5], [6], [7], [8], [9] measured magnetic field produced by laptops of different brands. The magnetic field values were considerable higher

<sup>1</sup>Darko Brodić is with the Technical Faculty in Bor, University of Belgrade, Vojske Jugoslavije 12, 19210 Bor, Serbia, E-mail: dbrodic@tf.bor.ac.rs

<sup>1</sup>Milena Jevtić are with the Technical Faculty in Bor, University of Belgrade, Vojske Jugoslavije 12, 19210 Bor, Serbia, E-mail: mjjevtic@tf.bor.ac.rs

<sup>2</sup>Jordan Radosavljević is with the Faculty of Technical Sciences, University of Priština, Kneza Miloša 7, 38220 Kosovska Mitrovica, Serbia, E-mail: jordan.radosavljevic@pr.ac.rs

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

than the values recommended by some recent guidelines for laptops magnetic field emissions [6], [10] and those considered the risk of tumor development. Refs. [4], [5] showed that the EMF values became alarmingly high when the laptop was used close to the body. According to ref. [5], the laptop induces currents that are within 34,2% to 49,8% ICNIRP [2] when is close to the body, which is very important to the adult's body and to the fetus in pregnant women. Also, the power adapter induces strong intra-corporal electric current densities in the fetus and in the adult subject, which are respectively 182-263% and 71-483% higher than ICNIRP basic restriction recommended to prevent adverse health effects [5].

Ref. [9] has shown that laptops may be harmful to sperm with samples exposed to a wireless internet connected laptop showing a significant decrease in sperm motility and an increase in sperm DNA fragmentation. These researchers considered that keeping a laptop connected wirelessly to the internet on the lap near the testes may result in decrease male fertility. Ref. [11] pointed that DNA reacts to electromagnetic fields, making the long-term consequences of repeated microwave exposures to genetic material. Laptop computers are the high-energy source of thermal, electric and electromagnetic fields. The thermal effects can be blocked by using a lap pad, an electric field can be blocked by reflective material or metal, but magnetic field components cannot be blocked. However, situation with tablets is even worse due to their higher portability.

During the measurement of low-frequency magnetic field emissions, magnetic flux density was characterized in terms of field amplitude and "weighted peak" (WP) index, [4], [12], [13], that implement the weighted peak approach, which is well suited in the case of complex waveforms. Security limits for exposure to a magnetic field are different according to the different authors: 1  $\mu$ T [14], 0,4  $\mu$ T [15], [16], 0,2  $\mu$ T [17]. Ref. [18] suggests that exposure to 1  $\mu$ T EMF alters sleep by reducing total sleep time and sleep efficiency. Ref. [19] postulates experiments that may be used to characterize the nature of the transduction step in which a magnetic or electric field is converted into a biological signal.

In this paper, we conducted the experiment on tablet to measure and evaluate their ELF magnetic field emission. Hence, any high frequency magnetic field emission from the tablets' component like WIFI, 3G/4G or similar is out of the scope of this paper. To our best knowledge, there have been no study of this problem yet. Because of wide-spreading of tablet computers among the younger population, it is of great importance. It should be pointed that the experimentation was conducted on typical tablets. It means that we tested tablets with screens from 7" to 10" wide which are usually made of gorilla glass in its front, and of plastic material in its bottom. Also, we made an assumption that the users' hands are in close contact with touch area as well as its back. Furthermore, the tablets are tested in their typical use, i.e. internet browsing.

The rest of the paper is organized as follows. Section 2 describes the measuring method and equipment. Section 3 defines the experiment. Section 4 presents the results and gives the discussion. Section 5 draws the conclusions.

## II. METHODS AND MATERIALS

### A. Static magnetic filed

The method consists of measuring the extremely low-frequency magnetic field produced by the tablet computers. In normal operation, the tablet's components are supplied by current  $I$ . The current  $I$  is flowing through electronic or electrical components, which induces the magnetic-field. The measuring devices usually register the scalar components of the magnetic induction:  $B_x$ ,  $B_y$  and  $B_z$ . Accordingly, the root mean square (RMS) of magnetic induction  $B$  is calculated as:

$$B = \sqrt{B_x^2 + B_y^2 + B_z^2} \quad (1)$$

### B. Measuring devices

Magnetic field measurement is performed by EMF measuring device Lutron EMF-828 with separate probe, including sensing head. Fig. 1 shows the measuring device Lutron EMF-828 with its probe.



Fig. 1. Measuring device Lutron EMF-828 with separate probe

The calibration of the measurement device is performed according to ISO 9001 by the producer of the equipment. Lutron EMF-828 measures the scalar components of the magnetic induction  $B_x$ ,  $B_y$  and  $B_z$  in the range between 0.01  $\mu$ T and 2 mT. It is measured in the extremely low-frequency range, i.e. between 30 and 300 Hz. The device has three measurement extents: 20  $\mu$ T, 200  $\mu$ T and 2 mT. The precision of the measurement is of the order of 0.01  $\mu$ T for the measurement extent of 20  $\mu$ T, 0.1  $\mu$ T for the measurement extent of 200  $\mu$ T and 1  $\mu$ T for the measurement extent of 2 mT.

### C. Tablet measuring positions

Similarly to [7], the magnetic field is measured at nine different positions at the tablet touchscreen at the top and at nine different positions at the bottom of the tablet. We use measuring in direct contact because it is a typical way of using the tablet. Fig. 2 shows the measurement positions at the top and at the bottom of the tablet computer.



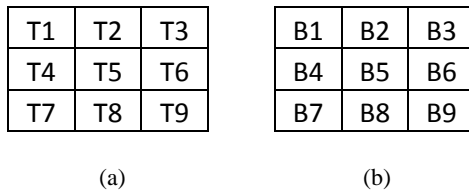


Fig. 2. Measurement positions: (a) at the top, (b) at the bottom

### III. EXPERIMENT

The experiment is conducted on 5 different tablet computer. All tablets are using the Android operating system. The tablets are tested in their typical working condition, i.e. internet browsing. The magnetic field is measured ten times at the measurement positions at the top: T1...T9, and at the bottom: B1...B9. The average values of the measurement are used as a reference. Hence, we measured magnetic field of the tablets at no distance from the touchscreen display or its bottom. We make this assumption taking into account typical use of tablets where users have fingers on tablet top and/or bottom positions all the time. Also, the measurement is carried out 30 cm away from the tablet in all four directions. At these points, the obtained magnetic field was below  $0.02 \mu\text{T}$ . It can be considered as the background magnetic field. Hence, it is negligible.

### IV. RESULTS AND DISCUSSION

The results of the measurement are given in Fig. 3.

Each area is given by the number representing the level of the emitted magnetic field in  $\mu\text{T}$ . As a thresholding (dangerous) level, we used TCO proposed reference level of  $0.2 \mu\text{T}$ .

Table 1 shows the levels of the extremely low frequency magnetic field that are emitted by the top parts of tablets in the min-max manner.

TABLE I  
THE MIN-MAX OF THE TOP MEASURED MAGNETIC FIELD (IN  $\mu\text{T}$ )

Tablet	Top Minimum	Top Maximum
Tablet 1	0.04	0.15
Tablet 2	0.06	0.43
Tablet 3	0.02	0.22
Tablet 4	0.05	0.86
Tablet 5	0.06	0.45

Table 2 shows the levels of the extremely low frequency magnetic field that are emitted by the bottom parts of tablets in the min-max manner.

TABLE II  
THE MIN-MAX OF THE BOTTOM MEASURED MAGNETIC FIELD (IN  $\mu\text{T}$ )

Tablet	Bottom Minimum	Bottom Maximum
Tablet 1	0.05	0.20
Tablet 2	0.07	0.86
Tablet 3	0.01	0.51
Tablet 4	0.05	0.96
Tablet 5	0.07	0.01

From Fig. 3 or from Tables 1-2 (see max levels), it is obvious that all tablets have dangerous areas, which emitted the magnetic field level above the reference one. Still, 3 out of 5 tablets have dangerous levels on the both sides of the tablets, i.e. the top as well as at the bottom. Also, it worth to note that left and right side areas are the most exposed to the tablet's users due to their way of working with tablets.

T1

0.15	0.09	0.04
0.06	0.10	0.06
0.05	0.08	0.04

T2

0.13	0.43	0.09
0.26	0.26	0.13
0.06	0.25	0.05

T3

0.09	0.15	0.13
0.02	0.22	0.22
0.09	0.12	0.06

T4

0.06	0.20	0.86
0.07	0.13	0.06
0.06	0.08	0.05

T5

0.08	0.23	0.09
0.45	0.23	0.07
0.31	0.06	0.07

T1

0.09	0.15	0.05
0.18	0.10	0.06
0.05	0.20	0.06

T2

0.25	0.49	0.13
0.31	0.26	0.07
0.11	0.86	0.08

T3

0.35	0.20	0.06
0.51	0.20	0.01
0.05	0.35	0.46

T4

0.06	0.29	0.96
0.05	0.08	0.07
0.05	0.13	0.07

T5

0.08	0.10	0.07
0.08	0.07	0.10
0.11	0.06	0.07

Fig. 3. Magnetic field measurement level (Top position represents measurement at the top of the tablet (touchscreen display area). Bottom position represents measurement at the bottom areas of the tablet. Colored fields represent the areas with higher emission than it is allowed)

T1	T2	T3	T4	T5																																													
<table><tr><td>0.06</td><td>-0.06</td><td>-0.01</td></tr><tr><td>-0.13</td><td>0.00</td><td>0.00</td></tr><tr><td>0.00</td><td>-0.12</td><td>-0.02</td></tr></table>	0.06	-0.06	-0.01	-0.13	0.00	0.00	0.00	-0.12	-0.02	<table><tr><td>-0.12</td><td>-0.06</td><td>-0.03</td></tr><tr><td>-0.05</td><td>-0.01</td><td>0.06</td></tr><tr><td>-0.05</td><td>-0.60</td><td>-0.04</td></tr></table>	-0.12	-0.06	-0.03	-0.05	-0.01	0.06	-0.05	-0.60	-0.04	<table><tr><td>-0.26</td><td>-0.05</td><td>0.07</td></tr><tr><td>-0.49</td><td>0.02</td><td>0.21</td></tr><tr><td>0.04</td><td>-0.23</td><td>-0.40</td></tr></table>	-0.26	-0.05	0.07	-0.49	0.02	0.21	0.04	-0.23	-0.40	<table><tr><td>0.00</td><td>-0.09</td><td>-0.09</td></tr><tr><td>0.02</td><td>0.05</td><td>-0.01</td></tr><tr><td>0.01</td><td>-0.05</td><td>-0.02</td></tr></table>	0.00	-0.09	-0.09	0.02	0.05	-0.01	0.01	-0.05	-0.02	<table><tr><td>0.00</td><td>0.13</td><td>0.02</td></tr><tr><td>0.37</td><td>0.16</td><td>-0.03</td></tr><tr><td>0.20</td><td>0.00</td><td>0.00</td></tr></table>	0.00	0.13	0.02	0.37	0.16	-0.03	0.20	0.00	0.00
0.06	-0.06	-0.01																																															
-0.13	0.00	0.00																																															
0.00	-0.12	-0.02																																															
-0.12	-0.06	-0.03																																															
-0.05	-0.01	0.06																																															
-0.05	-0.60	-0.04																																															
-0.26	-0.05	0.07																																															
-0.49	0.02	0.21																																															
0.04	-0.23	-0.40																																															
0.00	-0.09	-0.09																																															
0.02	0.05	-0.01																																															
0.01	-0.05	-0.02																																															
0.00	0.13	0.02																																															
0.37	0.16	-0.03																																															
0.20	0.00	0.00																																															
SUM	SUM	SUM	SUM	SUM																																													
<table><tr><td>-0.28</td></tr></table>	-0.28	<table><tr><td>-0.91</td></tr></table>	-0.91	<table><tr><td>-1.09</td></tr></table>	-1.09	<table><tr><td>-0.18</td></tr></table>	-0.18	<table><tr><td>0.84</td></tr></table>	0.84																																								
-0.28																																																	
-0.91																																																	
-1.09																																																	
-0.18																																																	
0.84																																																	

Fig. 4. The difference of the top and bottom measured magnetic field levels (with given sum of all below)

Fig. 4 shows the difference of the measured magnetic field levels between the top and bottom area. The sign – in difference represents a higher level of the magnetic field emission at bottom than at the top areas of the tablets. From Fig. 4, it is an obvious that 4 out of 5 tablets have higher levels of the magnetic field emission at their bottom. It can be concluded that such magnetic field emission is caused by the battery, which is mainly situated at the bottom part of the tablets.

From the given study, it is an easy to conclude that the tablets did not emit the level of the magnetic field like laptops [7]. However, their emission is also above the proposed reference level by TCO. Brief comparison with laptops (battery powered) showed that the level of ELF magnetic field emission is from 0.2  $\mu$ T to 4.5  $\mu$ T at the top and from 0.2  $\mu$ T to 3.5  $\mu$ T at the bottom [7]. Consequently, tablets emitted at the top from 0.02  $\mu$ T to 0.86  $\mu$ T and at the bottom from 0.01  $\mu$ T to 0.96  $\mu$ T. Still, the comparison is not quite fair because the laptops have much more computer processing power than tablets. Also, it is worth to note that tablets have higher levels of emission at the bottom. It can be assumed that the gorilla glass at the top is much more resistant to the magnetic field radiation than the plastic materials at the bottom of the tablets.

Hence, during the use of the tablets some precaution is necessary. Among them, the most important are: (i) Making breaks during the long working hours with tablets, (ii) Using tablet's stand whenever possible, (iii) Using external keyboard whenever possible, (iv) Buying tablets made from better plastic or composite materials, which did not allow high emission of the magnetic field, (v) Using gloves that reduce the level of the magnetic field exposure, and (vi) Do not use the tablets on your knees.

## V. CONCLUSION

The given study presented one of the very first experiments conducted on tablets in order to measure their emission of the magnetic field. Because, the tablets are very spread in the younger population, it is very important to determine the level of the magnetic field that they emit and compare them to the proposed safe limit standards like TCO. The results of the study showed that the level of the magnetic field emission is typically above the safe limits. Hence, some precautions during the tablet use are necessary. At the end, the list of suggestions for the safe work with tablets was proposed.

Future works will be oriented toward exploring the higher number of testing tablets in their normal use or in so-called under stress condition (gaming or similar).

## REFERENCES

- [1] <http://www.cdc.gov/niosh/topics/emf/>
- [2] ICNIRP, 2014. "Guidelines for Limiting Exposure to Electric Fields Induced by Movement of the Human Body in a Static Magnetic Field and by Time-Varying Magnetic Fields Below 1 Hz", Health physics, vol. 106, no. 3, pp. 418-425, 2014.
- [3] "TCO Certified Notebooks 4.0", TCO Development, 2012.
- [4] N. Zoppetti, D. Andreuccetti, C. Bellieni, A. Bogi and I. Pinto, "Evaluation and Characterization of Fetal Exposures to Low Frequency Magnetic Fields Generated by Laptop Computers", Progress in Biophysics and Molecular Biology, vol. 107, no. 3, pp. 456-463, 2011.
- [5] C. V. Bellieni I. Pinto, A. Bogi, N. Zoppetti, D. Andreuccetti and G. Buonocore, "Exposure to Electromagnetic Fields From Laptop Use of "Laptop" Computers", Archives of Environmental and Occupational Health, vol. 67, no. 1, 2012.
- [6] T. Koppel and P. Tint, "Reducing Exposure to Extremely Low Frequency Electromagnetic Fields from Portable Computers", Agronomy Research, vol. 12, no. 3, pp. 863-874, 2014.
- [7] D. Brodic, A. Amelio, "Classification of the Extremely Low Frequency Magnetic Field Radiation Measurement from the Laptop Computers", Measurement Science Review, vol. 15, no. 4, 2015.
- [8] D. Brodic, M. Jevtic, "The Analysis of the Laptop Characteristics with the Impact to the Laptop Magnetic Field Emission", International Scientific Conference "UNITECH 2015", Conference Proceedings, pp. 253-256, Gabrovo, Bulgaria, 2015.
- [9] C. Avendano, A. Mata, C. A. Sanchez Sarmiento, and G. F. Doncel, "Use of Laptop Computers Connected to Internet through Wi-Fi Decreases Human Sperm Motility and Increases Sperm DNA Fragmentation", Fertility and Sterility, vol. 97, no. 1, pp. 39-45, 2012.
- [10] SWEDAC (Swedish Board for Technical Assistance), User's Handbook for Evaluating Visual Display Units, MPR 1990:10, SWEDAC, Bords, Sweden, 1990.
- [11] M. Blank, R. Goodman, "DNA is a Fractal Antenna in Electromagnetic Fields", International Journal of Radiation Biology, vol. 87, no. 4, pp. 409-415, 2011.
- [12] ICNIRP, 2003. "Guidance on Determining Compliance of Exposure to Pulsed and Complex Non-Sinusoidal Waveforms below 100 kHz with ICNIRP Guidelines", Health Physics, vol. 84, no. 3, pp. 383-387, 2003.
- [13] ICNIRP, 2010. "Guidelines for Limiting Exposure to Time-Varying Electric and Magnetic Fields (1 Hz to 100 kHz)", Health Physics, vol. 99, no. 6, pp. 818-836, 2010.
- [14] IEEE: "Biological and Health Effects of Electric and Magnetic Fields from Video Display Terminals", IEEE Engineering in Medicine and Biology Society Committee on Man and Radiation (COMAR), vol. 16, no. 3, pp. 87-92, 1997.
- [15] A. Ahlbom, N. Day, M. Feychting, et al. "A Pooled Analysis of Magnetic Fields and Childhood Leukemia", Br J Cancer. vol. 83, no. 15, pp. 692-698, 2000.
- [16] A. Ahlbom, J. Bridges, R. De Seze, et al. "Possible Effects of Electromagnetic Fields (EMF) on Human Health—Opinion of the Scientific Committee on Emerging and Newly Identified Health Risks (SCENIHR)", Toxicology, vol. 246, no. 2-3, pp. 248-250, 2008.
- [17] Gurney JG, Mueller BA, Davis S, Schwartz SM, Stevens RJ, Kopecky KJ. "Childhood Brain Tumor Occurrence in Relation to Residential Power Line Configurations, Electric Heating Sources, and Electrical Appliance Use", Am J Epidemiol, vol. 143, no. 2, pp. 120-128, 1996.
- [18] Bellieni CV, Acampa M, Maffei M, et al. "Electromagnetic Fields Produced by Incubators Influence Heart Rate Variability in Newborns", Arch Dis Child Fetal Neonatal Ed., vol. 93, no. 4, pp. F298-F301, 2008.
- [19] S. Engström and R. Fitzsimmons, "Five Hypotheses to Examine the Nature of Magnetic Field Transduction in Biological Systems", Bioelectromagnetics vol. 20, no. 7, pp. 423-430, (1999).

# Functional and Timing Analysis of Improved Serial Pseudorandom/Natural Code Converter

Goran Miljković<sup>1</sup>, Dragan Denić<sup>2</sup>, Milan Simić<sup>3</sup> and Aleksandar Jocić<sup>4</sup>

**Abstract** – The pseudorandom/natural code converters are often part of systems and components which exploit pseudorandom binary code, such as absolute pseudorandom position encoders. Serial pseudorandom/natural code converters have advantage in simplicity of hardware implementation, but their conversion time is greater compared to parallel code converters. In the paper is applied Galois generator in the implementation of serial pseudorandom/natural code converter in order to reduce conversion time. One faster way for adjustment logic design which is necessary for proper functioning of Galois based serial code converter is proposed. It is detailed described functionality of serial code converter through simulation examples in NI Multisim software. Also, detailed timing analysis of Fibonacci and Galois based serial pseudorandom/natural code converter is given.

**Keywords** – pseudorandom binary sequence (PRBS), pseudorandom/natural code converter, Fibonacci and Galois architecture for generation of PRBS code

## I. INTRODUCTION

The absolute pseudorandom position encoders are well-known digital transducers which could be used for absolute position measurement in industry, robotics, electrical power engineering, elevators, telescopes/antennas, printers, etc. Absolute position determination in these encoders is based on property of  $n$ -bit pseudorandom binary sequence (PRBS) that sliding window of length  $n$ , which passes along a sequence, will extract unique code word in every moment [1, 2]. Also, the code words are now longitudinally lined, and two successive code words are overlapping and differ only in one bit. The PRBS signal is exploited in other fields, cryptography, telecommunications, testing of VLSI circuits [3], testing of gas sensors [4], measurement of the frequency response, etc. The absolute pseudorandom position encoder consists from following functional parts: code reading system [2, 5], code scanning [2, 6, 7], error detection [8] and pseudorandom/natural code conversion [2, 9, 10].

The pseudorandom/natural code conversion method is

<sup>1</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](mailto:goran.miljkovic@elfak.ni.ac.rs)

<sup>2</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](mailto:dragan.denic@elfak.ni.ac.rs)

<sup>3</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](mailto:milan.simic@elfak.ni.ac.rs)

<sup>4</sup>Aleksandar Jocić is with the Faculty of Electronic Engineering at University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: [aleksandar.jocic@elfak.ni.ac.rs](mailto:aleksandar.jocic@elfak.ni.ac.rs)

compromise between two opposite requests, minimal conversion time and low complexity and price of used hardware. The different pseudorandom/natural code conversion methods can be sorted to three distinct groups: parallel [7], serial [2, 9] and serial-parallel [7]. The parallel pseudorandom/natural code conversion is the fastest, but requests large ROM memory for storing of code conversion table especially in the case of high resolution encoder. The serial pseudorandom/natural code conversion is simplest, less hardware expensive, but slower method. The compromise solution is the serial-parallel code converter, which combines previous two methods.

In the paper are detailed described improved solution of serial pseudorandom/natural code converter based on Galois generator of PRBS which can be used in absolute pseudorandom position encoder or in other fields where PRBS is applied. The presented serial pseudorandom/natural code converter is then implemented in software NI Multisim, integrated capture and simulation design environment. These simulations are used for functional and timing analysis of serial pseudorandom/natural code converters.

## II. IMPROVED SERIAL PSEUDORANDOM/NATURAL CODE CONVERTER

Duration of pseudorandom/natural code conversion process, the conversion time, becomes critical at high resolution absolute pseudorandom position encoders. The serial pseudorandom/natural code converters are based on the fact that is possible to find the actual value of the position  $p$  by counting of the steps that shift register with reverse feedback needs until it reaches the initial state by successive shifting from the read pseudorandom  $n$ -bit code word. One solution is the serial pseudorandom/natural code converter based on Fibonacci generator [9]. Improved solution is based on faster Galois generator of PRBS which can be used in serial code converter instead of Fibonacci generator [9]. This generator also uses 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 improved solution of serial pseudorandom/natural code converter based on Galois generator can be seen in Fig. 1. A 31-bits long PRBS generated by a 5-bit shift register with the feedback set for direct generation law [5,4,3,1] is used for the encoding of the code track. It consists of the following bits: 1110000110101001000101111101100. The read pseudorandom code bits with one code reading head  $x_5$  are loaded into five-cell bidirectional code assembly register. The actual content of the code assembly register corresponding to the current position is loaded into a shift register of Galois generator.

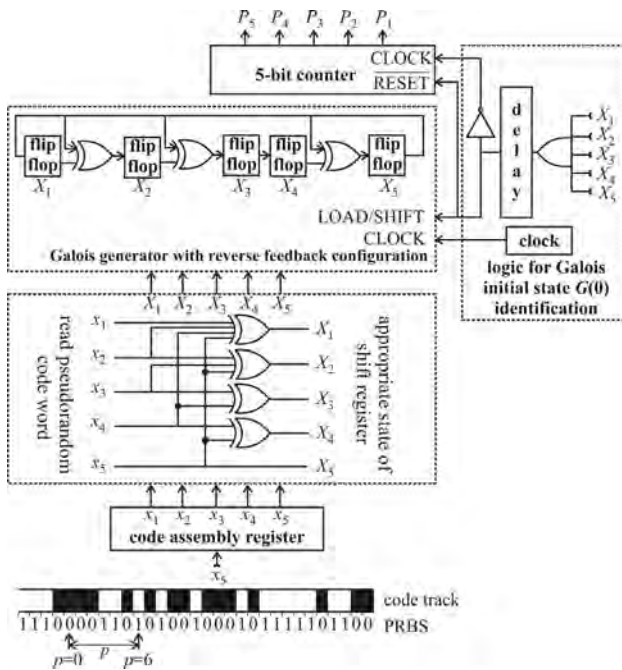


Fig. 1. Serial pseudorandom/natural code converter based on Galois generator

This solution is faster because it has reduced number of logic gates (XOR gates) in the feedback loop. The position of XOR gates between flip flops in the Galois shift register is determined by the used feedback set [5,4,2,1] for reverse generation law. Adequate feedback sets for different resolutions of PRBS sequence can be found in the literature [10]. Now, one XOR gate exists in the feedback configuration compared to three serial connected XOR gates in the case of Fibonacci generator. But, the additional logic that read code word converts to the appropriate content of the shift register is needed for proper functioning of this converter. This logic only participates on the beginning of each code conversion cycle and does not further participate in the code conversion process. The 5-bit counter and additional logic for Galois initial state  $G(0)$  identification are also needed in this serial pseudorandom/natural code converter.

Logic for initial adjustment of read pseudorandom code word is needed because the read  $n$ -bit pseudorandom code word in real time is not identical to the  $n$ -bit current content of the shift register, which was the case at Fibonacci generator. Procedure of this adjustment logic design is detailed explained in the reference [9] for  $n=5$  bit resolution and chosen feedback set. Following this procedure for logic design in the case of resolution  $n=5$  and feedback set for Galois generator [5, 4, 2, 1], the following set of equations is obtained:

$$X_5 = x_5 \quad (1)$$

$$X_4 = x_4 \oplus x_5 \quad (2)$$

$$X_3 = x_3 \oplus x_4 \quad (3)$$

$$X_2 = x_2 \oplus x_3 \oplus x_5 \quad (4)$$

$$X_1 = x_1 \oplus x_2 \oplus x_4 \oplus x_5 \quad (5)$$

Based on the previous logic equations (1) to (5) the following matrix can be written:

	$x_5$	$x_4$	$x_3$	$x_2$	$x_1$
$X_5$	1	0	0	0	0
$X_4$	1	1	0	0	0
$X_3$	0	1	1	0	0
$X_2$	1	0	1	1	0
$X_1$	1	1	0	1	1

where columns are in order  $\{x_5, x_4, x_3, x_2, x_1\}$ , and rows in order  $\{X_5, X_4, X_3, X_2, X_1\}$ . It can be seen that for coefficients [5, 4, 2, 1] in rows, the adequate diagonals are filled with "1" beginning from defined coefficients. It is highlighted diagonal filled with "1" for coefficient 5 in the feedback set. So, for any feedback set of coefficients this matrix can be easily written, and then from there adequate equations for adjustment logic can be easily obtained.

This procedure is checked for different resolutions from 3 to 16 bit, and various possible feedback sets [10]. Based on previous considerations it is confirmed that exist the clear rule for matrix filling for any chosen resolution and the feedback set. The numbers in feedback set defines adequate starting position in rows,  $X_1, \dots, X_n$ , from which starts with filling of diagonals with "1". Filled matrix directly gives set of logic equations which is needed for design of adequate logic for initial adjustment of read pseudorandom code word. This logic directly converts current content of Fibonacci shift register to adequate content of Galois shift register in adequate PRBS generator.

### III. SIMULATIONS AND ANALYSIS OF SERIAL CONVERTERS

For purpose of detailed functional and timing analysis of presented solution of serial pseudorandom/natural code converter, simulation of hardware realization is performed in software environment NI Multisim 11.0, Fig. 2. In Multisim software can be found several useful tools and parts for digital circuits. They range from digital gates (both families, TTL and CMOS are supported), indicators (such as LED's, LCD displays, Logic Analyzers and Word Generators), wires, data buses and power supplies. The most widely used subfamily for digital circuit simulation is the 74LSXX series where the digits after the LS indicate the gate's function. This subfamily is used for simulation of presented serial pseudorandom/natural code converters, and applied logic gates in the simulations with their propagation delays are listed in Table 1.

Table 1. Logic gates and adequate propagations delays

Logic gates	Used logic gates from 74LSXX series	Propagation delays
AND	74LS08D	15 ns
OR	74LS32D	22 ns
INV	74LS04D	15 ns
XOR	74LS86D	30 ns
D flip flop	74LS74D	$t_s=20$ ns $t_h=5$ ns $t_{CLK-Q}=25$ ns
AND with 4 inputs	74LS21D	15 ns



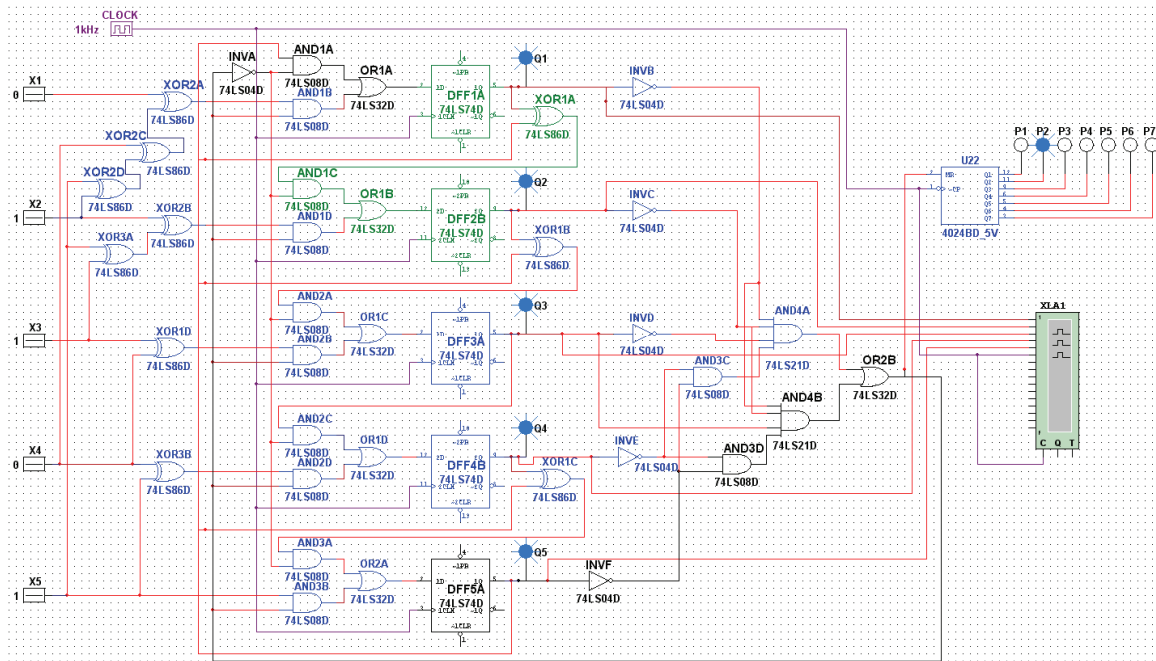


Fig. 2. Simulation of serial pseudorandom/natural code converter based on Galois generator in Multisim

The conversion process is executed for concrete example of read pseudorandom code word and for defined initial code word. 7-bit binary counter is used for counting of needed steps to shift register for finishing of one code conversion cycle. The read 5-bit code word  $x = x_1x_2x_3x_4x_5$  is loaded in the 5 D flip flops on the beginning of the code conversion process. It is used example where read pseudorandom code word is  $x = \{0,1,1,0,1\}$ , and initial code word is  $X(0) = \{1,1,1,0,0\}$ . It is needed six clock periods to the shift register when its state becomes equal to the initial state, which is output of the binary counter in the case of Fibonacci generator. In the case of Galois based code converter, Fig. 2, the read pseudorandom code word  $x = \{0,1,1,0,1\}$  is firstly converted to appropriate state of the Galois shift register  $X_{appr.state} = \{0,1,1,1,1\}$  based on equations (1)-(5), and then is also needed six clock periods to the shift register when its state becomes equal to Galois initial state  $G(0) = \{0,0,1,0,0\}$ . Galois initial state is obtained from initial state  $X(0) = \{1,1,1,0,0\}$  using equations (1)-(5). It is proved proper functioning of proposed hardware realization for different examples of read and initial pseudorandom code words. States of the shift register in Fibonacci and Galois generator in each clock period during shifting from read pseudorandom code word to the initial code word of presented example are listed in Table 2. In the simulations of serial converter, Fig. 2, is also used AND gate with 4 inputs together with one AND gate for initial state identification.

To find the maximal working frequency of the whole digital circuit, which is the most significant difference between two presented solutions, it is needed to analyze propagation delays of each path. Combinational propagation delays are additive. It is possible to determine the propagation delay of a larger combinational circuit by adding the propagation delays of the circuit components along the longest path. Frequency must be determined by locating the longest path among all the flip-flop paths in the circuit.

Table 2. Shift register content in Fibonacci and Galois generator

Output of the binary counter	Fibonacci generator shift register content: $\{X_1X_2X_3X_4X_5\}$	Galois generator shift register content: $\{X_1X_2X_3X_4X_5\}$
Read pseudorandom code word, P=0	01101	01111
P=1	00110	11010
P=2	00011	01101
P=3	00001	11011
P=4	10000	10000
P=5	11000	01000
Initial code word, P=6	11100	00100

Propagation delays along longest paths in both simulations are (Fig 2.):

- for Fibonacci based converter (from the output of the fifth flip flop DFF5A, through three XOR gates in the feedback loop, then AND1A, OR1A, to the input of the first flip flop DFF1A):

$$T_1 = t_{CLK-Q}(5) + 3 t_{pd}(XOR) + t_{pd}(AND) + t_{pd}(OR) + t_s(1) = 25 \text{ ns} + 3 \times 30 \text{ ns} + 15 \text{ ns} + 22 \text{ ns} + 20 \text{ ns} = 172 \text{ ns},$$

- for Galois based converter (from the output of the fifth flip flop DFF5A, through logic for initial state identification INVF, AND3D, AND4B, OR2B, INVA, then AND1A, OR1A to the input of the first flip flop DFF1A):

$$T_2 = t_{CLK-Q}(5) + t_{pd}(INV) + 2 t_{pd}(AND) + t_{pd}(OR) + t_{pd}(INV) + t_{pd}(AND) + t_{pd}(OR) + t_s(1) = 25 \text{ ns} + 15 \text{ ns} + 2 \times 15 \text{ ns} + 22 \text{ ns} + 15 \text{ ns} + 15 \text{ ns} + 22 \text{ ns} + 20 \text{ ns} = 164 \text{ ns}.$$

So, maximal working frequency of Fibonacci based converter would be,

$$f_{\max\text{-fibonacci}} = 1/T_1 = 5.81\text{MHz}. \quad (6)$$

- for Galois based converter (from the output of the first flip flop DFF1A, through one XOR gates in the feedback loop XOR1A to the input of the second flip flop DFF2B):

$$T_3 = t_{\text{CLK-Q}}(5) + t_{\text{pd}}(\text{XOR}) + t_{\text{pd}}(\text{AND}) + t_{\text{pd}}(\text{OR}) + t_s(1) = 25 \text{ ns} + 30\text{ns} + 15 \text{ ns} + 22 \text{ ns} + 20 \text{ ns} = 112 \text{ ns}$$

-  $T_2$  is the same as in the case of the Fibonacci based converter, so maximal working frequency of Galois based converter would be,

$$f_{\max\text{-galois}} = 1/T_2 = 6.09\text{MHz} \quad (7)$$

From previous timing analysis, it can be concluded that Galois generator is faster than Fibonacci. The pseudorandom/natural code converter based on Galois generator has additional propagation delay due to the logic for initial conversion of read pseudorandom code word which is equal to propagation delay of three serial connected XOR gates (XOR2A, XOR2C, and XOR2D). Also, the contamination delays along each path are greater than or equal to the destination flip flop hold time, so the circuit will operate as designed.

For presented example of 5-bit pseudorandom binary sequence, 31 unique pseudorandom code words define absolute position. So, for the worst case, if the code converter needs 30 clock periods to convert the read pseudorandom code word in relation to the initial pseudorandom code word which present zero position, maximal duration of one conversion cycle for Fibonacci and Galois generator would be,

$$T_{\text{conv-fibonacci}} = 30 \times 1 / f_{\max\text{-fibonacci}} = 5.16\mu\text{s} \quad (8)$$

$$T_{\text{conv-galois}} = 3 \times t_{\text{pd}}(\text{XOR}) + 30 \times 1 / f_{\max\text{-galois}} = 5.01\mu\text{s} \quad (9)$$

So, the serial pseudorandom/natural code converter based on Galois generator is relatively faster than Fibonacci based for,

$$\delta_{\%} = \frac{T_{\text{conv-fibonacci}} - T_{\text{conv-galois}}}{T_{\text{conv-fibonacci}}} 100\% = \frac{5.16 - 5.01}{5.16} 100\% = 2.9\% \quad (17)$$

Duration of code conversion time depends on used resolution of PRBS and used feedback set (for same resolution different feedback sets exist, with different number of coefficients).

## IV. CONCLUSION

Improved, Galois based implementation of serial pseudorandom /natural code converters is presented and analyzed. It is proved reducing of conversion time at Galois based serial converter, which is its main advantage, by using detailed timing analysis. It is examined full functionality of both, Fibonacci and Galois based serial converters through simulation in NI Multisim software. The solution for easier designing of necessary logic for initial adjustment of read pseudorandom code word at Galois based converter is also given.

## ACKNOWLEDGEMENT

Research activities presented in this paper, are supported by funds of the Ministry of Science and Technological Development of Republic of Serbia, having the reference project number TR32045.

## REFERENCES

- [1] F.J. MacWilliams, N.J.A Slone, "Pseudo-random sequences and arrays", *Proceeding of IEEE*, Vol. 64, No. 12, pp. 1715-1728, 1976.
- [2] E. M. Petriu, J. S. Basran, "On the position measurement of automated guided vehicles using pseudorandom encoding", in *IEEE Trans. IM*, vol. 38, no. 3, pp. 799-803, 1989.
- [3] J.C. Rau, P.H. Wu, Y.F. Ho, "A novel reseeding mechanism for improving pseudo-random testing of VLSI circuits", in *Tamkang Journal of Science and Engineering* vol. 11, no. 2, pp. 175-184, 2008.
- [4] M.E.H. Amrani, R.M. Dowdeswell, P.A. Payne, K.C. Persaud, "Pseudo-random binary sequence interrogation techniques for gas sensors", in *Sensor. Actuat. B-Chem.* vol. 47, pp. 118-124, 1998.
- [5] M. Arsić, D. Denić, "New pseudorandom code reading method applied to position encoders", in *Electron. Lett.* vol. 29, pp. 893-894, 1993.
- [6] D. Denić, G. Miljković, "Code reading synchronization method for pseudorandom position encoders", in *Sensor. Actuat. A-Phys.* vol. 150, pp. 188-191, 2009.
- [7] E.M. Petriu, J.S. Basran, F.C.A. Groen, "Automated guided vehicle position recovery", in *IEEE Trans. Instrum. Meas.* vol. 39, pp. 254-258, 1990.
- [8] D. Denić, M. Arsić, "Checking of pseudorandom code reading correctness", in *Electron. Lett.* vol. 29, pp. 1843-1844, 1993.
- [9] D. Denić, I. Stojković, "Pseudorandom/natural code converter with parallel feedback logic configuration", in *Electron. Lett.* vol. 46, pp. 921-922, 2010.
- [10] New Wave Instruments, "Linear feedback shift registers: Implementation, M-sequence properties, feedback tables", (2004), [http://www.newwaveinstruments.com/resources/articles/m\\_sequence\\_linear\\_feedback\\_shift\\_register\\_lfsr.htm](http://www.newwaveinstruments.com/resources/articles/m_sequence_linear_feedback_shift_register_lfsr.htm)

# Review and Contribution to the Best Measurement and Calibration Capabilities in the Field of Electrical Energy and Power

Marija Cundeva-Blajer<sup>1</sup> and Bojan Iliev<sup>2</sup>

**Abstract** – In this paper a review of the best measurement and calibration capabilities at international, regional and national level in the field of electrical energy and power will be made. The contribution of the Laboratory for Electrical Measurements at the Faculty of Electrical Engineering-Skopje to the enhancement of the national and regional electrical energy and power metrology infrastructure will be presented.

**Keywords** – electrical energy measurement, power measurement, metrology infrastructure, best measurement and calibration capability, measurement uncertainty.

## I. INTRODUCTION

The electrical energy and power measurements represent a very important segment in the trade of electrical energy, and they are in the scope of the legal metrology, [1, 2]. The insurance of measurement's accuracy is essential for safe and fair electrical energy delivery and trade. In R. Macedonia the coverage of this significant metrology segment is on a relatively basic level in comparison to the international state of the art, [2]. For the purpose of compliance to the international standards in this area, currently a procedure for establishment of national standard for electrical energy and power is in phase of implementation, [3, 4]. The national standard will enable calibration of secondary power and energy standards with accuracy up to class 0,01.

In this paper a comparison and review of the current state of the art at three levels, international, regional and national, of the best measurement and calibration capabilities will be conducted. This survey is necessary for further planning and development of the metrology infrastructure, especially at national and regional level, in the field of electrical energy and power as well as for an official proclamation of national standard for electrical energy and power, [4, 5]. The analysis can be conducted for different physical quantity ranges, but the main accent in this contribution will be on the best measurement and calibration capabilities of single phase active, reactive and apparent electrical power up to 400 Hz frequency range, [5]. The Laboratory of Electrical Measurements (LEM) at the Faculty of Electrical Engineering

and Information Technologies (FEIT) at the Ss. Cyril and Methodius University in Skopje comprises the laboratory for calibration of electromagnetic quantities and owns a reference standard for electrical energy and power ZERA COM3003 with accuracy class 0,01. The LEM is an already accredited calibration laboratory for electromagnetic quantities, including a certain scope of electrical energy and power, according to the international standard ISO 17205:2005, [6], by the Institute of Accreditation of R. Macedonia (IARM).

In this paper the methodology for introduction of the ZERA COM3003 standard, [7], as potential national standard in R. Macedonia (the top of the metrology pyramid in the field of electrical energy and power), through the chain of hierarchy calibrations of secondary standards of energy and power will be elaborated. This will contribute to the development of the metrology infrastructure in the field of electrical energy and power at national as well as at regional level.

## II. SURVEY OF THE CURRENT STATE OF THE ART

The survey of the current state of the art in the field of metrology of electrical energy and power is conducted at three levels: national, regional (South-East Europe) and international. The main source is the official database of the International Bureau of Weights and Measures (BIPM) of the calibration and measurement capabilities (CMC) [5]. The coverage factor of the expanded measurement uncertainty is  $k=2$  for level of confidence 95%.

### A. Best Measurement and Calibration Capabilities-International Level

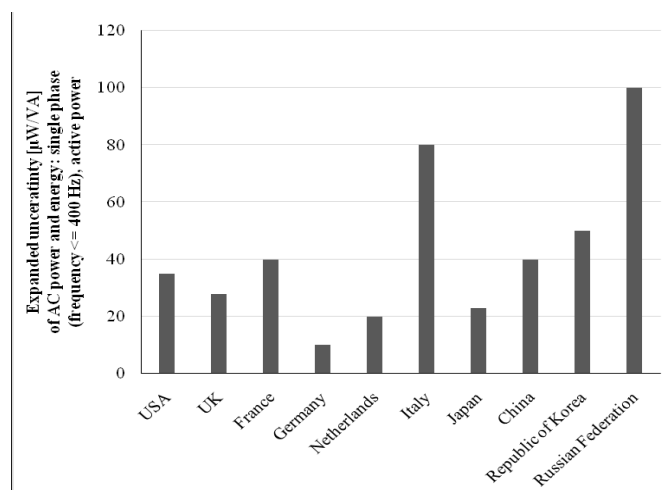


Fig. 1. Comparison of the best CMCs of the largest NMIs of active electrical power for frequencies up to 400 Hz.

<sup>1</sup>Marija Cundeva-Blajer is with the Faculty of Electrical Engineering and Information Technologies-Skopje at the Ss. Cyril and Methodius University in Skopje, ul. Ruger Boskovic br. 18, POB 574, 1000 Skopje, R. Macedonia, E-mail: mcundeva@feit.ukim.edu.mk.

<sup>2</sup>Bojan Iliev is with Elko Ing DOOEL-Strumica, ul. Goce Delcev br. Strumica, R. Macedonia, E-mail: bojaniliev83@yahoo.com

**TABLE I**

COMPARISON OF THE BEST MEASUREMENT AND CALIBRATION CAPABILITIES FOR AC POWER AND ENERGY: SINGLE PHASE (FREQUENCY  $\leq 400$  Hz), ACTIVE POWER OF THE NATIONAL METROLOGY INSTITUTES AT INTERNATIONAL LEVEL

	Measurand Level or Range				Measurement Conditions/Independent Variable		Expanded Uncertainty			
	Method of Measurement	Minimum value	Maximum value	Units	Parameter	Specifications	Value	Units	Coverage Factor	Level of Confidence
<b>USA</b> NIST (National Institute of Standards and Technology)	Comparison	0	60	kW	Voltage	60 V to 600 V	35	$\mu\text{W}/\text{VA}$	2	95%
					Current	0.1 A to 100				
					Phase shift	all				
					Frequency	50 Hz to 400				
<b>UK</b> NPL (National Physical Laboratory)	Digital sampling	0	130	kW	Voltage	1 V to 1000 V	28-40	$\mu\text{W}/\text{VA}$	2	95%
					Current	2 mA to 100				
					Phase shift	1 to 0, inductive or capacitive				
					Frequency	50 Hz to 400				
<b>France</b> LNE (Laboratoire national de métrologie et d'essais)	Direct comparison	0	60	kW	Voltage	60 V to 600 V	40 to 70	$\mu\text{W}/\text{VA}$	2	95%
					Current	0.05 A to 100				
					Phase shift	1 to 0, inductive or capacitive				
					Frequency	45 Hz to 65				
<b>Germany</b> PTB (Physikalisch-Technische Bundesanstalt)	Direct comparison	0	76,8	kW	Voltage	30 V to 480 V	10 to 140	$\mu\text{W}/\text{VA}$	2	95%
					Current	0.005 A to 160				
					Phase shift	-180 to 180				
					Frequency	16.7 Hz to 400				
<b>Netherlands</b> VSL	Direct comparison	0	1,2	kW	Voltage	120 V and 240 V	20	$\mu\text{W}/\text{VA}$	2	95%
					Current	1 A and 5 A				
					Phase shift	1 to 0, inductive or capacitive				
					Frequency	45 Hz to 65 Hz				
<b>Italy</b> INRIM (Istituto Nazionale di Ricerca Metrologica)	Comparison with reference standard	0	30	kW	Voltage	1 V to 600 V	80	$\mu\text{W}/\text{VA}$	2	95%
					Current	0.05 A to 50 A				
					Phase shift	1 to 0, inductive or capacitive				
					Frequency	50 Hz, 60 Hz				
<b>Japan</b> NMIJ (National Metrology Institute of Japan) and JEMIC	Direct comparison	0	0,6	kW	Voltage	100 V, 110 V, 120 V	23 to 28	$\mu\text{W}/\text{VA}$	2	95%
					Current	5 A				
					Phase shift	1 to 0 L/C				
					Frequency	50 Hz, 60 Hz				
<b>China</b> NIM (National Institute of metrology)	Direct comparison	0	40	kW	Voltage	50 V to 400 V	40	$\mu\text{W}/\text{VA}$	2	95%
					Current	0.5 A to 100 A				
					Phase shift	0.0 lead and lag, 0.5 lead and lag, 0.866 lead and lag, 1.0				
					Frequency	45 Hz to 65 Hz				
<b>South Korea</b> KRISS (Korea Research Institute of Standards and Science)	Direct comparison	1,5	12	kW	Voltage	60 V to 240 V	50 to 100	$\mu\text{W}/\text{VA}$	2	95%
					Current	/				
					Phase shift	0.25 to 1				
					Frequency	45 Hz to 65				
<b>Russian Federation</b> VNIIM (D.I. Mendeleyev Institute for Metrology, Rosstandard)	Direct comparison	0,25	20	kW	Voltage	50 V to 400 V	100	$\mu\text{W}/\text{VA}$	2	95%
					Current	0.01 A to 50				
					Phase shift	1.0 i/c to 0.5 i/c				
					Frequency	45 Hz to 55 Hz				



A comparison of the best measurement and calibration capabilities of the largest national metrology institutes (NMIs) in the field of single phase electrical power up to 400 Hz frequency is made, [5]. The voltage ranges are from 0 to 200 kV, electrical current from 0 to 2000 A, the power factor  $\cos\varphi$  is from 0 to 1 capacitive and inductive load. The comparison is displayed in Table 1.

The lowest measurement uncertainty of  $\pm 10 \mu\text{W/VA}$  of active electrical power at  $\cos\varphi=1$  is declared by PTB-Germany. However, the highest uncertainty among the largest NMIs is of  $\pm 100 \mu\text{W/VA}$ , as displayed in Figure 1.

Additionally, the lowest measurement uncertainty of  $\pm 20 \mu\text{W/VA}$  of electrical power at  $\cos\varphi=0$  (reactive power) is declared by VSL-Netherlands.

### *B. Best Measurement and Calibration Capabilities-Regional Level in South-East Europe*

In some of the countries in South-East Europe (SEE) the measurements and calibrations in the field of electrical energy and power are present, [5], because they a significant pillar in the legal metrology system.

The metrology infrastructure, in the field of energy and power, demands allocation of high and expensive resources (facilities, equipment, development of measurement methods and personnel training), [2, 4, 6]. Therefore, the investments and the development plans in this field must be optimized according to the national as well as regional needs of the economy.

TABLE II

COMPARISON OF THE BEST MEASUREMENT AND CALIBRATION CAPABILITIES FOR AC POWER AND ENERGY: SINGLE PHASE (FREQUENCY  $\leq 400$  Hz), ACTIVE POWER OF THE NATIONAL METROLOGY INSTITUTES IN SOUTH-EAST EUROPE

	Measurand Level or Range				Measurement Conditions/Independent Variable		Expanded Uncertainty			
	Method of Measurement	Minimum value	Maximum value	Units	Parameter	Specifications	Value	Units	Coverage Factor	Level of Confidence
<b>Slovenia</b> MIRS/SIQ/ Metrology (MIRS/Slovenian Institute of Quality and Metrology)	Direct voltage sampling	0	35	kW	Voltage	0.1 V to 700 V	25 to 800	$\mu\text{W/VA}$	2	95%
					Current	1 mA to 50 A				
					Phase shift	1 to 0, inductive or capacitive				
					Frequency	45 Hz to 65 Hz				
<b>Serbia</b> DMDM (Directorate of Measures and Precious Metals)	Direct comparison or sampling method	0	48	kW	Voltage	4 V to 700 V	30 to 129	$\mu\text{W/VA}$	2	95%
					Current	0.05 A to 100 A				
					Phase shift	1 to 0, inductive or capacitive				
					Frequency	16 Hz to 400 Hz				
<b>Turkey</b> UME (TÜBİTAK Ulusal Metroloji Enstitüsü)	Direct comparison	0	60	kW	Voltage	30 V to 500 V	100	$\mu\text{W/VA}$	2	95%
					Current	0.01 A to 120				
					Phase shift	1 to 0, inductive or capacitive				
					Frequency	45 Hz to 65				
<b>Bulgaria</b> BIM (Bulgarian Institute of Metrology)	Direct comparison	0	48	kW	Voltage	10 V to 480 V	16 to 150	$\mu\text{W/VA}$	2	95%
					Current	0.002 A to 100 A				
					Phase shift	1 to 0, inductive or capacitive				
					Frequency	46 Hz to 65				
<b>Greece</b> EIM (Hellenic Institute of Metrology)	Direct comparison	0,075	1,2	kW	Voltage	60 V to 240 V	150	$\mu\text{W/VA}$	2	95%
					Current	5 mA to 5 A				
					Phase shift	1 to 0.25, inductive or capacitive				
					Frequency	53 Hz				
<b>Romania</b> INM (National Institute of Metrology)	Direct comparison	0,15	57,6	kW	Voltage	60 V to 480 V	100 to 380	$\mu\text{W/VA}$	2	95%
					Current	0.005 A to 120 A				
					Phase shift	0.5 i to 1 to 0.5 c				
					Frequency	50 Hz				

So, for the purposes of creation of proper and optimal development strategy, a survey of the current state of the art in the area of measurements and calibrations of electrical energy and power in most of the SEE countries, is conducted, [5]. A comparison of the best CMCs of the SEE national metrology

institutes (NMIs) in the field of single phase electrical power up to 400 Hz frequency is made, [5]. The voltage ranges are from 0 to 700 V, electrical current from 0 to 120 A, the power factor  $\cos\varphi$  is from 0 до 1 capacitive and inductive load. The comparison is shown in Table 2. The lowest measurement

uncertainty of  $\pm 16 \mu\text{W/VA}$  of active electrical power at  $\cos\varphi=1$ , is declared by BIM-Bulgaria. The highest uncertainty among the SEE NMIs is  $\pm 150 \mu\text{W/VA}$ , as shown in Figure 2.

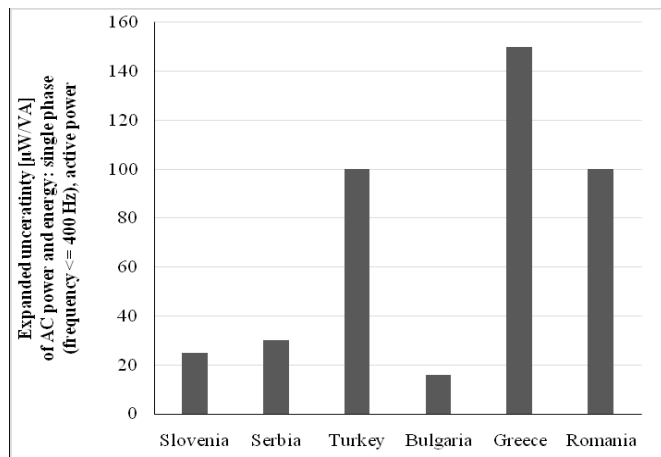


Fig. 2. Comparison of the best CMCs of the SEE NMIs of active electrical power for frequencies up to 400 Hz.

The other countries in the South-East Europe (Croatia, Bosnia and Herzegovina, Montenegro, Macedonia and Albania) have no CMC in the BIPM database in the field of AC electrical power and energy, [5]. It can be concluded that in the region of South-East Europe there are opportunities for further development of the metrology infrastructure in the field of electrical energy and power. However, this development must be planned and well adjusted to the regional and national measurement and calibration needs.

### III. CONTRIBUTION TO THE DEVELOPMENT OF NATIONAL METROLOGY SYSTEM FOR ELECTRICAL ENERGY AND POWER IN R. MACEDONIA

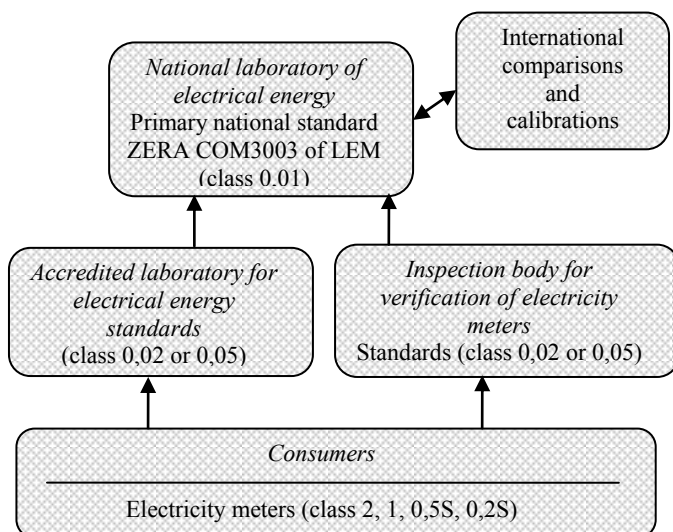


Fig. 3. Proposal-scheme of national metrology infrastructure in electrical energy and power in R. Macedonia.

The current state and the importance of the measurements of electrical energy for billing and other purposes requires

measures for upgrading of the Macedonian system of legal and industrial metrology for electrical energy. The procedure for establishment of a national standard of electrical energy and national laboratory for electrical energy, as well as traceability of the electrical energy measurements has already started. Namely, the Laboratory for Electrical Measurements (LEM) at FEIT has accomplished the accreditation in compliance to the international standard ISO 17025, [6] according to the legal conditions, of [3, 4]. The proposed scheme for practicing the legal metrology and establishing of the traceability chain in the measurements of electrical energy in R. Macedonia is given in Figure 3. The LEM owns a reference standard for electrical energy and power ZERA COM3003 with accuracy class 0,01, [7]. LEM is accredited with a scope in electrical energy and power as in Table III.

TABLE III  
SCOPE OF ACCREDITATION OF LEM IN ELECTRICAL ENERGY AND POWER

	Ranges	Rel. exp. uncertainty at $k=2, P=95\%$
AC Power	Voltage: 70, 140, 280, 560 V	$\pm 0,2\%$
	Current: 0,5, 6, 20, 120 A	$\pm 0,2\%$
	Power	$\pm 0,4\%$
AC Energy	Voltage: 70, 140, 280, 560 V	$\pm 0,2\%$
	Current: 0,5, 6, 20, 120 A,	$\pm 0,2\%$
	Time	$\pm 0,05\%$
	Energy	$\pm 0,4\%$

### IV. CONCLUSION

In this paper a survey of the current state of the art in the field of metrology of electrical energy and power at international, regional and national level is made as a basis for metrology infrastructure development planning. The proclamation of the LEM's ZERA COM3003 electrical energy comparator as a national standard will significantly fill the gap in the electrical energy/power legal metrology system and will upgrade the regional metrology landscape as well.

### REFERENCES

- [1] Klaus-Dieter Sommer, Stefan Sarge, "Energy measurement and standards – the contemporary challenge", 2<sup>nd</sup> IMEKO TC 11 International Symposium Metrological Infrastructure, Cavtat, Dubrovnik, Croatia, June 15-17, 2011.
- [2] L. Arsov, M. Cundeva-Blajer, "Establishing a Metrological Infrastructure and Traceability of Electrical Power and Energy in the R. Macedonia", ACTA IMEKO, Volume 2, Number 2, pp. 86-90, [www.imeko.org](http://www.imeko.org), December 2013.
- [3] Law on Metrology, Official Gazette of R. Macedonia No. 55/02, 84/07 and 120/09.
- [4] Rulebook of the Conditions for Realization, Storage and Maintaining of National Standards and the Conditions for Proclamation of the Reference Standards as National Standards, Official Gazette of R. Macedonia, No. 28/2009.
- [5] [www.bipm.org](http://www.bipm.org) (last access 30.03.2016)
- [6] EN ISO/IEC 17025 General requirements for the competence of testing and calibration laboratories, Cenelec, Brussels, 2005.
- [7] ZERA, "Operation manual—Three Phase Comparator COM 3003", ZERA GmbH – 53639 Koenigswinter, Germany, 2012.

# Analysis of the noise parameters of quad structure of MEMS linear accelerometers

Emil Iontchev<sup>1</sup>, Rosen Miletiev<sup>2</sup>, Vladimir Bashev<sup>3</sup> and Rumen Yordanov<sup>4</sup>

**Abstract** – The current paper discusses the analysis of the noise parameters of quad sensor structure of linear accelerometers in comparison to the same parameters of the single sensors. It is shown that the digital signal processing allows decreasing the main noise parameters (random walk, quantization noise and bias instability) of the quad sensor structure in comparison to the same parameters of the single sensors. This allows increasing the accuracy of the inertial navigation systems which are integrated to obtain the vehicle's position, velocity and attitude.

**Keywords** – Quad sensor structure, MEMS, Allan variance

## INTRODUCTION

Velocity random walk is the additive white noise component on the MEMS sensor output. The sensor noises, bias variations, drifts and scale factor instabilities are the main error sources which cause the accuracy degradation of the modern navigation systems especially based on low cost MEMS inertial sensors. The power spectrum density (PSD) is one of the commonly used for the investigation of sensors' stochastic process. PSD is a frequency-domain approach for modeling noise. The Allan variance (AV) method also may be used to determine the characteristics of the underlying random processes that give rise to the data noise. The Allan variance is a method of representing the root mean square (RMS) random-drift errors as a function of averaging times. Because of the close analogies to inertial sensors, this method has been adapted to random-drift characterization of a variety of devices. The AV method is straightforward and simple for calculation in comparison with other stochastic modeling methods, such as the autocorrelation method, which needs very long term static data. In this paper, this technique is used to characterize various types of noise terms in different inertial-sensor data.

The Allan variance is a time-domain-analysis technique originally developed in the mid-1960s to study the frequency

<sup>1</sup>Emil Iontchev is with the Higher School of Transport "T. Kableshkov" 158 Geo Milev Street, Sofia 1574, Bulgaria, E-mail: [e\\_iontchev@yahoo.com](mailto:e_iontchev@yahoo.com)

<sup>2</sup>Rosen Miletiev is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria. E-mail: [miletiev@tu-sofia.bg](mailto:miletiev@tu-sofia.bg)

<sup>3</sup>Vladimir Bashev is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria. E-mail: [v.bashev@gmail.com](mailto:v.bashev@gmail.com)

<sup>4</sup>Rumen Yordanov is with the Faculty of Electronic Engineering and Technology at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria. E-mail: [rsyordanov@yahoo.com](mailto:rsyordanov@yahoo.com)

stability of precision oscillators [1]. Due of the close analogies to inertial sensors, this method has been adapted to random-drift characterization of a variety of devices [2]–[5]. The most attractive feature of AV is its ability to sort out various noise components by the slopes on the root AV (Allan deviation) plot provided the different noise mechanisms are reasonably separated in the frequency and time domain [6]–[7].

The current paper discusses the noise parameters of the quad sensor structure, build by four identical digital output inertial sensors with synchronization inputs and microcontroller to compare the error values according to the single sensor.

## SYSTEM DESCRIPTION

The inertial measurement systems consists of four identical MEMS inertial sensors MPU-6000 [8], which are placed in such way that each two sensors have one common axis and two opposite axes (Figure 1) [9]. One of the most important special feature of the selected MEMS sensors is the FSYNC (Frame synchronization) input, which may be used to start the measurement process. The system synchronization is carried out by the multiple output clock generator. It is based on the Ultra-low Jitter LVCMOS Fanout Buffer/Level Translator LMK00105 [10]. The low noise VCMOS fanout buffer may distribute 5 ultra low jitter clocks from a differential, single ended or crystal input. The four of these 5 clocks are used in the quad sensor structure and the fifth one is connected to the microcontroller clock input. The four sensors are divided to two sections which are connected to two independent I<sup>2</sup>C interface of the microcontroller. This allows reading the inertial data from two sensors simultaneously and reducing the measurement time. The inertial data are transferred to PC via RS232 interface and may be analyzed later.

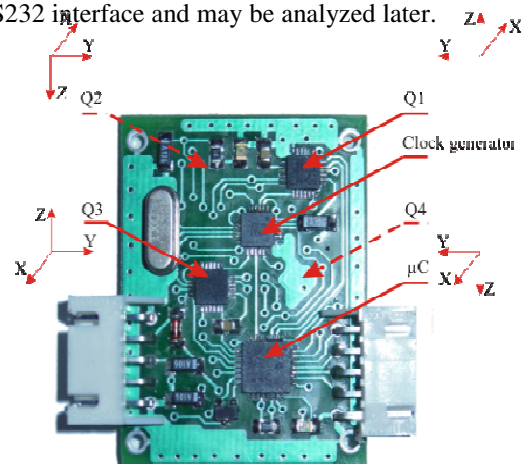


Figure 1. Measurement system design

## EXPERIMENTAL RESULTS

The inertial data are sampled in the synchronous mode, which is based on the simultaneous sampling of the inertial data of all four sensors. This mode is realized by a simultaneous activation of the FSYNC inputs of the sensors due to the circumstance that all FSYNC inputs are connected to one port of the microcontroller. The experimental data are obtained and processed in the MATLAB environment. The sensor update time is set to 1 kHz (accelerometers and gyroscopes) while the sampling frequency is set to 100Hz due to the limited time to transfer the inertial data to PC via RS232 interface. The system reads the XYZ linear and angular accelerations (12 bytes) and temperature (2 bytes) from each sensor and sends the data to PC. The data from the sensor couples (sensor #1 and #3) and (sensor #2 and #4) are read simultaneously because the couples are connected to the different hardware implemented I<sup>2</sup>C interfaces. The total amount of transferred data is equal to 57 bytes per one time sample (4 sensors multiplied by 14 bytes per sensor plus 1 SYNC byte). The measurement time is set to 2 hours.

The resulting accelerations of the quad sensor structure are calculated according to the equations:

$$\begin{aligned} A_x &= \frac{a_1 + a_2 - a_3 - a_4}{4} + g_x + a_x \\ A_y &= \frac{a_1 - a_2 - a_3 + a_4}{4} + g_y + a_y \\ A_z &= \frac{a_1 - a_2 + a_3 - a_4}{4} + g_z + a_z \end{aligned} \quad (1)$$

where  $a_x, a_y, a_z$  – acceleration of the structure  
 $g_x, g_y, g_z$  – earth acceleration  
 $a_i$  – bias of the  $i$ -th sensor

TABLE 1. Allan variance for X axis for single sensors and quad sensor structure

Error type	X acceleration		Curve slope	Units
	Sensor #1 Sensor #2 Sensor #3 Sensor #4	Quad sensor structure		
Bias instability $B = \frac{\sigma(f_0)}{0.664}$	1,18.10 <sup>-3</sup> 1,09.10 <sup>-3</sup> 1,58.10 <sup>-3</sup> 1,43.10 <sup>-3</sup>	2,1.10 <sup>-4</sup>	0	$m/s^2$
Acceleration random walk $K = \sigma(3)$	1.10 <sup>-4</sup> - 1,2.10 <sup>-4</sup> -	-	+1/2	$\frac{m/s^2}{\sqrt{s}}$
Velocity random walk $N = \sigma(1)$	7,0.10 <sup>-3</sup> 7,0.10 <sup>-3</sup> 7,5.10 <sup>-3</sup> 7,1.10 <sup>-3</sup>	3,8.10 <sup>-3</sup>	-1/2	$\frac{m/s}{\sqrt{s}}$

TABLE 2. Allan variance for Y axis for single sensors and quad sensor structure

Error type	Y acceleration		Curve slope	Units
	Sensor #1 Sensor #2 Sensor #3 Sensor #4	Quad sensor structure		
Bias instability $B = \frac{\sigma(f_0)}{0.664}$	9,03.10 <sup>-4</sup> 6,17.10 <sup>-4</sup> 9,03.10 <sup>-4</sup> 1,05.10 <sup>-3</sup>	2,93.10 <sup>-4</sup>	0	$m/s^2$
Acceleration random walk $K = \sigma(3)$	6.10 <sup>-5</sup> - 8.10 <sup>-5</sup> -	-	+1/2	$\frac{m/s^2}{\sqrt{s}}$
Velocity random walk $N = \sigma(1)$	7,5.10 <sup>-3</sup> 7,0.10 <sup>-3</sup> 6,9.10 <sup>-3</sup> 7,0.10 <sup>-3</sup>	3,8.10 <sup>-3</sup>	-1/2	$\frac{m/s}{\sqrt{s}}$

TABLE 3. Allan variance for Z axis for single sensors and quad sensor structure

Error type	Z acceleration		Curve slope	Units
	Sensor #1 Sensor #2 Sensor #3 Sensor #4	Quad sensor structure		
Bias instability $B = \frac{\sigma(f_0)}{0.664}$	3,01.10 <sup>-3</sup> 2,25.10 <sup>-3</sup> 3,76.10 <sup>-3</sup> 3.16.10 <sup>-3</sup>	1,28.10 <sup>-3</sup>	0	$m/s^2$
Acceleration random walk $K = \sigma(3)$	8,5.10 <sup>-4</sup> - 9,0.10 <sup>-4</sup> 6,0.10 <sup>-4</sup>	2,2.10 <sup>-4</sup>	+1/2	$\frac{m/s^2}{\sqrt{s}}$
Velocity random walk $N = \sigma(1)$	9,5.10 <sup>-3</sup> 9,5.10 <sup>-3</sup> 1,0.10 <sup>-2</sup> 9,5.10 <sup>-3</sup>	5,1.10 <sup>-3</sup>	-1/2	$\frac{m/s}{\sqrt{s}}$

The AV parameters are calculated for each axis acceleration of the single sensor and the resulting acceleration for the quad sensor structure according to the equation (1) and the results are shown at Table 1, Table 2 and Table 3 for the X, Y and Z linear acceleration respectively. It is clearly visible that all values of the quad sensor structure are lower than the corresponding noise parameter of the single sensor. The bias value theoretically may be reduced to zero according to the equation (1) because the bias values are summed with the opposite sign of the different values.



## CONCLUSION

The Allan Variance is a simple and efficient method for identifying and characterizing different stochastic processes and their coefficients. It is shown by this method that the quad sensor structure is distinguished with better noise parameters (acceleration random walk, velocity random walk and bias instability) in comparison with the single MEMS accelerometer. If the structure is manufactured as a single chip it may be used in many different applications like free-fall detection, precise tracking, kinetic measurements, tilt measurements, vehicle vibration monitoring, etc.

## REFERENCES

- [1]. D. W. Allan, "Time and frequency (time-domain) characterization, estimation, and prediction of precision clocks and oscillators," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. UFFC-34, no. 6, pp. 647–654, Nov. 1987. IEEE Std 1293-1996,
- [2]. IEEE Standard Specification Format Guide and Test Procedure for Linear, Single-axis, Nongyroscopic Accelerometers, IEEE 1996.
- [3]. IEEE 952-1997, IEEE Standard Specification Format Guide and Test Procedure for Single-Axis Interferometric Fiber Optic Gyros
- [4]. Haiying Hou - *Modeling Inertial Sensors Errors Using Allan Variance*, A Thesis, Department of Geomatics Engineering, Calgary, Alberta, 2004
- [5]. Songlai Han, Jinling Wang, Nathan Knight - *Using Allan variance to determine the calibration model of inertial sensors for GPS/INS integration*, 6th International Symposium on Mobile Mapping Technology, Presidente Prudente, São Paulo, Brazil, July 21-24, 2009
- [6]. H. Kim, J. G. Lee, C. G. Park, *Performance Improvement of GPS/INS Integrated System Using AV Analysis*, International Symposium on GNSS/GPS, Sydney, Australia, 2004
- [7]. N. El-Sheimy, H. Hou and X. Niu - *Analysis and Modeling of Inertial Sensors Using Allan Variance*, *IEEE Trans. Instr. Meas.*, vol. 57, no. 1, pp. 140-149, 2008
- [8]. MPU-6000/6050 datasheet - <http://invensense.com/mems/gyro/documents/PS-MPU-6000A-00v3.4.pdf>
- [9]. Emil Iontchev, Radostin Kenov, Rossen Miletiev, Ivaylo Simeonov - *Hardware implementation of quad microelectromechanical sensor structure for inertial systems*, 37<sup>th</sup> International Spring Seminar on Electronics Technology, May 7-11, 2014, Drezden, Germany
- [10]. LMK00105 datasheet - <http://www.ti.com/lit/ds/symlink/lmk00105.pdf>

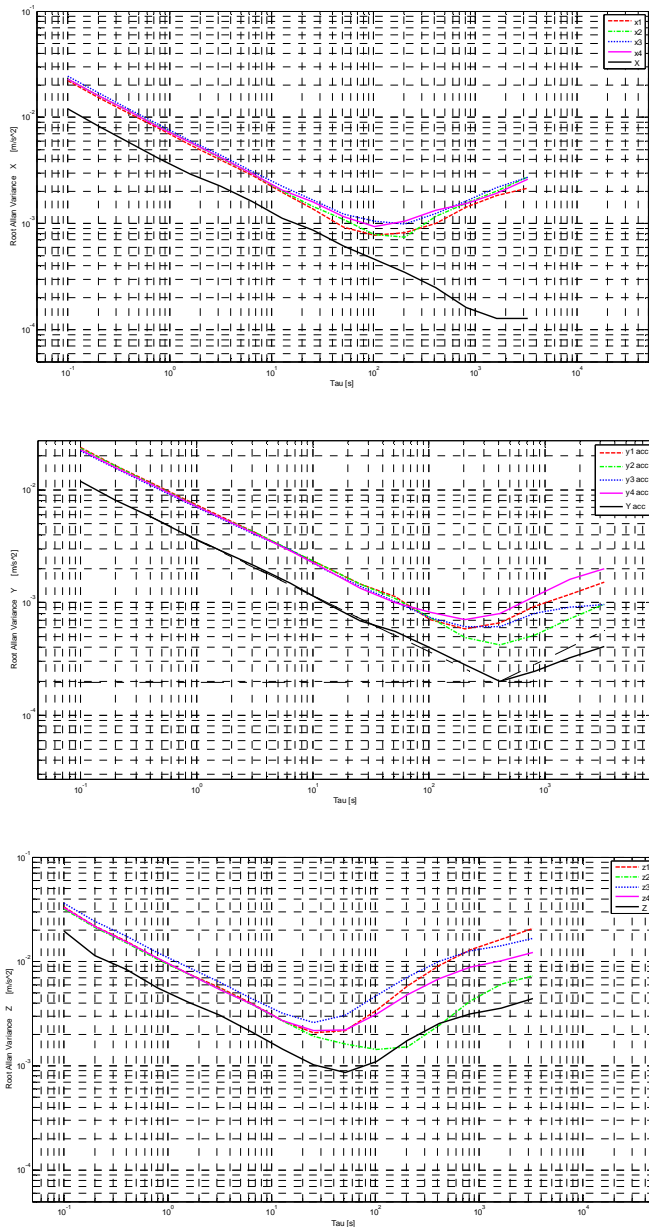


Figure 2. Allan deviation plot for all axes

The Figure 2 represents the AV plots for all axes of the single sensors and the quad sensor structure. The graphics shows that the noise components are lower for all values of the time cluster.

This Page Intentionally Left Blank

# Analysis of the noise parameters of quad structure of MEMS angular rate sensors

Rosen Miletiev<sup>1</sup>, Emil Iontchev<sup>2</sup>, Vladimir Bashev<sup>3</sup> and Rumen Yordanov<sup>4</sup>

**Abstract** – The current paper discusses the analysis of the noise parameters of quad sensor structure of angular rate sensors in comparison to the same parameters of the single sensors. It is shown that the signal processing of quad gyroscopes may result to the improved noise parameters of the whole structure. The Allan variance method is used to calculate the bias instability, rate random walk and angle random walk according to the curve slope.

**Keywords** – Quad sensor structure, MEMS, Allan variance

## INTRODUCTION

Inertial sensors, gyroscopes and accelerometers are widely applied. The sensor noises, bias variations, drifts and scale factor instabilities are the main error sources which cause the accuracy degradation of the modern navigation systems especially based on low cost MEMS inertial sensors. Two main parameters representing the quality of inertial sensors are the angle random walk and the bias instability for gyroscopes. These two parameters may be obtained by the Allan variance (AV) method [1-3] or by power spectral density (PSD) [4-6].

The PSD is the most commonly used representation of the spectral decomposition of a time series. It is a powerful tool for analyzing or characterizing data and stochastic modeling. The PSD, or spectrum analysis, is also better suited for analyzing periodic or aperiodic signals than other methods.

Allan Variance (AV) is a simple and efficient method for verifying and modeling these errors by representing the root mean square (RMS) random drift error as a function of averaging time. The AV can be used to determine the characteristics of different random processes. The details of the AV and how it is applied to the frequency and time metrology can be found in the literatures e.g. [7]. The most attractive feature of AV is its ability to sort out various noise components by the slopes on the root AV (Allan deviation) plot provided the different noise mechanisms are reasonably separated in the frequency and time domain

<sup>1</sup>Rosen Miletiev is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria. E-mail: [miletiev@tu-sofia.bg](mailto:miletiev@tu-sofia.bg)

<sup>2</sup>Emil Iontchev is with the Higher School of Transport "T. Kableshkov" 158 Geo Milev Street, Sofia 1574, Bulgaria, E-mail: [e\\_iontchev@yahoo.com](mailto:e_iontchev@yahoo.com)

<sup>3</sup>Vladimir Bashev is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria. E-mail: [v.bashev@gmail.com](mailto:v.bashev@gmail.com)

<sup>4</sup>Rumen Yordanov is with the Faculty of Electronic Engineering and Technology at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria. E-mail: [rsyordanov@yahoo.com](mailto:rsyordanov@yahoo.com)

The current paper discusses the noise parameters of the quad sensor structure, build by four identical digital output inertial sensors with synchronization inputs and microcontroller to compare the error values according to the single sensor using AV method.

## SYSTEM DESCRIPTION

The inertial measurement systems consists of four identical MEMS inertial sensors MPU-6000 [8], which are placed in such way that each two sensors have one common axis and two opposite axes (Figure 1) [9]. The proposed quad sensor structure is based on MPU-6000 MEMS sensors which combines a 3-axis gyroscope, 3-axis accelerometer, and a Digital Motion Processor™ (DMP). One of the most important special feature of the selected MEMS sensors is the FSYNC (Frame synchronization) input, which may be used to start the measurement process. The sync signal connected to the FSYNC pin may support image, video and GPS synchronization. The system synchronization is carried out by the multiple output clock generator. The MPU-6000 inertial sensor has external clock inputs of 32.768kHz or 19.2MHz, but the first one is too slow to obtain high speed measurements. Therefore the 19.2MHz oscillator is connected to the crystal oscillator interface of the clock generator. It is based on the Ultra-low Jitter LVC MOS Fanout Buffer/Level Translator LMK00105 [10] with 5 ultra low jitter clocks from a differential, single ended or crystal input. The four of these 5 clocks are used in the quad sensor structure to ensure the sensors synchronization and the fifth one is connected to the microcontroller clock input (Figure 1).

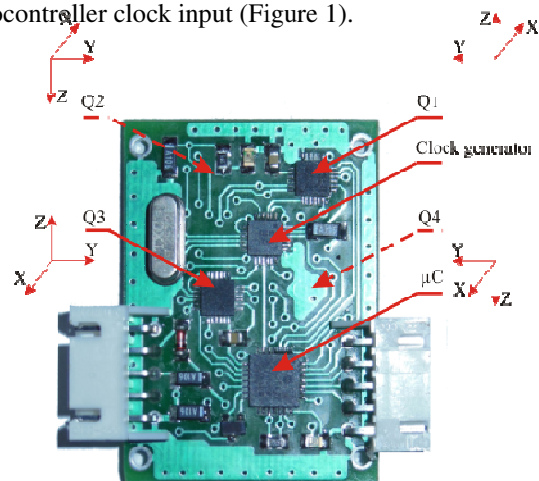


Figure 1. Measurement system design

The synchronous mode is distinguished with simultaneous sampling of the signal for all four sensors. This mode is realized by a simultaneous activation of the FSYNC inputs of the MEMS sensors. This action is possible because all FSYNC inputs are connected to the same port (PORTA) of the microcontroller and only one instruction is required to set all PORTA outputs simultaneously.

Two of the inertial sensors (Q1 and Q3) are placed on the top plate of PCB, while the other pair of the inertial sensors (Q2 and Q4) are placed on the bottom plate. The inertial sensor pairs (Q1-Q3) and (Q2-Q4) are connected to the different hardware based I<sup>2</sup>C based ports of the microcontroller type PIC18F25K22 ( $\mu$ C). This allows reading the inertial data from two sensor pairs simultaneously.

## EXPERIMENTAL RESULTS

The experimental data are obtained and processed in the MATLAB environment. The sensor update time is set to 1 kHz (accelerometers and gyroscopes) while the sampling frequency is set to 100Hz due to the limited time to transfer the inertial data to PC via RS232 interface. The system reads the XYZ linear and angular accelerations and temperature (14 bytes) from each sensor and sends the data to PC. The total amount of transferred data is equal to 57 bytes per one time sample (4 sensors multiplied by 14 bytes per sensor plus 1 SYNC byte). The measurement time is set to 2 hours.

The errors in typical inertial sensors include white noise, quantization noise, white rate noise, correlated (Markov) random drift, bias instability (1/f or flicker rate), flicker rate ramp (ramp instability), and random rate ramp. The AV method is used to analyze the bias instability, rate random walk and angle random walk of the quad sensor structure and the single sensors.

TABLE 1. Allan variance for X axis for single sensor and quad sensor structure

Error type	X gyro axis		Curve slope	Units
	Sensor #1 Sensor #2 Sensor #3 Sensor #4	Quad sensor structure		
Bias instability $B = \frac{\sigma(f_0)}{0.664}$	7,22.10 <sup>-3</sup> 6,47.10 <sup>-3</sup> 7,37.10 <sup>-3</sup> 6,32.10 <sup>-3</sup>	1,65.10 <sup>-3</sup>	0	deg/ s
Rate random walk $K = \sigma(3)$	1,7.10 <sup>-3</sup> 9,0.10 <sup>-4</sup> - 1,0.10 <sup>-3</sup>	1,2.10 <sup>-4</sup>	+1/2	$\frac{\text{deg/s}}{\sqrt{s}}$
Angle random walk $N = \sigma(1)$	2,3.10 <sup>-2</sup> 2,8.10 <sup>-2</sup> 2,2.10 <sup>-2</sup> 2,4.10 <sup>-2</sup>	1,2.10 <sup>-2</sup>	-1/2	$\frac{\text{deg}}{\sqrt{s}}$

TABLE 2. Allan variance for Y axis for single sensor and quad sensor structure

Error type	Y gyro axis		Curve slope	Units
	Sensor #1 Sensor #2 Sensor #3 Sensor #4	Quad sensor structure		
Bias instability $B = \frac{\sigma(f_0)}{0.664}$	7,37.10 <sup>-3</sup> 8,28.10 <sup>-3</sup> 6,92.10 <sup>-3</sup> 8,88.10 <sup>-3</sup>	2,4.10 <sup>-3</sup>	0	deg/ s
Rate random walk $K = \sigma(3)$	1,8.10 <sup>-3</sup> 2,1.10 <sup>-3</sup> - 2,2.10 <sup>-3</sup>	2,0.10 <sup>-4</sup>	+1/2	$\frac{\text{deg/s}}{\sqrt{s}}$
Angle random walk $N = \sigma(1)$	2,2.10 <sup>-2</sup> 2,5.10 <sup>-2</sup> 2,3.10 <sup>-2</sup> 2,5.10 <sup>-2</sup>	1,3.10 <sup>-2</sup>	-1/2	$\frac{\text{deg}}{\sqrt{s}}$

TABLE 3. Allan variance for Z axis for single sensor and quad sensor structure

Error type	Z gyro axis		Curve slope	Units
	Sensor #1 Sensor #2 Sensor #3 Sensor #4	Quad sensor structure		
Bias instability $B = \frac{\sigma(f_0)}{0.664}$	7,22.10 <sup>-3</sup> 6,92.10 <sup>-3</sup> 7,22.10 <sup>-3</sup> 6,32.10 <sup>-3</sup>	1,65.10 <sup>-3</sup>	0	deg/ s
Rate random walk $K = \sigma(3)$	1,6.10 <sup>-3</sup> 1,0.10 <sup>-3</sup> 1,3.10 <sup>-3</sup> 1,0.10 <sup>-3</sup>	1,2.10 <sup>-4</sup>	+1/2	$\frac{\text{deg/s}}{\sqrt{s}}$
Angle random walk $N = \sigma(1)$	2,3.10 <sup>-2</sup> 2,6.10 <sup>-2</sup> 2,2.10 <sup>-2</sup> 2,5.10 <sup>-2</sup>	1,2.10 <sup>-2</sup>	-1/2	$\frac{\text{deg}}{\sqrt{s}}$

The resulting angular accelerations of the quad sensor structure are calculated according to the equations:

$$\begin{aligned}
 G_{x,q} &= \frac{G_1 + G_2 - G_3 - G_4}{4} + G_x \\
 G_{y,q} &= \frac{G_1 + G_2 - G_3 - G_4}{4} + G_y \\
 G_{z,q} &= \frac{G_1 + G_2 - G_3 - G_4}{4} + G_z
 \end{aligned} \quad (1)$$

where  $G_{x,q}$ ,  $G_{y,q}$ ,  $G_{z,q}$  – angular rate of the structure,  $G_i$  – bias of the  $i$ -th sensor,  $G_x$ ,  $G_y$ ,  $G_z$  – angular rate



The AV parameters are calculated for each axis angular rate of the single sensor and the resulting angular rate for the quad sensor structure according to the equation (1) and the results are shown at Table 1, Table 2 and Table 3 for the X, Y and Z angular rate respectively. It is clearly visible that all values of the quad sensor structure are lower than the corresponding noise parameter of the single sensor. The bias value is also less than the single sensor value and theoretically may be reduced to zero according to the equation (1) because the bias values are summed with the opposite sign of the different sensors.

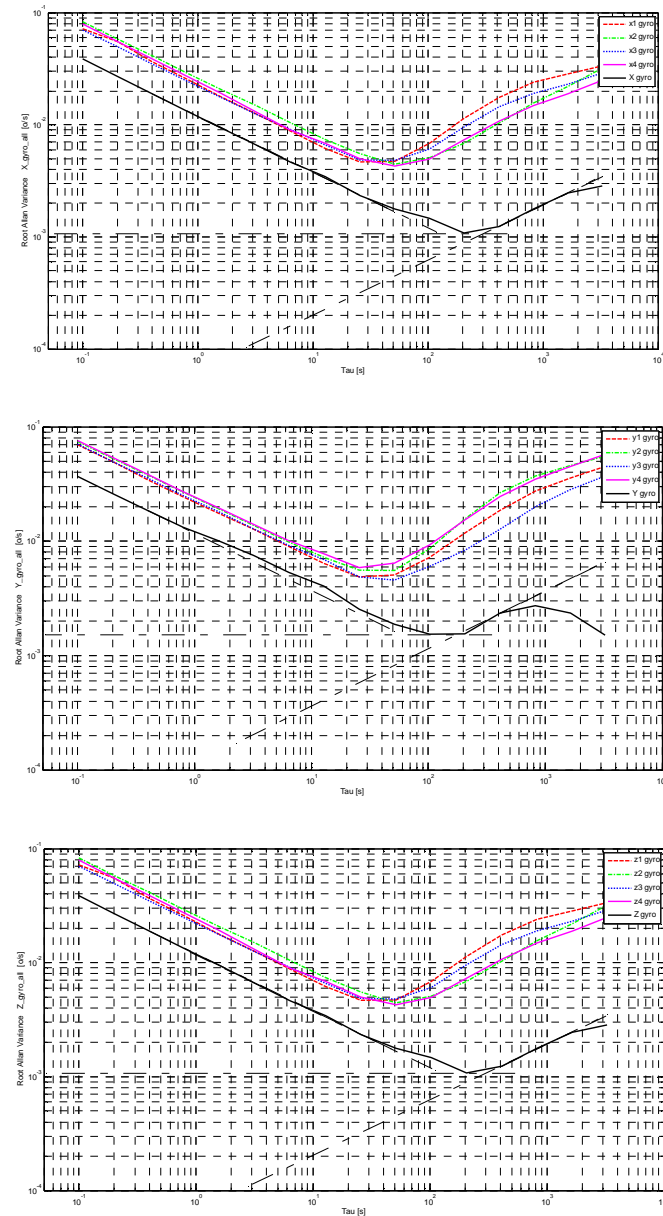


Figure 2. Allan deviation plot for all axes

The Figure 2 represents the AV plots for all axes of the single sensor and the quad sensor structure. The graphics shows that the noise components are lower for all values of the time cluster.

## CONCLUSION

The Allan-variance method presented in this paper allows a systematic characterization of the various random errors contained in the output data of the inertial sensor. By performing a simple operation on the entire length of data, a characteristic curve is obtained whose inspection facilitates the determination of the different types and magnitude of error terms existing in the inertial sensors.

It is shown by this method that the quad sensor structure is distinguished with better noise parameters (rate random walk, angle random walk and bias instability) in comparison with the single MEMS gyroscopes. If the structure is manufactured as a single chip it may be used in many different applications like free-fall detection, precise tracking, kinetic measurements, tilt measurements, vehicle vibration monitoring, etc.

## REFERENCES

- [1]. IEEE Standard Specification Format Guide and Test Procedure for Linear, Single -axis, Nongyroscopic Accelerometers, IEEE 1996.
- [2]. IEEE 952-1997, IEEE Standard Specification Format Guide and Test Procedure for Single-Axis Interferometric Fiber Optic Gyros
- [3]. Haiying Hou - *Modeling Inertial Sensors Errors Using Allan Variance*, A Thesis, Department of Geomatics Engineering, Calgary, Alberta, 2004
- [4]. IEEE Standard Specification Format Guide and Test Procedure for Single-Axis Laser Gyros, IEEE Std. 647-2006
- [5]. N. El-Sheimy, "Analysis and Modeling of Inertial Sensors Using AV", IEEE Transaction on instrumentation and measurement, vol. 57, No.1, JANUARY 2008
- [6]. H. Kim, J. G. Lee, C. G. Park, *Performance Improvement of GPS/INS Integrated System Using AV Analysis*, International Symposium on GNSS/GPS, Sydney, Australia, 2004
- [7]. N. El-Sheimy, H. Hou and X. Niu - *Analysis and Modeling of Inertial Sensors Using Allan Variance*, IEEE Trans. Instr. Meas., vol. 57, no. 1, pp. 140-149, 2008
- [8]. MPU-6000/6050 datasheet - <http://invensense.com/mems/gyro/documents/PS-MPU-6000A-00v3.4.pdf>
- [9]. Emil Iontchev, Radostin Kenov, Rossen Miletiev, Ivaylo Simeonov - *Hardware implementation of quad microelectromechanical sensor structure for inertial systems*, 37<sup>th</sup> International Spring Seminar on Electronics Technology, May 7-11, 2014, Drezden, Germany
- [10]. LMK00105 datasheet - <http://www.ti.com/lit/ds/symlink/lmk00105.pdf>

This Page Intentionally Left Blank

# Laboratory Performance Studies of WPA IEEE 802.11a PTP 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** – The increasing importance of wireless communications, involving electronic devices, has been widely recognized. Performance is a crucial issue, leading to more reliable and efficient communications. Security is also critically important. Laboratory performance measurements are presented for Wi-Fi IEEE 802.11 a WPA point-to-point links. Our study contributes to performance evaluation of this technology under WPA encryption, using available equipments (HP V-M200 access points and Linksys WPC600N adapters). New detailed results are presented and discussed, namely at OSI levels 4 and 7, from TCP, UDP and FTP experiments: TCP throughput, jitter, percentage datagram loss and FTP transfer rate. Comparisons are made to corresponding results of Open point-to-point links. Conclusions are drawn about the comparative performance of the links.

**Keywords** – Wi-Fi, WLAN, IEEE 802.11a, WPA Point-to-Point Links, Wireless Network Laboratory Performance.

## I. INTRODUCTION

The development of wireless communications technologies, e.g. Wi-Fi and FSO, has been possible using electromagnetic waves in several frequency intervals, propagating in the air. Their importance and utilization have been increasing.

Wi-Fi uses microwaves. It is an important and versatile technology that gives mobility and favorable prices. Wi-Fi has been increasing used for complementing wired networks. Although it can be used in ad-hoc, its main use is in infrastructure mode. Here an access point, AP, permits communications of Wi-Fi electronic devices with a wired based LAN through a switch/router, resulting in a WLAN, based on the AP. Wi-Fi has reached the personal home, where

a WPAN allows personal devices to communicate. Point-to-point (PTP) and point-to-multipoint (PTMP) setups are used both indoors and outdoors, with specific directional and omnidirectional antennas. PTP and PTMP links use microwaves in the 2.4 and 5 GHz frequency bands and IEEE 802.11a, b, g, n standards [1]. The 2.4 GHz band has been intensively used and is having increasing interferences. The 5 GHz band overcomes this problem, however, absorption is higher and ranges diminish.

Specifications permit nominal transfer rates up to 11 (802.11b), 54 Mbps (802.11 a, g) and 600 Mbps (802.11n). The medium access control is CSMA/CA. Studies were published on wireless communications, wave propagation [2,3], practical implementations of WLANs [4], performance analysis of the effective transfer rate for 802.11b PTP links [5], 802.11b performance in crowded indoor environments [6].

Performance has been a very important criterion to evaluate communications quality, 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].

Wi-Fi security is essential. However, as microwave radio waves propagate in the air, they can be very easily captured. WEP was initially intended to provide security like that of a wired LAN. In spite of its weaknesses, WEP is still widely used in Wi-Fi networks for security reasons, mainly in point-to-point links. More advanced and reliable security methods exist to provide authentication e.g., by increasing security order, WPA and WPA2. WPA implements the majority of the IEEE 802.11i standard [1]. It includes a message integrity check, replacing the CRC used in WEP.

Several performance measurements have been made for 2.4 and 5 GHz Wi-Fi open [8,9], WEP [10,11] and WPA links [12], as well as very high speed FSO [13]. In the present work new Wi-Fi (IEEE 802.11 a) results arise, using WPA links, namely through OSI levels 4 and 7. Performance is evaluated through laboratory measurements of WPA PTP links, using available equipments. Comparisons are made to corresponding results obtained for Open PTP links.

The rest of the paper is structured as follows: Section II presents the experimental details i.e. the measurement setup and procedure. Results and discussion are presented in Section III. Conclusions are drawn in Section IV.

## II. EXPERIMENTAL DETAILS

The measurements used a HP V-M200 access point [14], with three external dual-band 3x3 MIMO antennas, IEEE

<sup>1</sup>José Pacheco de Carvalho is with the Remote Sensing 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 Sensing 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 Sensing 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 Sensing 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.

802.11 a/b/g/n, software version 5.4.1.0-01-16481 and a 100-Base-TX/10-Base-T Allied Telesis AT-8000S/16 level 2 switch [15]. Two out of three PCs were used having a PCMCIA IEEE.802.11 a/b/g/n Linksys WPC600N wireless adapter with three internal antennas [16], to enable PTMP (two-node) links to the access point. In every type of experiment, interference free communication channels were used (ch 8 for 802.11 b, g). This was checked through a portable computer, equipped with a Wi-Fi 802.11 a/b/g/n adapter, running Acrylic WiFi software [17]. WPA encryption was activated in the AP and the wireless adapters of the PCs with a key composed of twenty six hexadecimal characters.. The experiments were made under far-field conditions. No power levels above 30 mW (15 dBm) were required, as the wireless equipments were close.

A laboratory setup has been planned and implemented for the measurements, as shown in Fig. 1. It can involve up to three wireless links to the AP. At OSI level 4, measurements were made for TCP connections and UDP communications using Iperf software [18]. 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 others, with IP 192.168.0.6, and 192.168.0.50 could be the Iperf clients. 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 [19]. The scheme of Fig. 1 was also used for FTP measurements, where FTP server and client applications were installed in the PCs. Another PC, with IP 192.168.0.20, was used to control the settings in the AP. The laboratory setup permitted three types of experiments to be made: PTP, using the client1 and the 192.168.0.2 PC as server; PTMP, using the client1 and the 192.168.0.2 PC as server; 4N-PTMP, using simultaneous connections/communications between the two clients and the 192.168.0.2 PC as server.

The scheme of Fig. 1 was also used for FTP measurements, where FTP server and client applications were installed in the PCs.

The server and client PCs were HP nx9030 and nx9010 portable computers, respectively, running Windows XP Professional. They were set to optimize the resources allocated to the present work. Batch command files have been re-written to enable the new 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 R&D Unit network, via switch.

### III. RESULTS AND DISCUSSION

The PC wireless network adapter were manually configured, for IEEE 802.11 a, with typical nominal transfer rates (6, 9, 12, 18, 24, 36, 48, 54 Mbps). For every fixed transfer rate, data were obtained for comparison of the

laboratory performance of WPA and Open PTP 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, signal to noise ratios (SNR, in dB) and noise levels (N, in dBm) were monitored and typical values are shown in Fig. 2. The main average TCP and UDP results are summarized in Table I, both for WPA and Open PTP links. The statistical analysis, including calculations of confidence intervals, was carried out as in [20].

In Figs. 3-4 polynomial fits were made to the 802.11 a TCP throughput data for WPA and Open PTP links, respectively, where  $R^2$  is the coefficient of determination. It was found that, on average, TCP throughput is slightly better for Open than WPA PTP links (Table I). This is due to increase in data length due to WPA encryption. In Figs. 5-6, the data points representing 802.11 a jitter data for WPA and Open PTP links, respectively, were joined by smoothed lines. It was found that, on average, the best jitter performances are for Open PTP links (Table I). Figs. 5-6 show decreases in jitter with increasing nominal transfer rate. In Fig. 7, percentage datagram loss data for 802.11 a are shown for WPA PTP links. The best performances were found for Open PTP links (Table I).

At OSI level 7 FTP transfer rates were measured versus nominal transfer rates configured in the access point and the wireless network adapters of the PCs for IEEE 802.11 a, as in [10]. The results show the same trends found for TCP throughput.

In comparison to Open links, for WPA links, TCP throughput, jitter and percentage datagram loss were found to show performance degradations, as data length is increased due to encryption.

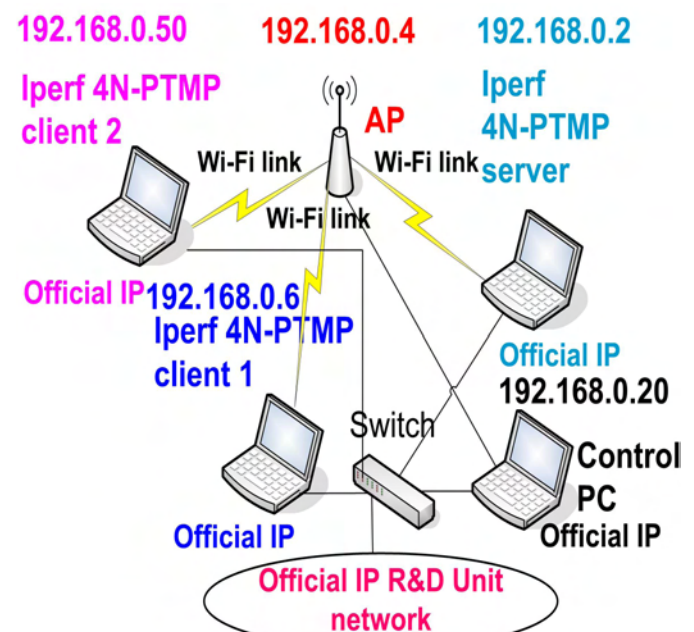


Fig. 1- Wi-Fi laboratory setup scheme.



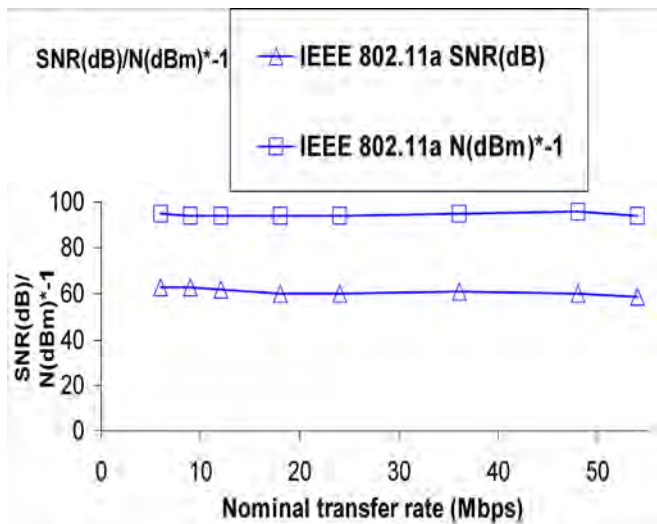


Fig. 2- Typical SNR (dB) and N (dBm); WPA.

TABLE I  
AVERAGE WI-FI (IEEE 802.11A) RESULTS;  
WPA AND OPEN PTP LINKS.

Link type	WPA	Open
TCP throughput (Mbps)	14.6 +- 0.4	15.0 +- 0.5
UDP-jitter (ms)	2.3 +- 0.2	2.2 +- 0.1
UDP-% datagram loss	1.9 +- 0.3	1.4 +- 0.1

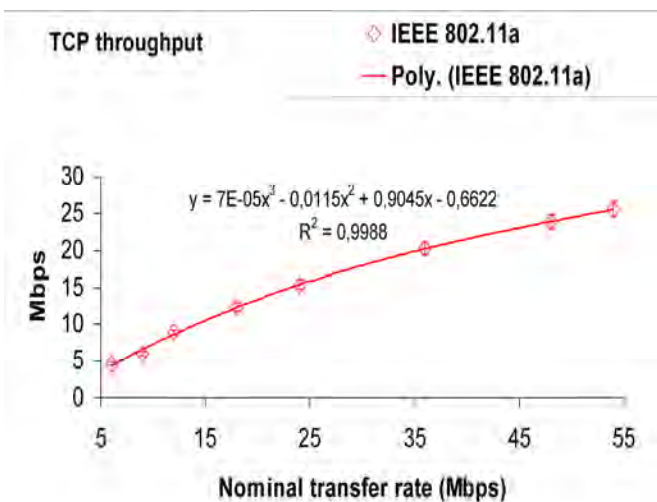


Fig. 3- TCP throughput versus technology and nominal transfer rate; WPA, PTP.

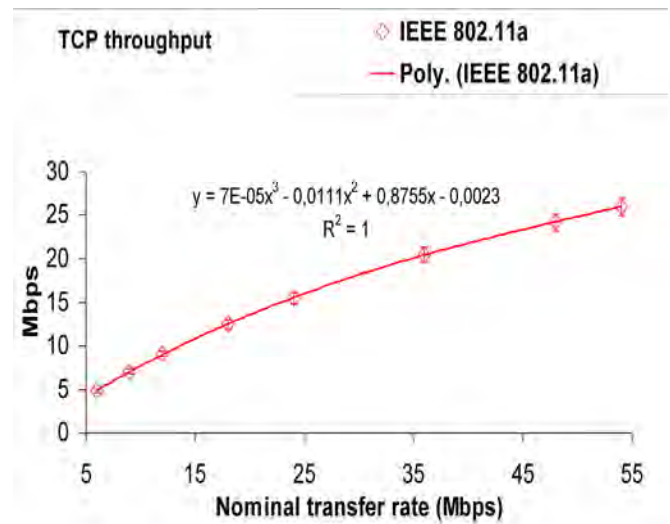


Fig. 4- TCP throughput versus technology and nominal transfer rate; Open, PTP.

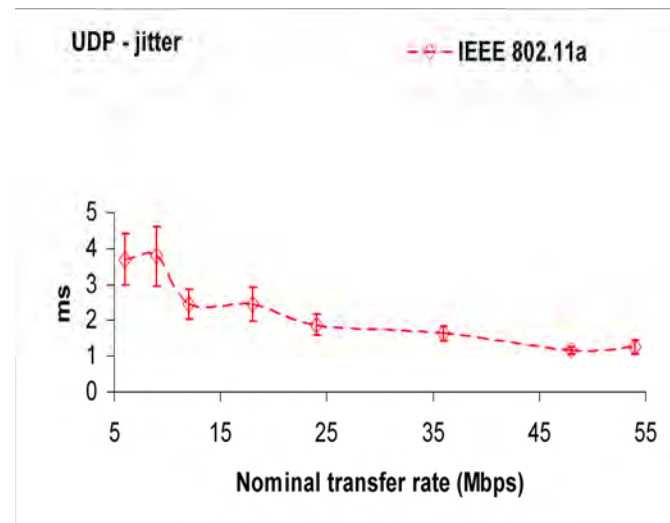


Fig. 5- UDP - jitter results versus technology and nominal transfer rate; WPA, PTP.

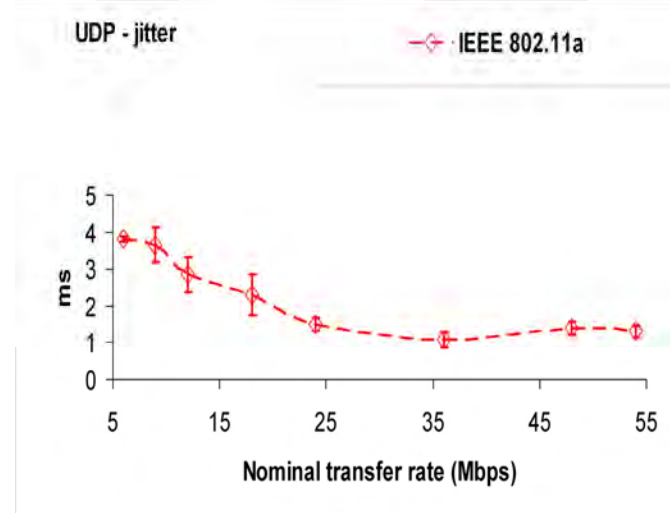


Fig. 6- UDP - jitter results versus technology and nominal transfer rate; Open, PTP.

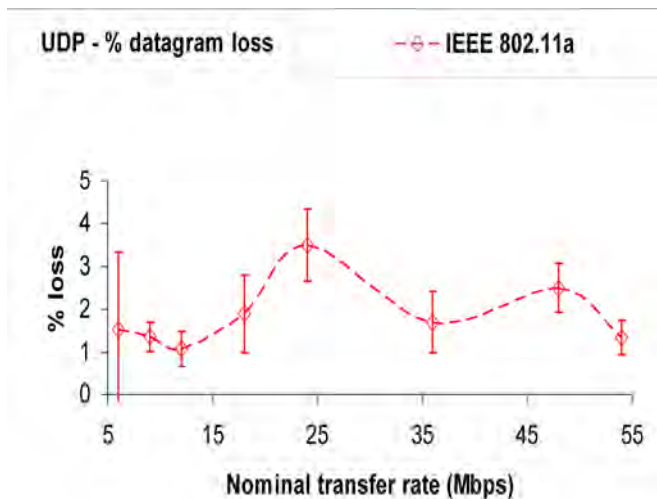


Fig. 7- UDP – percentage datagram loss versus technology and nominal transfer rate; WPA, PTP.

#### IV. CONCLUSION

A versatile laboratory setup arrangement has been planned and implemented, that permitted systematic performance measurements of available wireless equipments (V-M200 access points from HP and WPC600N adapters from Linksys) for Wi-Fi (IEEE 802.11 a) in WPA PTP links.

Through OSI layer 4, TCP throughput, jitter and percentage datagram loss were measured and compared for each standard to corresponding results obtained for Open PTP links. In comparison to Open links, TCP throughput, jitter and percentage datagram loss were found to show performance degradations for WPA links, where data length is increased due to encryption. However performance decreases in TCP throughput were found to be within the experimental error.

At OSI layer 7, FTP performance results have shown the same trends found for TCP throughput.

Further performance studies are planned using several equipments, topologies, security settings and noise conditions, not only in laboratory but also in outdoor environments involving, mainly, medium range links.

#### ACKNOWLEDGEMENT

Supports from Universidade da Beira Interior and FCT (Fundação para a Ciência e a Tecnologia)/PEst-OE/FIS/UI0524/2014 (Projecto Estratégico-UI524-2014) are acknowledged.

#### REFERENCES

- [1] Web site <http://standards.ieee.org> Web site; IEEE 802.11a, 802.11b, 802.11g, 802.11n, 802.11i standards; accessed 10 Jan 2014.
- [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, Guilin, China, 27-30 November 2006, pp. 1-4.
- [7] E. Monteiro, F. Boavida, Engineering of Informatics Networks, 4th ed., FCA-Editor of Informatics Ltd., 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, Vol. 60, pp. 507-514, 2010.
- [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, Vol. 90, pp. 105-114, 2011.
- [11] J. A. R. Pacheco de Carvalho, C. F. Ribeiro Pacheco, A. D. Reis, H. Veiga, "Laboratory Performance Measurements of IEEE 802.11 b, g WEP PTP Links", *Proc. WCE 2014 - World Congress on Engineering 2014*, Imperial College London, London, England, 2-4 July 2014, pp. 724-727.
- [12] J. A. R. Pacheco de Carvalho, H. Veiga, C. F. Ribeiro Pacheco, A. D. Reis, "Performance Evaluation of Laboratory Wi-Fi IEEE 802.11a Wpa Point-to-multipoint Links", *Proc. Conference on ENTERprise Information Systems CENTERIS 2013 / HCIST 2013*, Lisboa-Portugal 23-25 October 2013, SciVerseScienceDirect, Procedia Technology, 2013, 9, pp. 146 – 151.
- [13] J. A. R. Pacheco de Carvalho, Nuno Marques, H. Veiga, Cláudia F. F. P. Ribeiro Pacheco, A. D. Reis, "Contributions to Experimental Performance Studies of Wi-Fi and FSO Links", *Proc. CISTI 2010-5th Iberian Conference on Information Systems and Technologies*, Santiago de Compostela, Spain, 16-19 June 2010, pp. 221-226.
- [14] Web site <http://www.hp.com>; HP V-M200 802.11n access point management and configuration guide; 2010; accessed 3 Jan 2016.
- [15] Web site <http://www.alliedtelesis.com>; AT-8000S/16 level 2 switch technical data; 2009; accessed 10 Dec 2015.
- [16] Web site <http://www.linksys.com>; WPC600N notebook adapter user guide; 2008; accessed 10 Jan 2012.
- [17] Acrylic WiFi software; 2016; <http://www.acrylicwifi.com>; accesse 8 Jan 2016.
- [18] Iperf software; 2016; <http://iperf.fr>; accessed 16 Feb 2016.
- [19] Network Working Group. "RFC 1889-RTP: A Transport Protocol for Real Time Applications", <http://www.rfc-archive.org>; 1996; accessed 3 Jan 2014.
- [20] P. R. Bevington, Data Reduction and Error Analysis for the Physical Sciences, Mc Graw-Hill Book Company, 1969.

# **Remote Ecological Monitoring**





# Frequency characteristics and impedance as a criterion for purity by the monitoring of soil

Svetlana Gerganova-Savova<sup>1</sup>, Daniela Toneva-Zheynova<sup>2</sup> and Yanko Yanev<sup>3</sup>

**Abstract** – Electrical Impedance Spectroscopy (EIS-method) is reliable method, widely used in various fields of science and practice, allowing to measure impedance. Impedance measurement scheme can be presented as Randles model. The frequency characteristics of the clean and contaminated soil are substantially different.

**Keywords** – EIS-method, Randles model, Soil monitoring, Frequency characteristics, impedance.

The UN Standing Committee on Ecology defined environmental monitoring as a system of repeated observations on components and factors of the environment in spatially scale and consistently in time with specific purposes in accordance with monitoring programs [7]. Monitoring usually refers to processes in which observations are made over time and it is a special system for repeated observations and analysis of one or more elements of the environmental conditions, intended to register, evaluate and predict.

In the Bulgarian national system for environmental monitoring one the land - shaft, lands and soils are specific monitoring objects. The measurement generally refers to processes in which qualitative or quantitative properties are determined, as a rule. The main soil parameters, subject to measuring, with regard to soil contamination include: presence and concentration of heavy metals and metalloids, total nitrogen content ( $N_{tot}$ ), nitrate nitrogen, ammonia, phosphorus, organic carbon, robust response of the soil (pH), electrical conductivity, total carbon and persistent organic pollutants - 16 PAH, 6 PCB, 15 – chloro - organic pollutants, and others [6]. The needs of environmental monitoring, the frequency of sampling and measurement, allocations of sampling points and polygons, applied monitoring methodology should be determined correctly [8].

A wide variety of approaches and methods are applied in

the soil monitoring systems, but most of the conventional methods for soil monitoring are direct, purely laboratory methods, expensive. Therefore, the use and maintenance of calibrated and verified tools for monitoring and measurement should be ensured, as well as the development and application of devices and software. The Electrical Impedance Spectroscopy (EIS-method) is a commonly adopted method in many areas and in practice it is used for the measurement of electrical resistance, also for obtaining the frequency characteristics of the impedance of porous materials, as is the soil. The method allows to create a profile of soils differing in composition, which shows the moisture content (moisture regain) and is indicative regarding mineral content and in some cases pollution level [2], [3].

The scheme of measurement can be depicted in the form of Figure 1 and it shows the equivalent scheme, originally proposed by John Randals (John Edward Brough Randles), for modeling of surface electrochemical processes [1], [4], [5]. Randals model is simple, while the real electrochemical systems are usually more complex, but it is the basis for more complex cases.

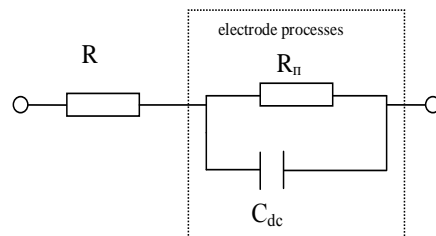


Fig. 1. Randles model

where:

$R$  - soil resistance [ $\Omega$ ],

$R_{\pi}$  - polarization resistance [ $\Omega$ ],

$C_{dc}$  - double layer capacity [F].

When the circuit has not only active but reactive elements as well, and the current is sinusoidal with circular frequency  $\omega$ , then the Ohm's law can be generalized, and the participating variables become complex.

In the frequency domain the dependence for the impedance has the form:

<sup>1</sup>Svetlana Gerganova-Savova - Automation of Manufacturing Department, Technical University of Varna, address: 1 Studentska, str., Varna 9010, Bulgaria, e-mail: gerganova-savova@lycos.com

<sup>2</sup>Daniela Toneva-Zheynova - Dept. Ecology and Environmental Protection Technical University of Varna, Varna, Bulgaria.

<sup>3</sup>Yanko Yanev - Automation of Manufacturing Department, Technical University of Varna, Varna, Bulgaria, e-mail: yanyanov@gmail.com

$$Z(j\omega) = R + j\omega X \quad (1)$$

Frequency characteristics for Z can be written as complex numbers in algebraic form.

For the Impedance readings at Randals model is formed [5]:

$$Z = R + \frac{1}{\frac{1}{R_n} + \frac{1}{Z_{dc}}} \quad (2)$$

$$Z = \left( R + \frac{R_n}{1 + \omega^2 R_n^2 C_{dc}^2} \right) - j \frac{\omega R_n^2 C_{dc}}{1 + \omega^2 R_n^2 C_{dc}^2}$$

The frequency characteristics - real frequency, imaginary frequency, amplitude - phase characteristic, amplitude - frequency, a phase - frequency are calculated and analyzes regarding the purity of the soil.

For this purpose five measurements of different depths have been made on soils gathered from the North-Eastern regions of Bulgaria [3]. The measurements for the non-contaminated soil are at frequencies: 2000, 6000, 10000, 15000, 20000 Hz, at a temperature of the air 28 degrees , 25.6 degrees for the soil and a depth of 0-20cm and 40-60cm.

The measurements of the contaminated soil are made for the same frequency diapason of depths: 0-20cm, 20-40cm, 40-60cm. They are processed in Matlab and some of the obtained results are systematized and presented below.

```
>> f=[2000,6000,10000,15000,20000]
```

```
>> w=2*pi*f
```

```
w=
```

```
12560    37680    62800    94200    125600
```

```
>> Re1=[1271,1268,1265,1260,1249]
```

```
>> Im1=[-195,-101,-23,73,167]
```

```
>> Re2=[890,909,909,909,901]
```

```
>> Im2=[-188,-68,-9,65,131]
```

```
>> hold on
```

```
>> plot(Re1,Im1,'b')
```

```
>> plot(Re2,Im2,'r')
```

```
>> Re3=[110,105,103,102,101]
```

```
>> Im3=[-80,-35,-19,-6,4]
```

```
>> Re4=[96,91,90,89,88]
```

```
>> Im4=[-60,-27,-15,-4,5]
```

```
>> Re5=[103,98,97,95,94]
```

```
>> Im5=[-60,-28,-16,-5,5]
```

```
>> plot(Re3,Im3,'g--')
```

```
>> plot(Re4,Im4,'c--')
```

```
>> plot(Re5,Im5,'m--')
```

```
.....
.....
```

The results are presented in fig1 and fig.2. The top two graphics are for clean soil. The bottom three graphics are for contaminated soil.

In fig1 are presented amplitude – frequency characteristics.

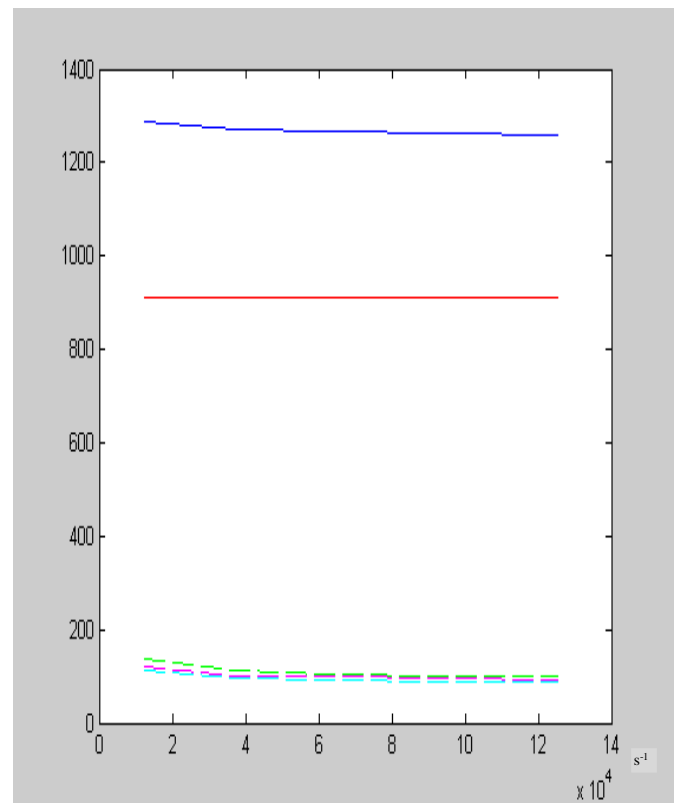


Fig.2. Amplitude – frequency characteristics

In fig. 2 are presented phase – frequency characteristics. The graphics are presented in Mathlab.

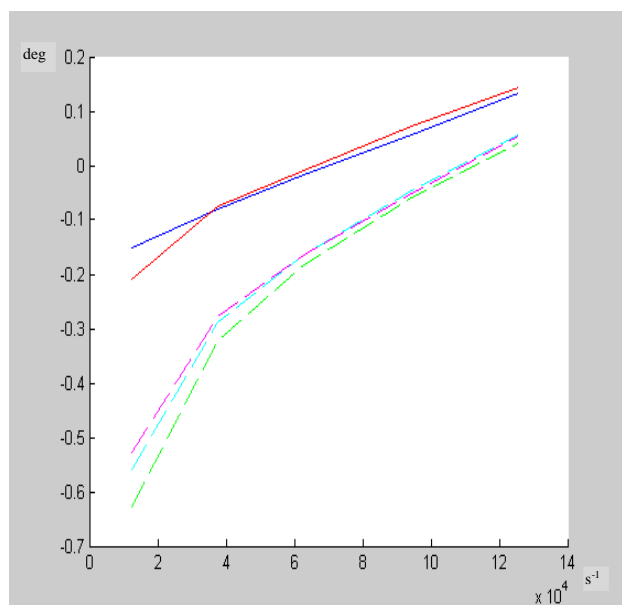


Fig.3. Phase – frequency characteristics

The frequency characteristics amplitude - frequency and phase - frequency of the clean and contaminated soil extract are substantially different - fig.2, fig.3. In rainy weather the soil has good conductive abilities, under certain chemical contaminants and in these cases  $Z=(101-102) \Omega$ . The soil contaminated with nitrates is a good conductor and the frequency characteristics are not affected by the depth of the probes, or by the frequency. The conductivity of water soluble electrolytes in the soil correlates with the concentration of the organic and micro organic elements. In terms of the environment increased levels of conductivity of the soil shows possible non-organic mineral pollution. The ammonium nitrate changes the electrical characteristics of the soil [4]. The graphics of the amplitude - phase characteristic are also indicative [5], that can be used with the same success.

Thus, by using the EIS method the content of fertilizer in the soil can be observed. This is especially important in the sanitary-protection zones of water catchments for drinking purposes, because the monitoring includes not only observation, but an assessment of the actual condition of the environment as well as forecasting its changes, and depending on the range, the tasks and directions. The purpose of monitoring is not only a passive observation of facts but also for conducting experiments, as well as process modelling as a basis for forecasting.

## REFERENCES

- [1] Янев Я., Д. Йонова, Относно използването на уреда Z-Meter III за изследвания по метода на електро импедансна спектроскопия, Годишник на ТУ-Варна, 2013г.
- [2] Parilkova J. Z-Meter III – User's Manual, Brno 2011.
- [3] Toneva – Zheyнова D., D. Ionova, A. Staneva, N. Atanasov, Application of EIS to environmental monitoring of soil, 2-nd Conference and Working Session within the Framework in the International Programme EUREKA, Brno, October 30-31, 2014.
- [4] Савова С., Д. Тонева-Жейнова, Мониторинг на почвата чрез измерване на електрическия импеданс и прилагането на метода EIS, XXV национален научен симпозиум с международно участие "Метрология и метрологично осигуряване 2015", 357-362.
- [5] Toneva-Zheyнова D., S. Savova, Y. Yanev, Deming cycle and soil monitoring by Z-Meter III, EUREKA2015, 3-rd Conference and Work Session, Brno University of Tehnology, 109-114.
- [6] <http://eea.government.bg/bg/nsmos>
- [7] <http://razrabotkite.blog.bg/biznes/2010/09/16/monitoring-na-okolnata-sreda-opredelenie-syshnost.606632>
- [8] <http://www.iso.org/iso/home.html>

This Page Intentionally Left Blank



# Client-Server System for Improving Ecological Conditions of the Environment

Dijana Jagodić<sup>1</sup>, Dejan Vujičić<sup>1</sup>, Siniša Randić<sup>1</sup>

**Abstract** – This paper describes the client-server system that was designed to improve ecological awareness of the society. Illegal landfills represent a big problem nowadays, and communal services cannot deal with them all, particularly because some of them are on unknown locations. The presented system can help communal services to better track illegal landfills and to react accordingly, but also any other communal problems can be reported. System is composed of Android client application, which is used to send information about ecological problems. These records are sent to the server, where the dispatchers review the issue and react accordingly to the problem. Users can send coordinates and description of the problem location, to facilitate problem solving process.

**Keywords** – Android, client-server, ecology, illegal landfills, waste management.

## I. INTRODUCTION

Natural resources are the key factor of the economic prosperity of a country. In order to be used at a maximum level, they need to be preserved and exploited in ecologically acceptable manner. However, with the technological development and increase in manufacturing production, the problem of waste management emerges as one of the major issues. The resolving of this problem not only involves the state legislations and regulations, but also the social responsibility of the companies and the whole society. While the companies' ecological responsibilities are primarily regulated by laws and inspections, the awareness of the society, particularly in developing countries, such as Serbia, is at low level. There have been numerous campaigns to increase ecological education in Serbia, but unfortunately, the reality is less than satisfactory.

Illegal landfills and undesirable disposal of communal waste are major factors in disturbing of natural balance. Large number of metal cans, plastic packaging, and even hazardous toxic materials are just dropped in nature and left to the mercy of ecologically responsible individuals. The degradation period of these materials can be longer than hundred years, some of them cannot be degradable at all. Their presence influences not only the soil, but also the watercourses and the ecological environment at a whole.

This paper describes the client-server system that is designed to help the community and state to overcome these ecological issues by suggesting larger involvement of the

society and individuals in this process. The people are given a chance to prove their ecological consciousness by reporting any illegal landfills or any sort of dump sites they stumble upon. In this way, the proper cooperation between the society and communal services can bring long expected balance to the environment.

The chapters in this paper are organized as follows. The second chapter brings the issues in waste management, particularly with the communal waste, packaging waste, and illegal landfills. In the third chapter the client-server system for reporting of ecological problems is described. The final chapter summarizes the whole matter.

## II. WASTE MANAGEMENT – MAJOR CHALLENGE OF THE MODERN SOCIETY

The waste management policy at European Union level is defined in Waste management strategy. It is based on hierarchy of principles, where the biggest priority is given to the prevention of waste formation, then to reusability and recycling, energy regeneration, and finally, to the waste disposal [1].

It is considered that Republic of Serbia belongs to the group of countries with low level of waste management. The following problems have been emphasized [1]:

- Omissions in applying the laws regarding waste management and ecological preservation;
- Lack of infrastructure;
- High costs and irrational organization of waste management;
- Low level of quality of service;
- Insufficient concern and enrollment of society and individuals into environmental problem solving.

Chapter 27 in the Serbia's application for membership of the European Union is regarding environmental conditions. In its document from 2011, it is said that Serbia has implemented crucial laws in line with the recommendations from EU, particularly in terms of waste prevention, reuse, recycling, and recovery. However, the major problems are emphasized as low waste collection rates in rural communities and vast number of illegal landfills [2]. In 2012, Serbia was successful in terms of registration and reporting on waste, but still, the waste collection rate in rural areas remains an issue [3]. By opening a regional waste management center in Pirot in 2013, Serbia moved on toward better waste management society. However, European Commission suggested that landfilling should be used only as a necessity, if there are no other waste management methods [4]. In 2014, a new waste management facility was opened and Serbia has had seven EU compliant regional landfills. European Commission emphasized the necessity of closing illegal landfills [5]. A regulation on the

<sup>1</sup>Dijana Jagodić, Dejan Vujičić, and Siniša Randić are with the Faculty of Technical Sciences in Čačak at University of Kragujevac, Sv. Save 65, Čačak 32000, Serbia, E-mails: [dijana7j@gmail.com](mailto:dijana7j@gmail.com), [dejan.vujicic@ftn.kg.ac.rs](mailto:dejan.vujicic@ftn.kg.ac.rs), [sinisa.randjic@ftn.kg.ac.rs](mailto:sinisa.randjic@ftn.kg.ac.rs)

plan for reducing packaging waste in period 2015 – 2019 was adopted by Serbia in 2015, but its implementation mainly depends on resolving “systematic weaknesses in the implementation of environmental projects” [6]. This paper addresses this problem, by suggesting larger enrolment of society and individuals in the preservation of ecological environment.

### A. Communal Waste

The data about communal waste are collected via reports from local communal services. Some indicators regarding communal waste for the period 2011 – 2014, are given in Table I [7].

TABLE I  
SOME INDICATORS REGARDING COMMUNAL WASTE FOR THE PERIOD  
2011 – 2014

Year	2011	2012	2013	2014
Amount of generated waste (millions of tons)	2.71	2.62	2.41	2.13
Amount of collected waste (millions of tons)	2.09	1.83	1.92	1.67
Amount of reused waste (tons)	N/A	N/A	24843	14521
Average percentage of waste collection (%)	77	~70	80	~80
Average daily amount of waste per capita (kg)	1.01	0.99	0.92	0.81
Average yearly amount of waste per capita (tons)	0.37	0.36	0.34	0.30

From Table I, it can be seen that the total amount of generated waste is decreasing, starting with the year 2011, which is in accordance with the data from the European Environment Agency, where many countries reported decreased amount of communal waste in 2014 [7].

### B. Packaging Waste

Packaging waste management is primarily related to preservation of natural resources. Some indicators regarding packaging waste for the period 2011 – 2014 are given in Table II [7-10].

TABLE II  
SOME INDICATORS REGARDING PACKAGING WASTE FOR THE PERIOD  
2011 – 2014

Year	2011	2012	2013	2014
Amount of market released packaging (thousands of tons)	343.66	344.25	321.58	327.71
Amount of reused waste (thousands of tons)	96.12	126.20	87.95	102.67
Percentage of collected paper and cardboard packaging (%)	79.51	72.35	63	62.8
Percentage of collected plastic packaging (%)	3.83	13.8	15	16.3
Percentage of collected wood packaging (%)	8.54	2.08	8	7.5
Percentage of collected metal packaging (%)	0.16	0.46	4	4.7
Percentage of collected glass packaging (%)	7.96	11.32	10	8.7

Packaging waste management has important ecological, social, and economic aspect. Packaging waste, depending on its category, has very long time of recycling. For example, it takes 500 years for degradation of one aluminum can. Degradation of PET packaging takes 100 years. Glass packaging will never degrade [11].

As can be seen from Table II, reusability of packaging waste has increased over years, but still it represents less than a third of an amount of market released packaging. Also, paper and cardboard packaging represent the major part of packaging waste. This issue can be resolved by building more recycling factories and waste management facilities, which is key principle in the Strategy of the waste management for the period 2010 – 2019 of the Republic of Serbia [12].

### C. Illegal Landfills

Data from 2014 show that there are total of seven sanitary landfills in the Republic of Serbia, and two are in the process of getting a working license. Communal services dispose communal waste to a total of 120 landfills. It has been located 3085 old and wild landfills. However, these data are collected from total of 142 municipalities. Landfills were not located in 15 municipalities, while 13 municipalities didn't send any data [7]. The location of illegal landfills in the Republic of Serbia, not including province of Kosovo and Metohija which is under temporary control of United Nations, is shown in Fig. 1 [13]. Yellow markers represent illegal landfills, and red markers represent communal services' landfills.

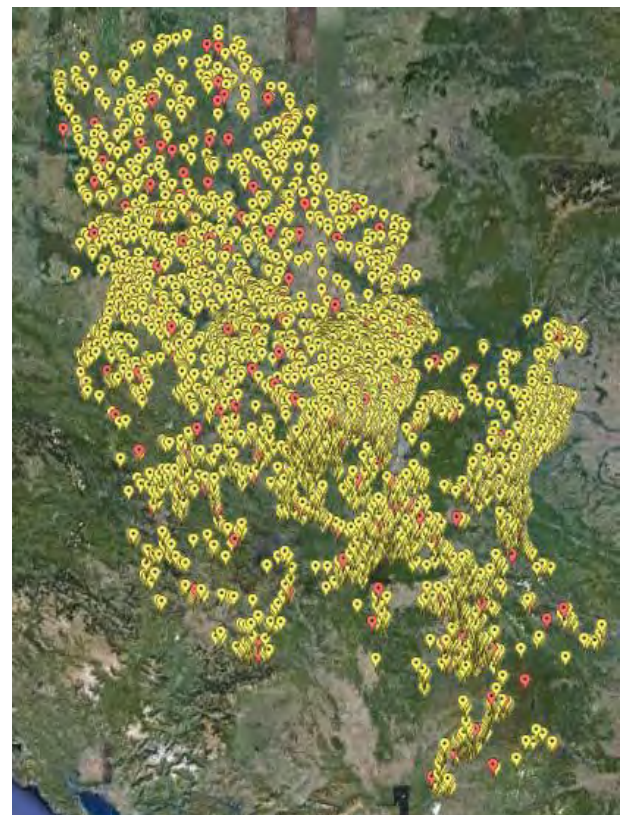


Fig. 1. Locations of illegal and communal landfills in the Republic of Serbia



### III. PRACTICAL REALIZATION OF THE SYSTEM

The system for reporting of ecological problems is comprised of client Android application and server for data acquisition and representation.

#### A. Client Android Application

The Android application layout is shown in Fig. 2.

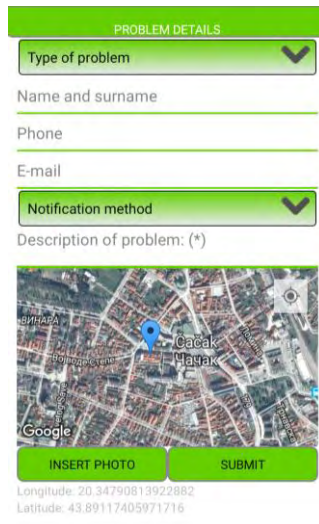


Fig. 2. The layout of Android client application

The field *Type of problem* is designed to select one of the three categories that the system is able to report:

1. Illegal landfill;
2. Other communal problems;
3. Inspection services.

The primarily target of reporting are illegal landfills, but some minor communal problems such as garbage on the public areas, can be reported as well. The third category is designed for reporting other communal problems not necessarily related to landfills.

The user may enter its name, phone, and e-mail address. If the user enters its phone or e-mail address, than by selecting the option from *Notification method* drop-down field, it can be notified about taken actions on the problem solving. The only obligatory field is *Description of problem*, in which user enters text describing the nature of the problem. The location of the troubleshooting place can be retrieved via GPS service inside the mobile phone, or by manual selection by placing the marker on the included map.

Another useful feature is the image of the problem, which user can take with camera on its mobile phone or select from device storage and send it together with other data. By clicking on the button *Submit*, the problem is sent to the server.

#### B. Server Side of the System

The data are sent to the server by using the SOAP – based Web service. Image is sent as an array of bytes. The Web

service is written in C# and it stores all the data in the database, together with the current date and time. The Web service acts as a mediator and represents more secure and reliable solution to accessing database.

Server – side application is written in PHP. It is intended for use in the dispatcher center of the communal services. Pictures are retrieved from the database by using *base64\_encode* PHP function, which encodes the byte representation of the stored image and displays it in PNG file format. For geolocation, the Google Maps API is used.

At first, users must enter their credentials to access the system. When accessed, they can view all the reported problems in two manners: as a list sorted by time of reporting, or on a map. The list view is shown in Fig. 3. (Please note that the images are taken from [14] and [15] and together with locations are used solely as a demonstration and may not represent the actual state on the field.)







Date	Type of problem	Description	Image	Location
2016-04-12 13:42:36	Other communal problems	Look at this! How embarrassing!		
2016-03-07 12:21:11	Illegal landfill	Near the local road, a huge amount of garbage has been dropped! Please do some surveillance because this is not the first time.		
2015-10-26 10:02:03	Illegal landfill	There is a massive junkyard not too far from the """" It is a great shame and something should be done immediately to resolve this embarrassment.		

Fig. 3. List of reported problems

The list shown in Fig. 3 consists of date and time of reporting, category (*Type of problem* field), problem description, its image if available, and location on the map.

The map view of reported problems is shown in Fig. 4. (Please note that locations and images are arbitrary and may not represent the actual state).

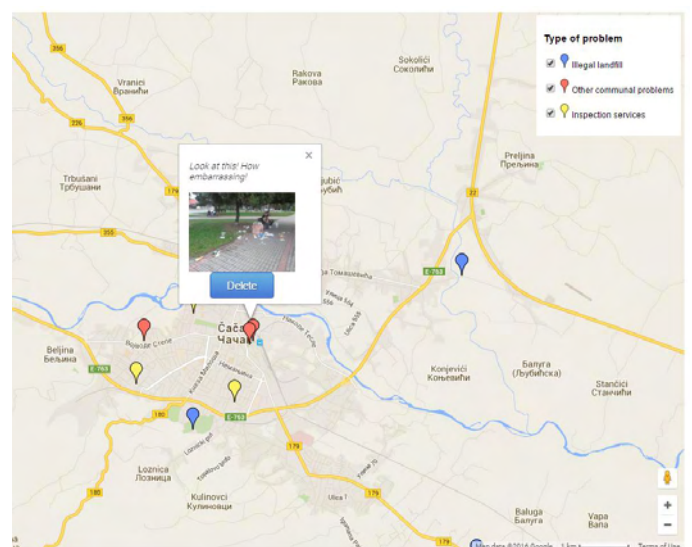


Fig. 4. Map view of reported problems

As can be seen from Fig. 4, different types of problems are represented with different colored markers. By checking the categories from top right corner, all corresponding markers are shown on the map. Clicking on the marker shows the info window in which the description and image of the problem is displayed. The user can click on the image and the bigger picture is displayed. Also, the user can delete the entry by clicking on the button *Delete*.

#### IV. CONCLUSION

Ecological environment preservation is one of the biggest challenges of the modern society. The enrollment of every part of the system is necessary in order to achieve this goal. Naturally, it all ends with the common citizens and their understanding of key principles of ecological balance. The education system needs to emphasize the major problems and to provide proper solutions and recommendations to solve this issue.

The amount of waste is constantly increasing and its management is of crucial importance. The planned construction of landfills and recycling facilities is of biggest importance. However, low ecological awareness of the society and lack of discipline and regulation enforcement leads to the formation of illegal landfills and dump sites that are not properly supervised. Furthermore, this brings misbalance to the ecological environment and can cause pollution of soil and watercourses.

In this paper, the client-server system for reporting ecological problems is presented. It consists of a client Android application that is intended for use among citizens, and server side, that is supposed to be used within communal services. Users can take a picture of a problem and send its coordinates to facilitate the issue solving process.

There are some indications that this system will be used by the communal service in the city of Čačak. If this becomes true, then the system can be slightly modified to adapt to the real situation on the field. Furthermore, by implementing this system, the citizens will have better insight in the work of communal service and entire municipality management.

#### ACKNOWLEDGEMENT

The work presented in this paper was funded by grant TR32043 for the period 2011-2016, by the Ministry of Education, Science, and Technological Development of the Republic of Serbia.

#### REFERENCES

- [1] Statistics Department of the Republic of Serbia, *Waste statistics and waste management in the Republic of Serbia for the period 2008 – 2010*, Belgrade, 2012, in Serbian
- [2] European Commission, *Commission staff working paper; Analytical report accompanying the document Communication from the commission to the European parliament and the council commission opinion on Serbia's application for membership of the European Union*, Brussels, 2011
- [3] European Commission, *Commission staff working document; Serbia 2012 progress report accompanying the document Communication from the commission to the European parliament and the council; Enlargement strategy and main challenges 2012-2013*, Brussels, 2012
- [4] European Commission, *Commission staff working document; Serbia 2013 progress report accompanying the document Communication from the commission to the European parliament and the council; Enlargement strategy and main challenges 2013-2014*, Brussels, 2013
- [5] European Commission, *Commission staff working document; Serbia 2014 progress report accompanying the document Communication from the commission to the European parliament, the council, the European economic and social committee and the committee of the regions; Enlargement strategy and main challenges 2014-2015*, Brussels, 2014
- [6] European Commission, *Commission staff working document; Serbia 2015 report accompanying the document communication from the commission to the European parliament, the council, the European economic and social committee and the committee of the regions; EU enlargement strategy*, Brussels, 2015
- [7] Serbian Environmental Protection Agency, *The report on the conditions of the environment in the Republic of Serbia for the year 2014*, Belgrade, 2015, in Serbian
- [8] Serbian Environmental Protection Agency, *The report on the conditions of the environment in the Republic of Serbia for the year 2013*, Belgrade, 2014, in Serbian
- [9] Serbian Environmental Protection Agency, *The report on the conditions of the environment in the Republic of Serbia for the year 2012*, Belgrade, 2013, in Serbian
- [10] Serbian Environmental Protection Agency, *The report on the conditions of the environment in the Republic of Serbia for the year 2011*, Belgrade, 2012, in Serbian
- [11] Serbian Environmental Protection Agency, *The report on the packaging and packaging waste management in the year 2014*, Belgrade, 2015, in Serbian
- [12] Republic of Serbia, *The strategy of waste management for the period 2010 – 2019*, Službeni glasnik Republike Srbije, no. 29/10, 2010, in Serbian
- [13] [www.sepa.gov.rs](http://www.sepa.gov.rs)
- [14] [www.caglas.rs](http://www.caglas.rs)
- [15] [www.ozonpress.net](http://www.ozonpress.net)



# **Engineering Education**



# An Overview of the Possibilities of the Variational Analysis for Exploration of Electric Circuits

Emil Panov<sup>1</sup> and Miroslava Doneva<sup>2</sup>

**Abstract** – The paper presents an overview of the variational analysis in combination with the transferring coefficients of the electric circuits (EC). In the reference papers [1 - 9] the methodologies for DC, AC and transient analysis of linear and non-linear EC are developed. Some examples are presented in order to illustrate the approach.

**Keywords** – variational approach, variational analysis of electric circuits, basic theorems for variational analysis, methodologies for variational analysis.

## I. INTRODUCTION

The variational approach is a powerful tool for analysis in mechanics, field theory, quantum mechanics and some other areas of modern science. The first author of the present paper developed four new theorems [1, 5] and the first systematical methodologies for variational analysis of EC in circuit theory [1-6].

Two theorems [1] can be used directly for the variational analysis of all types of regimes in a given linear EC (LEC), because they introduce the bases of such a type of analysis for EC.

**First theorem:** For each EC, among all sets of currents, which formally satisfy Kirchhoff's current law (KCL) for the nodes of the circuit, there is only one set of currents, for which the instantaneous power of each current source has an extremum (a minimum or a maximum), if this power is not equal to zero. And this set of currents is the only one, which satisfies the equations using Ohm's and Kirchhoff's laws for the circuit.

**Second theorem:** For each EC, among all sets of voltages, which formally satisfy Kirchhoff's voltage law (KVL) for the loops of the circuit, there is only one set of voltages, for which the instantaneous power of each voltage source has an extremum (a minimum or a maximum), if this power is not equal to zero. And this set of voltages is the only one, which satisfies the equations using Ohm's and Kirchhoff's laws for the circuit.

The variational analysis of harmonic LEC is comfortably to be introduced in combination with the phasor approach, because of the possibility for algebrization of the solutions. That can be achieved by the help of the next two theorems [5].

**Third theorem:** For each harmonic LEC, there is only one set of complex branch currents, for which the apparent pseudo-power of any of the current sources in

the explored circuit  $\left| \dot{S}_{j_{pseudo}} \right|$  has an extremum (a minimum or a maximum), where  $\dot{S}_{j_{pseudo}} = \dot{U}_j \cdot \dot{J}$  is the complex pseudo-power of the given source and  $\dot{U}_j$  is the complex voltage drop across it.

**Fourth theorem:** For each harmonic LEC, there is only one set of complex voltage drops across its elements, for which the apparent pseudo-power of any of the voltage sources in the explored circuit  $\left| \dot{S}_{E_{pseudo}} \right|$  has an extremum (a minimum or a maximum), where  $\dot{S}_{E_{pseudo}} = \dot{E} \cdot \dot{I}_E$  is the complex pseudo-power of the given source and  $\dot{I}_E$  is the complex current, flowing through it.

## I. VARIATIONAL ANALYSIS OF EC

### A. SHORT THEORY OF THE VARIATIONAL ANALYSIS OF DC AND AC LEC

The variational analysis of one LEC working at DC regime by the help of the the basic laws for EC can be conducted by the following methodology:

1) A reference source (of e.m.c. with a current  $j_1$  or e.m.f. with a voltage  $e_1$ ) is selected in the circuit being studied. A pair of terminals (a) and (b), separating the circuit into a reference source and a resistive part, is introduced, while all other sources are considered to be resistors according to the compensation theorem with positive or negative resistances  $R_{eq}$  or  $R_{js}$ .

2) A number of  $m$  transferring coefficients  $k_1, k_2, \dots, k_m$  are introduced in relation to the currents through the elements (the voltage drops upon them) having in mind KCL (or KVL).

3) One equation is created by the help of the balance of powers, where the power of the reference source is on the left hand side (LHS) and on the right hand side (RHS) the powers of the resistive elements are situated:

$$\begin{aligned} P_{j1} &= u_{j1} j_1 = const. = \\ &= R_1 (k_1 j_1)^2 + R_2 (k_2 j_1)^2 + \dots + R_p (k_p j_1)^2 + \\ &+ \sum_{q=1}^{vs} (\pm R_{eq} (k_q j_1)^2) + \sum_{s=1}^{cs} (\pm R_{js} (k_s j_1)^2) \end{aligned} \quad (1)$$

or

<sup>1</sup>Emil Panov is with the Department of Theoretical Electrical Engineering and Instrumentation at Technical University of Varna, 1 Studentska Str., Varna 9010, Bulgaria, E-mail: [epanov@yahoo.com](mailto:epanov@yahoo.com).

<sup>2</sup>Miroslava Doneva is with the Department of Theor. Electr. Eng. and Instrumentation at Technical University of Varna, 1 Studentska Str., Varna 9010, Bulgaria, E-mail: [m\\_grisheva@abv.bg](mailto:m_grisheva@abv.bg).

$$P_{e1} = e_1 i_{e1} = \text{const.} =$$

$$= G_1 (k_1 e_1)^2 + G_2 (k_2 e_1)^2 + \dots + G_p (k_p e_1)^2 + \dots \quad (2)$$

$$+ \sum_{q=1}^{vs} (\pm G_{eq} (k_q e_1)^2) + \sum_{s=1}^{cs} (\pm G_{js} (k_s e_1)^2)$$

4) A system of  $(m-1)$  equations is created by the help of KVL (or KCL) and the system is solved taking into account one of the coefficients to be a parameter, for example  $k_1$ .

5) The equation for the power (1) or (2) is differentiated by  $k_1$ :  $\frac{\partial P_{j1}}{\partial k_1} = 0$  or  $\frac{\partial P_{e1}}{\partial k_1} = 0$ . In the last equation, we have to

substitute  $R_{eq} = \frac{e_q}{i_q(k_1)}$  and  $R_{js} = \frac{u_{js}(k_1)}{j_s}$ , and that equation can be solved in respect of  $k_1$  analytically or by a suitable numerical method. Based on the obtained value of the transferring coefficient  $k_1$ , the other  $(m-1)$  coefficients are calculated, as well as the currents and voltages of the circuit.

**Example 1:** A circuit is given in Fig. 1 with the following parameters:  $R_1 = 30\Omega$ ;  $R_2 = 20\Omega$ ;  $R_3 = 10\Omega$ ;  $j = 5A$ ;  $e = 40V$ . Determine the branch currents  $i_1$ ,  $i_2$  and  $i_3$  by the variational approach.

**Solution:**

The source of e.m.c.  $j$  is selected for a reference source. The resistive part of the circuit comprises the resistances  $R_1$ ,  $R_2$ ,  $R_3$  and  $R_e = \frac{e}{i_3}$ . A transferring coefficient  $k$  is introduced for the current  $i_2$ , i.e.  $k = \frac{i_2}{j}$ , and applying formally KCL for node (1), we have:  $i_1 = j$ ,  $i_2 = k \cdot j$  and  $i_3 = (1-k) \cdot j$ .

Then, the equation for the balance of powers for the circuit is created - its LHS represents the power of the circuit's reference source, which is assumed to be a constant, and its RHS is a sum of the resistive elements' powers.

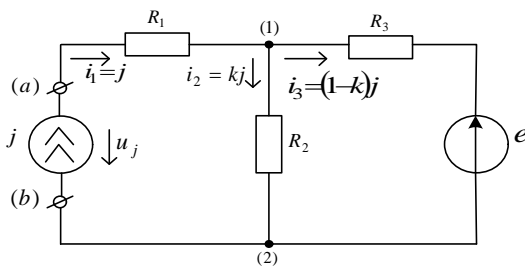


Fig. 1. The circuit from example 1.

$$P_j = u_j j = \text{const.} =$$

$$= R_1 j^2 + R_2 (kj)^2 + R_3 [(1-k)j]^2 + R_e [(1-k)j]^2 \quad (3)$$

Equation (3) is differentiated by  $k$  and the result is as follows:  $\frac{\partial P_j}{\partial k} = 0$ . After the differentiation of equation (3), we substitute  $R_e(k) = \frac{e}{(1-k)j}$  and then, it follows, that  $k = 0,6$ .

The branch currents we've been looking for in the circuit are the following:  $i_1 = j = 5A$ ,  $i_2 = k \cdot j = 3A$  and  $i_3 = (1-k) \cdot j = 2A$ .

In Fig.2 the dependence of the power of the reference source  $P_j$  versus the value of the transferring coefficient  $k$  is presented.

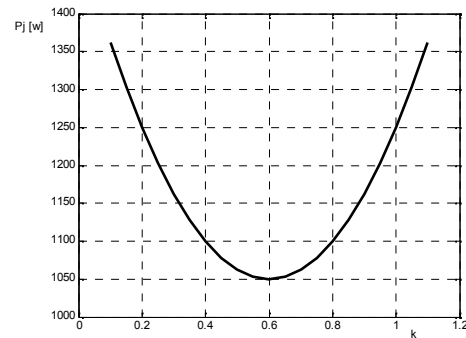


Fig. 2. The power of the reference source  $P_j$  versus  $k$ .

The minimum of  $P_j$  is very well seen from that graph.

The variational analysis of harmonic LEC is similar to the DC analysis and it must be conducted in combination with the phasor approach. The methodology for analysis of that regime needs the introduction of the complex pseudo-power of the reference source:  $\dot{S}_{j_{pseudo}}$  or  $\dot{S}_{\dot{E}_{pseudo}}$ . The methodologies, which combine the variational approach with the loop analysis or the nodal approach for both regimes (DC and AC) are presented in papers [5, 6].

#### B. SHORT THEORY OF THE VARIATIONAL ANALYSIS OF DC NON-LINEAR EC

In circuit theory there are no well-developed overall analytical methods for exploration of DC regimes by the help of the variational approach. So far, the known methods for variational approach are particular and they provide a solution to specific types of non-linear problems. The application of the variational approach with transferring coefficients for analysis of non-linear EC can be done by applying common methodologies, the same for each non-linear circuit working at DC regime. The variational analysis with transferring coefficients can be applied to any non-linear electric circuit, regardless of the number and the type of the included non-linear elements. The three methodologies for analysis, which were introduced in paper [8], allow the application of the



variational approach for analysis of non-linear EC using directly the basic laws for EC, the loop analysis or the nodal approach.

### C. VARIATIONAL APPROACH FOR ANALYSIS OF TRANSIENT PROCESSES IN NON-LINEAR EC

The analysis of transient processes in EC by the help of the variational approach needs a little bit different procedure, because the basic calculation technology is numeric one (especially for the analysis of non-linear electric circuits) [9].

The methodology for numeric variational analysis of the transient processes in one EC (for example with one reactive element) has the following steps:

1) A reference source is selected in the circuit being studied. The other sources can be substituted by positive or negative resistances  $R_{eq}(t)$  or  $R_{js}(t)$  according to the compensation theorem.

2) A number of  $m$  transferring coefficients  $k_1, k_2, \dots, k_m$  are introduced in relation to the currents flowing through the elements (the voltage drops upon them).

3) One differential equation can be created for the explored circuit of the following normalized form:

$$\frac{dk_1(t)}{dt} = f[k_1(t); R_1; \dots; R_q; C \text{ (or } L)] \quad (4)$$

which will be the predictor for the calculating process. Then, several initial values of  $k_1(t)$  can be calculated by the help of the method Runge – Kutta – 4 for the first  $p$  steps with a step size  $h$  in the time interval  $(t_0; t_p)$ .

4) Another equation can be created on the base of the balance of powers for the instantaneous powers of the elements of the explored circuit:

$$\int_{t_0}^{t_p} p_{R.E.}[k_1(t)]dt = W_{R.E.}[k_1(t_p)] - W_{R.E.}[k_1(t_0)] \quad (5)$$

where  $p_{R.E.}$  is the instantaneous power of the reactive element and  $W_{R.E.}$  is the corresponding accumulated energy. That equation will be the corrector of the calculation process. The procedure is maintained by the usage of an iteration procedure and a numeric integration formula of Newton – Cotes of rank higher than 4 (for example 6<sup>th</sup> order integration formula).

5) By the help of the predictor – equation (4), we can get the next value  $k_1(t_{p+1})$ , which we can improve by the corrector – equation (5).

**Example 2:** A non-linear circuit is presented in Fig. 3, where:  $R_1 = 120 \Omega$ ;  $R_2 = 5 \Omega$ ;  $R_3 = 0,2 \Omega$ ;  $j = 25/12 \text{ A}$ ;  $i_L(t) = 0,5 \cdot \Psi_L^2(t)$ . Find the current flowing through the coil  $i_L(t)$  by the help of the variational approach.

**Solution:**

The exact solution of that task is as follows:

$$i_L(t) = k(t) \cdot j = 0,5 \left( \frac{3 - 2e^{-250t}}{1,5 + e^{-250t}} \right)^2 \text{ A} .$$

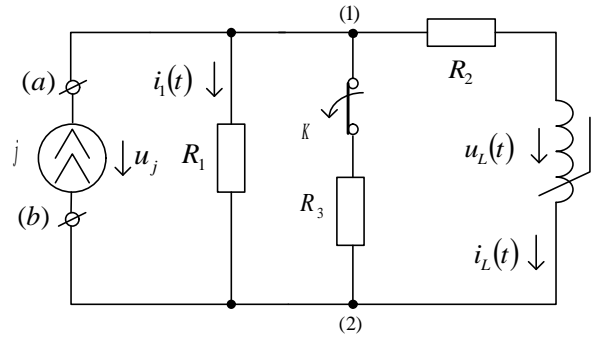


Fig. 3. The circuit from example 2.

The explored circuit contains a single non-ideal current source. The current flowing through the non-linear coil can be expressed as:  $i_L(t) = k(t) \cdot j$ .

Then:

$$u_L(t) = \frac{d\Psi_L(t)}{dt} = \frac{d}{dt} [\sqrt{2i_L(t)}] = \frac{dk(t)}{dt} \cdot \sqrt{\frac{j}{2k(t)}}$$

and

$$i_1(t) = j[1 - k(t)].$$

Using KCL and KVL we can express the voltage drop  $u_j(t)$ , i.e.  $u_j(t) = R_1 i_1(t) = R_1 \cdot j[1 - k(t)]$  and we can also find the connection among  $u_j(t)$ ,  $i_L(t)$  and  $u_L(t)$ :

$$u_j(t) = R_2 i_L(t) + u_L(t).$$

From here, we receive the predictor equation:

$$\frac{dk(t)}{dt} = \sqrt{2k(t)} \cdot j \cdot [R_1 - (R_1 + R_2)k(t)] .$$

The corrector for the calculation process will have the following form:  $\int_{t_0}^{t_6} [p_j(t) - p_{R_1}(t) - p_{R_2}(t)]dt = \int_{t_0}^{t_6} p_L(t)dt$  ,

$$\text{i.e. } k(t_6) = \left\{ 2,5\sqrt{1,5} \int_{t_0}^{t_6} [R_1 k(t) - (R_1 + R_2)k^2(t)]dt + k^{3/2}(t_0) \right\}^{2/3} .$$

Here, the numeric integration can be fulfilled by the 6<sup>th</sup> order Newton-Cotes formula. After we have already improved the six initial values of the transferring coefficient  $k(t)$ , we can improve the seventh one -  $k(t_7)$ , too.

In Fig. 4 the relative error  $d$  is presented for two calculation procedures – the Runge – Kutta - 4 method and the optimized numeric solution of the variational approach. Here,

the step size is accepted to be  $h = \frac{1}{3} \text{ ms}$  .

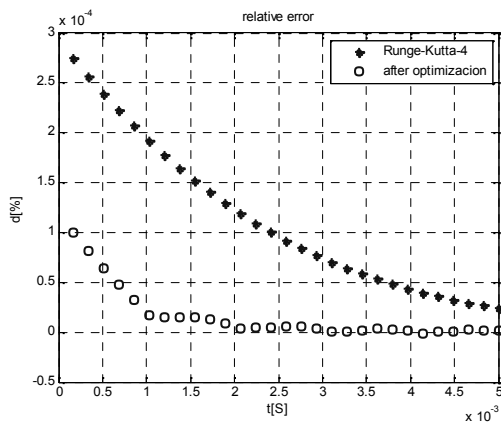


Fig. 4. The relative error d versus time t.

## CONCLUSION

The variational analysis with the transferring coefficients is a new approach for analysis of EC. May be it is the eleventh one according the classification of the authors of the paper. Its application is supported by the introduction of four new theorems for EC, two rules for introduction of the transferring coefficients and several methodologies for analysis of DC, AC regimes in EC and transient processes, too.

The variational analysis is more difficult to use compared with the well-known methods for analysis of EC, but it has something additional compared with the rest of the methods. All methods obey to two basic laws of nature - the charge conservation law (with its consequence - KCL) and the energy conservation law (with its consequence - KVL). The variational analysis except them uses also the balance of powers, which is a consequence of the least action principle, which all physical processes in nature obey to. This fact gives an additional tool to the variational approach to have an exclusive instrument to precise the errors of the calculated values of the currents and the voltages in the explored EC. Except that, the variational analysis can be accomplished in matrix form, too [3, 4] – a fact, which makes that approach able to use automated computer-aided calculations. So, circuit theory may become only more powerful after the development and the introduction of the new methodologies for variational analysis of EC by the help of the transferring coefficients.

The proposed methodology for variational analysis of transient processes in non-linear EC can be ever introduced successfully, because we can always choose a corrector of higher order, compared with the order of the predictor, having in mind that the integration equations of Newton-Cotes form an infinite family of high-precision formulas. The truncation

error of the method Runge –Kutta-4 is  $\varepsilon_{T_{R.K.}} = \frac{h^5}{5!} f^{(v)}(\xi)$ ,

where  $\xi$  is some point within the time interval of the last step

$h$  of the calculation process. The sixth order integration formula of Newton-Cotes has an error:

$$\varepsilon_{T_{N.C.}} = -\frac{h^7}{1400} [10 \cdot f^{(vi)}(\xi) + 9h^2 \cdot f^{(viii)}(\eta)],$$

where  $\xi$  and  $\eta$  are some points within the integration interval  $[t_k; t_{k+6}]$ . So, for  $h \ll 1$  it is clear, that  $\varepsilon_{T_{R.K.}} \gg \varepsilon_{T_{N.C.}}$ , which allows Newton-Cotes formulas of higher order to be in the base of the corrector integration equations.

The proposed variational approach is more complex compared with the classic solutions. It is better to use it for analysis of transient processes in non-linear electric circuits, especially in cases when there are no exact classic solutions. The optimization procedures implemented in the variational method gives the possibility to improve the numeric solutions and to increase the accuracy of the final results.

## REFERENCES

- [1] E. I. Panov, "Variation Method for Analysis of Electric Circuits", Proceeding of the Union of the Scientists - Varna, 2, 1995, 1, 1996, pp. 56 – 59, 1995.
- [2] E. I. Panov, "Exploration on the Possibilities of the Variation Method for Analysis of Electric Circuits", Proceeding of the Union of the Scientists - Varna, 2, 1995, 1, 1996, pp. 60 – 62, 1995.
- [3] E. I. Panov, "Computer – Aided Analysis of Electric Circuits by Variation Method", Proceeding of the 11-th International Conference "Systems for Automation of Engineering and Research" and DECUS NUG Seminar (SAER'97), 20-21 September, Varna, pp. 227 – 231, 1997.
- [4] E. I. Panov, "New Approaches for Implementation of the Variation Method for Analysis of Electric Circuits", Int. Conf. "Days of the Science of Republic of Bulgaria and Republic of Macedonia", 27-29 May, Sofia, Bulgaria, pp. 348 – 361, 1999.
- [5] E. I. Panov, M. Gr. Doneva, "Phasor Approach for Variation Analysis of Electric Circuits", International Scientific and Technical Conference "Electrical Power Engineering"-2014, 11-13 September, Varna, Bulgaria, pp. 71 – 80, 2014.
- [6] E. I. Panov, M. Gr. Doneva, "Exploration of the Possibilities for Variation Analysis of Complex Electric Circuits Combined with the Loop Analysis and the Nodal Approach", International Scientific and Technical Conference "Electrical Power Engineering"-2014, 11-13 September, Varna, Bulgaria, pp. 81 – 88, 2014.
- [7] E. I. Panov, M. Gr. Doneva, "Exploration of the Possibilities of the Method with the Transferring Coefficients Combined with the Variation Approach for Analysis of Electric Circuits", Annual of the Tehnical University of Varna, 2014, (in print).
- [8] Doneva M., Vasileva V., "Exploration of the Possibilities for Analysis of Non-linear Electric Circuits Working at DC Regimes Using Variational Approach", XIV-th International Conference on Electrical Machines, Drives and Power Systems – ELMA'2015, October, Varna, Bulgaria, pp. 81-90, 2015.
- [9] Panov E., Doneva M., "Exploration of the Possibilities of the Variational Approach for Analysis of Transient Processes in Electric Circuits", XIV-th International Conference on Electrical Machines, Drives and Power Systems – ELMA'2015, October, Varna, Bulgaria, pp. 305-312, 2015.

# Laboratory Equipment for Energy Accumulation from Renewable Sources

Silviya Letskovska<sup>1</sup>, Kamen Seymenliyski<sup>1</sup>, Ginko Georgiev<sup>1</sup>

**Abstract** – The goal of this work is to analyse the system for energy production using renewable energy sources. One of the main aspects is education of students. The method of work is based on production of hydrogen, accumulation and reverse transformation using photovoltaic and fuel cell.

**Keywords** – Photovoltaic, Fuel cells, Hydrogen, Electrical energy, Renewable energy sources.

## I. INTRODUCTION

The supply of electrical energy for the users can be successfully realized with the complex use of the energy from renewable energy sources (RES) and energy accumulation. This gives additional possibilities for developing of renewable energetics, particularly – nuclear stations.

The accumulation of the energy is not a new idea in energetics. The fossil sources (petrol, coal, etc.) are effective accumulators with high density of energy. As a fact, of shortage of the traditional energy sources and increasing of their price it is necessary to reach other method for accumulation.

One of the methods is production and accumulation of renewable fuel. One variant is chemical accumulation. The energy could be stored in the chemical connection.

Hydrogen energy is one of the important directions in the developing of renewable energetics. The main priorities touch dissolving of problems connected with the use of hydrogen as a energy carries and accumulators.

The hydrogen could be derived using hydrolyze of a water and as a gas could be kept, transported and used for energy. The only residual product of burning of hydrogen is water, there is no environment pollution.

The storage of hydrogen in big quantities is not easy process, even using high pressures and this requires significant volumes [1-5, 8].

Gas hydrogen is stored in balloons (in high pressure) and transported via pipes from special materials with high prices.

The direct used of hydrogen as a gas could be realized using hydrating process. Thus it could be stored for a long time and to be used in the traditional engines. In the last years there are intensive works on creating new materials for hydrogen storage.

As a perspective method for storage of hydrogen, is its adsorption in hydride metals (up to 3%) and in inter metals (up to 5%).

Hydride metals are compounds, type Me-H. The energy of destruction of the connection for same metals is low and lots of them - Ni, Cr, Co, Ti are destructed in the temperature 150÷250 °C.

High accumulating capabilities have noble metals – Pt and Pd but they are very expensive.

The practical accumulating of the hydrogen have elements from IV period group, but lots of them (from Se to Ni) form hydrides with very low concentration of hydrogen.

The modern accumulating systems based on metals give possibility to keep the hydrogen as a hydride of the metal in the ratio 1:10 by weigh.

Good results is reached with  $\text{MgH}_2$ –0,07 kg  $\text{H}_2$ /1 kg metal [3].

TABLE I  
MATERIALS FOR STORAGE OF HYDROGEN

Material	Composition	Working interval	
		T, °C	P, atm.
Metals	Mg	300-400	1-10
	V	0-200	1-200
	Ti	500-600	1-10
Alloys	Mg-Ni	250-400	1-10
	V-Cr-Mn	0-200	1-150
	Ti-Al-Ni	200-600	1-10

When store the hydrogen in a type of a hydride the volume of the systems decreases about three times compared with the volume of the balloon, transportation simplification and no expenses for liquefaction.

The best materials for use are low temperatures (working temperature: -20÷100 °C) hydride of intermetallic compounds as a type AB<sub>5</sub> (A–La, Ce; B–Ni, Co, Fe, Cu, Mn, Al), AB<sub>2</sub> (A–Ti, Zr; B–Mn, Cr, Fe, V), AB (A–Ti, Zr; B–Fe, Co) and compounds based on Vanadium which have high volume concentration of hydrogen, but not sufficient capacity as a mass (less than 3% mass.).

In Table I the data for same perspective materials are shown [6].

The base advantages of metal hydride systems are high volume density of hydrogen, good interval of working pressure and temperatures, possibility for regulation of pressure and speed of hydrogen derivation, high purity and etc.

<sup>1</sup>Silviya Letskovska - Burgas Free University, San Stefano 62, 8000 Burgas, Bulgaria, E-mail: silvia@bfu.bg

<sup>1</sup>Kamen Seimenliyski –Burgas Free University, San Stefano 62, 8000 Burgas, Bulgaria, E-mail: silvia@bfu.bg

<sup>1</sup>Ginko Georgiev - Burgas Free University, San Stefano 62, 8000 Burgas, Bulgaria, E-mail: ginkoeve@abv.bg

Including in the energy system of hydrogen accumulators of energy (complex of electrolyzer of water, contained for hydrogen and oxygen and accumulator for fuel cell) appears to be useful (Fig. 1).

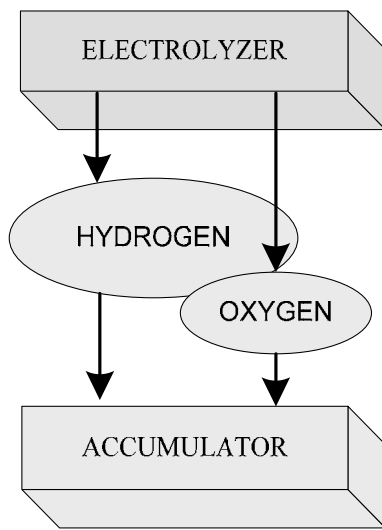


Fig. 1. Functional working diagram of fuel cells with reservoir, supplied by electrolyzer.

Such technical decision will insure creating of equipment for long time keeping of energy practically without losses. This leads to additional decreasing of system price.

Using of fuel cells in energy systems increase working capability.

The fuel cells have low toxic, low noise, variety of fuel wide interval of power (Fig. 2).

Still now the price of the energy is very high and relative low resource. The best resource, have polymer materials as PEM FC - 2÷5000 h work.

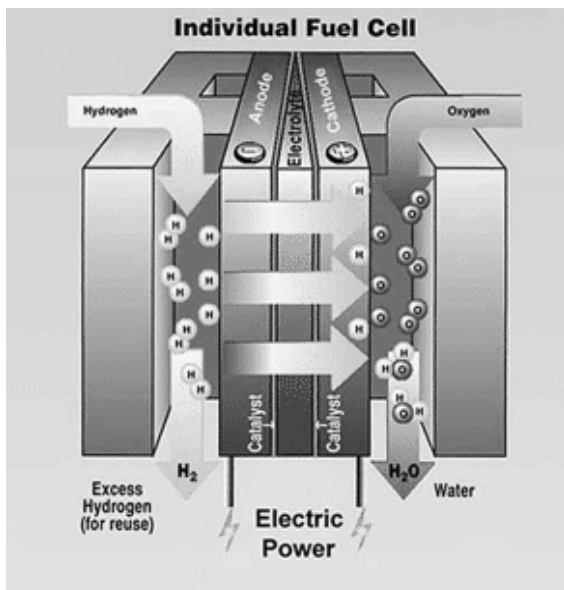


Fig. 2. Fuel cell.

The accumulating of energy with the use of hydrogen can be realized in energy system with renewable sources (wind, sun, hydro energy).

The technology for hydrogen production from primary sources includes the use of electrolyzing equipment [1]. The main work of such type systems is converting of electrical energy from renewable sources into chemical energy of hydrogen.

This energy can be used as a type of electrical or thermal energy depending on the need of the user.

The base element of the accumulating system is the equipment for hydrogen production. It is connected principally with the power of the renewable energy sources.

The use of the system electrolyzer – fuel cell gives the possibility to realize significant accumulating effect.

The problem in RES is the fact, that energy production depends on the meteorology conditions and it could not be planed.

The fuel cell (FC) is a electrochemical energy current source, in which the direct transforming of energy from fuel and oxide.

This process is a result of uninterrupted supply of fuel to the electrodes (Fig. 2).

In hydrogen fuel cell the distribution is as follows: hydrogen to anode, oxygen to cathode.

Fuel cell with proto reaction, rapidly reach to the maximum working power and they generate current with high density – 2A/cm<sup>2</sup>.

They have high dynamic range – after turning on FC immediately come in the nominal power from several microwatts to hundred wats.

## II. INVESTIGATION THE POSSIBILITIES FOR HYDROGEN PRODUCTION USING PV-ENERGY

### 1.1. Experimental equipment.

The investigation of the possibilities for accumulating an energy using system electrolyzer – fuel cell was done in the laboratory of Faculty of computer science and engineering, Burgas Free University (BFU).



Fig. 3. PV-central on the roof of BFU.



This laboratory has different equipment as follows:

- PV-central (Fig. 3), mounted in the roof of the university with six photovoltaic panels and total installed power 1428 Wp (6x238 Wp).

The roof space gives the possibility 50 PV-panels to be installed. The central can work in two modes: as a direct sublayer of electricity for the needs of BFU building or for hydrogen production;

- Meteostation - BFU-METEO, system for monitoring the parameters of sun radiation, temperature, wind in real time [7].

The system has possibilities for writing, storage and analyzing of data received from the sensors.

The diagrams and graphs are used; file exportation of data for the concrete period of time is supported.

- Electrolizer – model HYDROFILL-FCH-010 from the company “Horizon Fuel Cell Technologies”. It is polymer electrolyte fuel cell with reverse action [9-11].

The produced hydrogen is kept in special reservoir-cylinders from alloy type AB2 for hydrogen absorption.

Represent a group of AB2 alloys containing titanium, zirconium or hafnium as a base and a metal (nickel, chromium, vanadium) in the lattice structure. Has the ability to absorb hydrogen to form hydrides full and saturation. The internal pressure of fully loaded cartridge remains 30 Bar (435 pounds per square inch-PSI).

The electolizer is supplied from PV, then the system is fully autonomous.

- Fuel cell, model H-12 (H-20) (FCS-B12) from the company “Horizon Fuel Cell Technologies”.

It is a module of fuel cells type PEM FC and includes thirteen serial connected single cells with total power 20 W [9].

### 1.2. Investigation the yield of hydrogen using electrical energy from PV central.

With the help of meteostation BSU-MS the experimental data for sun radiation for seven months were received (Table 2).

TABLE II  
MONTHLY VALUES OF SUN RADIATION

Sun radiation (kWh/m <sup>2</sup> )	BFU-MS	PVGIS
August	187,15	211,73
September	135,17	163,8
October	84,62	123,38
November	36,43	82,5
December	36,31	59,52
January	63,85	66,34
February	75,65	82,04

The data were compared with the information automated system PVGIS gives for a point with geographical coordinates of Burgas Free University.

On the base of received data for the intensity and duration of sun radiation in the time of investigated period the theoretical values of generated energy were determined.

On Fig. 4 the radiation data for chosen day with the biggest intensity (19 august) are shown.

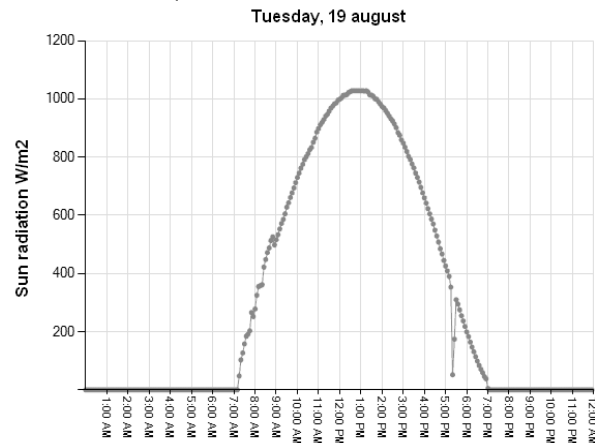


Fig. 4. The radiation data for chosen day with the biggest intensity.

TABLE III  
GENERATED ELECTRICAL ENERGY (KWH) PER MONTH

Month	Number of panels			
	1	6	10	50
August	22,271	133,625	222,709	1113,543
September	16,085	96,511	160,852	804,262
October	10,070	60,419	100,698	503,489
November	6,069	36,415	60,692	303,462
December	6,049	36,295	60,492	302,462
January	10,637	63,824	106,374	531,871
February	12,603	75,620	126,033	630,165

The quantity of the energy, which could be received from the panels was determined using Eq.1.

$$W = (k \times P_w \times N) / 1000 \quad (1)$$

Where:

- W – produced electrical energy (W/h);
- k – coefficient reading the power losses as a result for photovoltaic heating, as well as and angle change of sun beams dropping onto the photovoltaic during

the day. For winter months the value is 0,7; for summer the value is 0,5;

- $P_w$  – power of the panel;
- $N$  – number of panels.

The received results are shown in Table III.

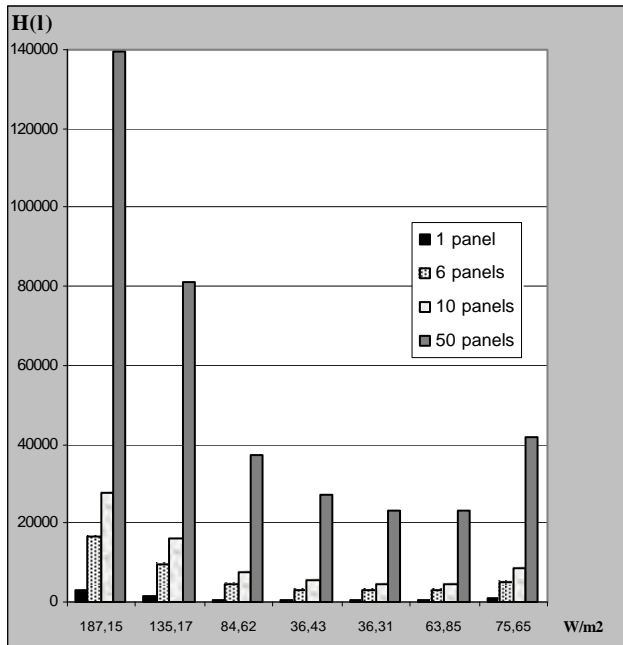


Fig.5. Data for hydrogen in liters, produced by different number PV panels depending on sun monthly radiation.

Theoretical calculation for quantity hydrogen, which can be produced using electrolyzer were done.

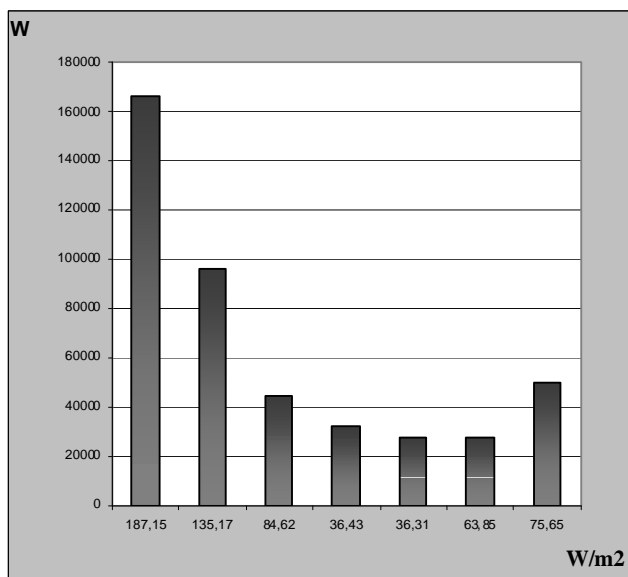


Fig. 6. Quantity electrical energy from the accumulated hydrogen.

The choice of the electrolyzer power is in connection with the quantity electrical energy produced for one hour from PV.

The data for hydrogen which can be produced with different number of panels for months with experimental data for sun radiation are shown on Fig. 5.

The accumulated hydrogen could be used for production of electrical energy with the help of FC.

Data for quantity electrical energy which can be received from the accumulated hydrogen are shown on Fig. 6.

### III. CONCLUSION

So proposed complex systems, is suitable for practical student (laboratory) classes.

It gives the possibility to follow and analyze a cycle of technological processes – production of electrical energy from renewable energy sources, accumulation and reverse converting in electrical.

The other advantage of the proposed system is the possibility the produced electrical energy to be used for the needs of university building.

Independently that the effectiveness of transformation hydrogen into electricity is not high, the received energy successfully could be used in emergency switch off of power supply – for supporting of servers, emergency light etc.

### REFERENCES

- [1] N. M. Mkhitarian, S. A. Kudrya, L. Yatsenko, L. Y. Shinkarenko, M. D Tkalenko, V. I. Budko, Using Integrated renewable energy sources, Scientific Journal for Alternative Energy and Ecology № 17 (139) 2013.
- [2] Hydrogen Fuel Cell Engines and Related Technologies: College of the Desert, Palm Desert, CA, USA Energy Technology Training Center, Rev 0, December 2001.
- [3] Tarasov B. P, Lototsky M. V., Hydrogen Energy: Past, Present and prospects for the future, Ros. Chem. Well. (J. Roth. Chem. Mendeleev on the islands it.), 2006.
- [4] Materials for Hydrogen Energy, Ural State University. A. M. Gorky, Faculty of Chemistry, Ekaterinburg, 2008.
- [5] Tarasov B. P, Burnacheva V. V., Lototsky M. V., Jartis V. A., Hydrogen method meals and opportunities to use metal hydrides, © 2005 Scientific Technical Centre «TATA».
- [6] Survey methods meals hydrogen, Institute problem material science NAN Ukraine.
- [7] Silvija Letskovska and Kamen Seymenliyski, Monitoring system for evaluation solar energy potential, Burgas Free University, ISSN: 1311-221-X, vol. XXX, 208-214.
- [8] Eldar Zaerov, Sylvia Letskovska, Kamen Seymenliyski, Burgas Free University-energy producer, National Conference with international participation "Horizons First human resource development and knowledge", Burgas, Burgas Free University, 2015, t. II, pp. 396-401.
- [9] <http://www.horizonfuelcell.com>
- [10] <http://shp.by.ru>
- [11] Guide, Hydrogen. Properties, receiving, steering, transportation, application, „Moscow "Chemistry" – 1989.

# Project Based Learning

Danica Milosevic<sup>1</sup>

**Abstract** – This article discusses the nature of learning in PBL and examines the empirical evidence supporting it. This paper shares some relevant published evidence that active problem solving by learning teams contributes to critical thinking and effective team communications, which can be carried forward into professional roles in the work place. The curiosity and inspiration that inspired this paper was a research question “How does facilitated problem-based learning benefit a team of distributed learners when collaborating on designing and producing a practical solution?”

**Keywords** – Project Based Learning, Self Directed Learning, Matlab Ford-Fulkerson algorithm, Heuristicslab.

## I. INTRODUCTION

The old-school model of learning facts passively and reciting them out of context is no longer sufficient to prepare students to survive in today’s world. Solving highly complex problems requires students to have both fundamental skills and Digital Age skills.

This paper begins with the basic thoughts on Learning based on the Project, as part of the theory of learning from experience. To develop such attitudes and the theory, a long period of time was needed over which knowledge has been divided into theoretical and practical knowledge, with these two segments being integrated only in recent times. In this approach to learning, knowledge is considered to be the product of experience.

This approach is broadly applicable in scientific research where the emphasis is not placed on learning, but rather on problem solving and knowledge transfer between members of the research group. PBL is a big part of modern learning environments. It takes students on a cool journey of creativity, exploration, and real-world relevance.

Problem-based learning (PBL) is an instructional method in which students learn through facilitated problem solving. In PBL, student learning centers on a complex problem that does not have a single correct answer. Students work in collaborative groups to identify what they need to learn in order to solve a problem. They engage in Self-Directed Learning (SDL) and then apply their new knowledge to the problem and reflect on what they learned and the effectiveness of the strategies employed. Problem Based teaches students 21 st century skills as well as content. These skills include communication and presentation skills, organization and time management skills, research and inquiry skills, self-assessment and reflection skills, and group participation and leadership skills. By bringing real-life context and technology to the curriculum through a Project-based Learning approach,

students are encouraged to become independent workers, critical thinkers, and lifelong learners. If students learn to take responsibility for their own learning, they will develop in the way to work with others in their adult life. Project-based Learning is not just a way of learning, but a way of working together, Fig.1.

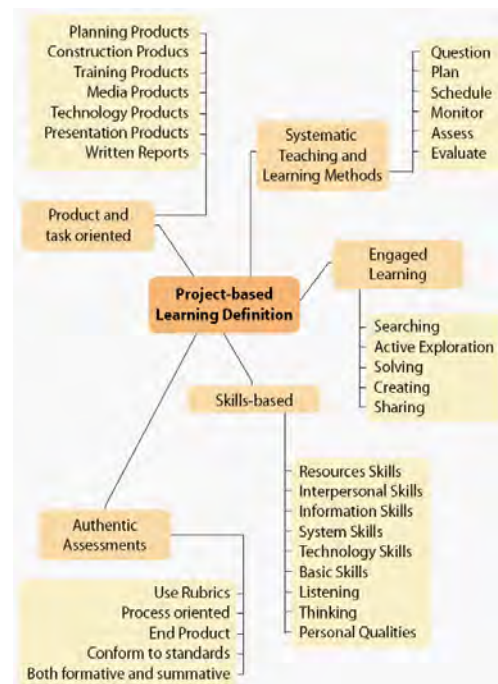


Fig. 1: Project Based Learning at a glance

For many students, the appeal of this learning style comes from the authenticity of the experience. Students take on the role and behavior of those working in a particular discipline. Whether they are making a documentary video about an environmental concern, designing a travel brochure to highlight sites of historical significance in their community, or developing a multimedia presentation about the pros and cons of building a shopping mall, students are engaged in real-world activities that have significance beyond the classroom.

The teacher acts to facilitate the learning process rather than to provide knowledge. Teachers who are used to traditional instructional methods know how to manage a classroom when they’re directing a lesson, supervising individual assignments, or leading a discussion. But managing the varied and more team-based, lively activities found in a project might seem daunting to teachers new to PBL.

Project-based Learning is a model for classroom activity that shifts away from the usual classroom practices of short, isolated, teacher-centred lessons. PBL learning activities are long-term, interdisciplinary, student-centred, and integrated with real-world issues and practices. It is a method that fosters abstract, intellectual tasks to explore complex issues. It

<sup>1</sup> Danica Milosevic is with the College of Applied Technical Sciences University Nish, A. Medvedeva 20, Nish 18000, Serbia, E-mail: danicamil@yahoo. com

promotes understanding, which is true knowledge. In PBL, students explore, make judgments, interpret, and synthesise information in meaningful ways. It is more representative of how adults are asked to learn and demonstrate knowledge.

The goals of PBL include helping students develop:

- 1) flexible knowledge,
- 2) effective problem-solving skills,
- 3) SDL skills,
- 4) effective collaboration skills, and
- 5) intrinsic motivation.

## II. HOW TO IMPLEMENT PROJECT-BASED LEARNING?

A basic and simple overview of the PBL Process is illustrated in Fig. 2, which starts with a buy-in decision that the PBL protocol will add value to a lesson agenda and ends with the problem being solved and documenting a lesson learned. As seen by the arrows, steps are looped until the results are satisfying.

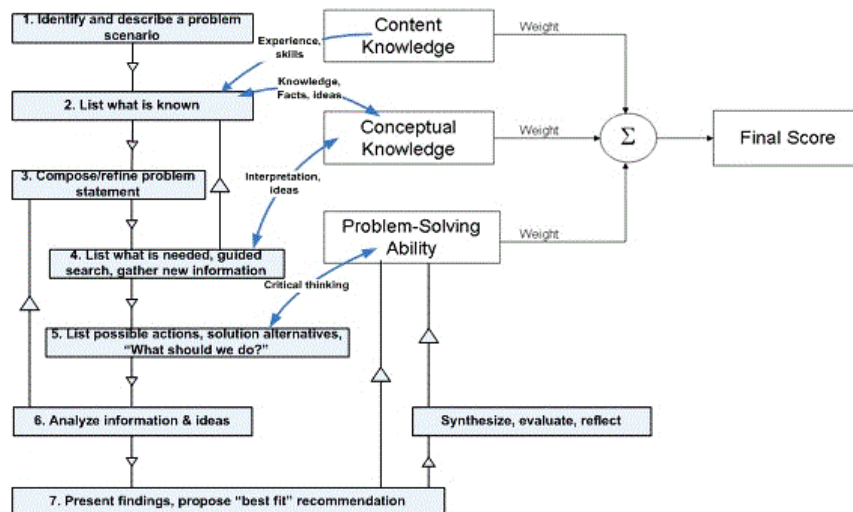


Fig 2. PBL action steps, integrated with course assessment.

Real PBL, by contrast, is deep, complex, rigorous, and integrated where each stakeholder in the school plays an important role. In implementing PBL, its fundamentals are fourfold:

- ✓ Create teams of three or more students to work on an in-depth project for three to eight weeks.
- ✓ Introduce a complex entry question that establishes a student's need to know, and scaffold the project with activities and new information that deepens the work.
- ✓ Calendar the project through plans, drafts, timely benchmarks, and finally the team's presentation to an outside panel of experts drawn from parents and the community.
- ✓ Provide timely assessments and/or feedback on the projects for content, oral and written communication, teamwork, critical thinking, and other important skills.

Characteristics on which project based learning is based are:

- ✓ Any group or individual, within a set of problems, getting special tasks to solve independently,

- ✓ It is desirable that the group is small, or heterogeneous, so that the advanced students can help less advanced students,
- ✓ Project is organized around topical issues that the teacher and students find important,
- ✓ Content topics are organized around problems, not about discipline,
- ✓ Students manage the process of learning. The group carried out field research, studying literature and the like.

The project sets the student in the center of attention, as well as the promotion of his creative talents and possibilities. The student has a maximally active role. When the student is intellectually and functionally active, then he is intellectually developed.

Students address the terms of reference used by the same mindset and the same procedures that are used in science. The project continues to develop the skills necessary in life and work of adults and in everyday life.

Scenario consists of several phases, Table I.



TABLE I  
PHASES OF PBL FOR SCENARIO PLANNING

Phase	Purpose
Goal Description	Define a problem scenario to demonstrate course learning outcomes. Establish an ultimate goal for team project.
Specify Criteria	Specify criteria to be met. What is the focus of the testing? How will you know when you have reached your target?
Background Knowledge	Identify knowledge needed to accomplish the goal. Ask experts? Research and adapt best practices.
Generate Ideas	Brainstorm & generate ideas.
Implement Solution	Generate, develop and implement an adapted solution(s) for verifying the requirements are validating the system.
Reflect	Evaluate and reflect on your solution(s) and on the process we went through the lesson learned.
Generalize	Conceptualize, integrate and generalize previous knowledge about system and new learning.

Reflection by students is a time of pause to decide to stay in the loop or to complete the exercise because enough is known, or the time has expired. Reflecting on the relationship between problem solving and learning is a critical component of PBL and is needed to support the construction of extensive and flexible knowledge (Salomon and Perkins, 1989). This reflection should help learners understand the relationship between their learning and problem-solving goals. Reflection helps students to relate their new knowledge to their prior understanding, mindfully abstract knowledge, and understand how their learning and problem-solving strategies might be reapplied.

Project-based learning offers a wide range of benefits to both students and teachers. A growing body of academic research supports the use of project-based learning in school to engage students, cut absenteeism, boost cooperative learning skills, and improve academic performance (George Lucas Educational Foundation, 2001). For students, benefits of project-based learning include:

- Increased attendance, growth in self-reliance, and improved attitudes toward learning (Thomas, 2000)
- Academic gains equal to or better than those generated by other models, with students involved in projects taking greater responsibility for their own learning than during more traditional classroom activities (Boaler, 1997; SRI, 2000)
- Opportunities to develop complex skills, such as higher-order thinking, problem-solving, collaborating, and communicating (SRI, 2000)
- Access to a broader range of learning opportunities in the classroom, providing a strategy for engaging culturally diverse learners (Railsback, 2002)

### III. SOLVING PROBLEM OF THE COLLABORATIVE GROUPS

Collaborative problem-solving groups are a key feature of PBL. One assumption of PBL is that the small group structure helps distribute the cognitive load among the members of the group, taking advantage of group members' distributed expertise by allowing the whole group to tackle problems that would normally be too difficult for each student alone. In PBL groups, the students often work together to construct collaborative explanations. Most PBL groups need some help to collaborate effectively.

To solve this problem we can use heuristic method and a modified Ford-Fulkerson algorithm. This algorithm we use to make workflow formation model of heterogeneous groups of students who are learning in a team. To make this possible, it is necessary to create a MATLAB mathematical model. In this model the heterogeneity refers to the characteristics of students. We have taken into account the following features that are easy to identify on the basis of surveys and preliminary results of students: interest in the subject, the performance of the studies (the average), motivation, attitude toward work in a group performance studies on matters related to the subject of learning, Table II.

In order to simplify each of these characteristics a range of values from 1 to 5 is defined where 1 means a low and 5 means a high grade. In this way, each student presents a feature vector whose value ratings from 1 to 5. For example, student  $S_1$  can be represented as  $S_1 (5, 5, 5, 5, 5)$ . Total student score on the basis of heterogeneity weights is simply the sum of the elements of the vector characteristics.

TABLE II  
CHARACTERISTICS OF STUDENTS

Students Group	Interest for the Course	Average Score	Motivation	Attitude toward work in a Group	Matters related to the Subject of Learning	Average
S1	5	5	5	5	5	5
S2	3	4	3	3	4	3,4
S3	4	4	4	4	4	4
S4	3	3	3	5	3	3,4
S5	3	2	3	4	2	2,8
S6	2	2	4	5	3	3,2
S7	2	1	3	5	2	2,6

Fig. 3 shows vector distribution of the students group:

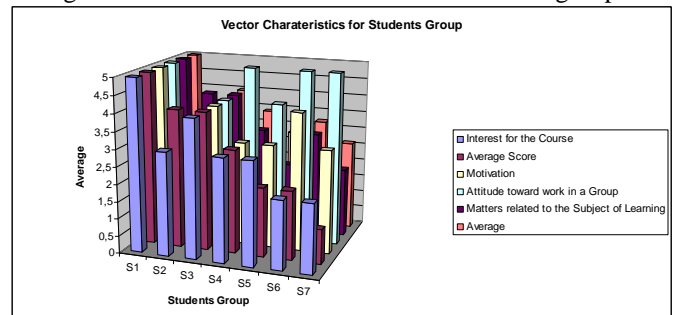


Fig 3. Vector characteristics for Students Group.

In this case, the matrix of mutual links looks like:

Source	1	1	1	3	2	4	5	6	7
Destination	4	6	7	5	4	6	1	2	6
Weights (exaple)	4	3	2	2	3	3	2	3	3

In Matlab we introduce described connected links:

```
cm = sparse([1 1 1 3 2 4 5 6 7],[ 4 6 7 5 4 6 1 2 6],[ 4 3 2 2 3 3 2 3 3 3 2 3],7,7)
cm=
(5,1) 2; (6,2) 3; (1,4) 4; (2,4) 3; (3,5) 2;
(1,6) 3; (4,6) 3; (7,6) 3; (1,7) 2
```

And draw the diagram group, Fig. 4:

```
h = view(biograph(cm,[],'ShowWeights','on'))
```

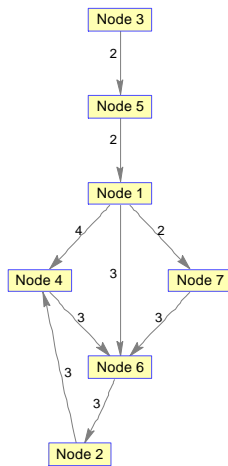


Fig 4. PBL action steps, integrated with course assessment.

Now we are introducing a measure of the quality of the heterogeneity of the group according to the following principle: if from group of  $N$  students we remove the best and worst student on the basis of characteristic vectors, the results of other students should be somewhere in the middle compared to these two "border" students. It can be shown that the following formula for quality and heterogeneity of the group  $i$  or  $GH_i$ , gives greater value as a group is more heterogeneous.

The measure of  $GH$  can be computed as follows. Let  $ADi$  be the average of the maximum and the minimum student-score in the  $i$ -th group.

$$AD_i = \frac{\max score(S_1, S_2, \dots, S_n) + \min score(S_1, S_2, \dots, S_n)}{2}$$

The measure of goodness of heterogeneity is then defined as:

$$GH_i = \frac{\max score(S_1, S_2, \dots, S_n) - \min score(S_1, S_2, \dots, S_n)}{1 + \sum_j |AD_i - score(S_{j(i)})|}$$

where  $S_1, S_2, \dots, S_n$  are students who belong to the group  $i$ ,  $ADi$  represents the mean of the results of the best and worst student of the group  $i$  and a sum of  $j$  is the index of student

group, wherein the index does not take value of the best and the worst student. If only large values of  $GH$  were pursued, it would lead to forming groups with extremely high coefficient of heterogeneity and groups with extremely low coefficient so that the final formula includes a coefficient of variance  $CV$  and then suitable algorithm is used to perform optimization for this formula. The final formula looks like this:

$$F = \omega_{GH} GH + \omega_{CV} CV$$

where  $GH$  is the sum of  $GH_i$  for all groups, and  $CV$  is the coefficient of variance that represents the ratio of the standard deviation and the arithmetic mean of the results of all students, and  $w$  represents the weight. The process of optimization of this formula with the help of a Heuristiclab boils down to the definition of external tool for evaluation of solutions that are in fact based on a predefined final formula. Since this is the iterative process of finding solutions, this function is called for with each iteration. Next, we can select a suitable algorithm to optimize and adjust its parameters. In the end, it is necessary to carry out simulations with the modification of parameters (input, the algorithm and the weight of the formula) to provide a better solution whereby the operation can be performed using a suitable graphic interface of this environment.

## IV. CONCLUSION

Given the current state of research on Project-Based Learning, what can we conclude about the relative merits of PBL as a teaching and learning method? Research on PBL implementation is largely limited to research on project-based science administered by teachers with limited prior experience with PBL. There is some evidence that students have difficulties benefiting from self-directed situations, especially in complex projects. The effectiveness of PBL as an instructional method may depend, to a greater extent than we recognize, on the incorporation of a range of supports to help students learn how to learn. There is direct and indirect evidence, both from students and teachers, that PBL is a more popular method of instruction than traditional methods. Additionally, students and teachers both believe that PBL is beneficial and effective as an instructional method.

## REFERENCES

- [1] PROJECT-BASED LEARNING HANDBOOK, Educational Technology Division, Ministry of Education, Kuala Lumpur, Malaysia, 2015.
- [2] Barbara Lauridsen, Problem-based Learning applied to Team Environments: Capella University, USA, 2016..
- [3] Designing Effective Projects: Characteristics of Projects Benefits of Project-Based Learning, Intel® Teach Program, 2014
- [4] Cindy E. Hmelo-Silve, Problem-Based Learning: What and How Do Students Learn? Educational Psychology Review, Vol. 16, No. 3, September 2004
- [5] John R. Mergendoller, John W. Thomas, Managing Project Based Learning: Principles from the Field, Buck Institute for Education, Novato, California, USA, 2016.

# Collaborative Learning in Higher and Secondary Education

Tania Vasileva<sup>1</sup>, Stela Stefanova<sup>2</sup> and Vassiliy Tchoumatchenko<sup>3</sup>

**Abstract** – The paper considers introduction of collaborative learning approach in university and secondary school settings. Two case studies of re-designed compulsory courses to promote students' knowledge work practices by following the trialogical design principles are presented. The results from pilots courses conducted in ASIC Design and Networking Technologies are also highlighted.

**Keywords** – Collaborative learning, Knowledge work practices, Computer cloud technologies.

The paper discusses efforts done at the Department of Electronics of the Technical University of Sofia (TU-Sofia) and the Technology School "Electronic Systems", associated with the TU-Sofia to reconstruct pedagogical practices to promote necessary competences. Two case studies of applying trialogical learning approach [3] for university and secondary school education are discussed. The results from pilots conducted in the ASIC Design and Networking Technologies courses are also highlighted.

## I. INTRODUCTION

In an increasingly global economy, effective creation, use and dissemination of knowledge is the key to success. In order to manage changes in the society and in the work life, new types of competencies are needed, such as collaborative learning, cultural awareness, self-leadership and flexibility. Widespread pedagogical methods and practices do not usually support these new challenges because the focus is on content learning rather than on fostering higher-order knowledge work competencies [1]. Many educational analysts and industry representatives report that students leave higher education with underdeveloped abilities to collaborate, manage their work processes, use computers, or solve open-ended problems [2]. Current technology now offers many more possibilities for sharing, archiving, retrieving, combining and generating new knowledge. Particularly in secondary education, technology is used infrequently, only by some teachers and often for previously established teaching methods but not for transforming practices.

Rapid changes and demands of the knowledge society, acceleration of technology and networking challenge educational institutions to reconsider pedagogical practices to ensure that students acquire necessary future competencies during upper secondary and university education. New strategies are needed for introducing pedagogical models addressing the use of collaborative work and cloud information technology into the educational systems. Students need interdisciplinary, goal-oriented projects, where cross-fertilization between schools, higher education institutions and professional organizations is enabled.

<sup>1</sup>Tania Vasileva is with the Faculty of Electronic Engineering and Technologies at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: tkv@tu-sofia.bg.

<sup>2</sup>Stela Stefanova is with the Technology School "Electronic Systems" associated with Technical University of Sofia, 2 Vl. Pashov Str, Sofia 1750, Bulgaria, E-mail: sstefanova@elsys-bg.org.

<sup>3</sup>Vassiliy Tchoumatchenko is with the Faculty of Electronic Engineering and Technologies at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: vpt@tu-sofia.bg

## II. COURSES RECONSTRUCTION

In order to promote the development of new competencies through education, specific design principles based on the trialogical approach were created for supporting the design of pedagogy [4].

The problem was how to re-design our courses to better promote students' knowledge work competencies and how to implement the trialogical design principles in own teaching in order to increase motivation of students to learn, to enhance their professional competences and soft skills:

- Abilities to resolve complex problems;
- Better practical training (experience in using professional software and CAD tools);
- Working on multidisciplinary tasks, utilizing multiple knowledge sources;
- Abilities to work in group;
- Abilities to use modern computer technologies and environments;
- Networking.

In order to achieve these objectives and resolve problems a trialogical educational approach was introduced with using cloud computing technologies, up-to-date communication tools for student-teacher connection, continuous monitoring and assistance students' activities. We decided to reconstruct the courses and to adapt design principles as summarized below:

DP1: Organizing activities around shared object - Collaborative development of a semester long project and, preparation of a shared report

- Activities: face to face meeting for the distribution of project tasks, defining and preliminary review of the tools used; getting students acquainted with the phases of the design cycle.
- Team organization: students are free to choose team partners.

DP2: Supporting integration of personal and collective agency and work

- Coordinate the participants' interests – team members to choose an appropriate project they want to develop and are interested in.

- Motivate students to manage tasks distribution between team members, having respect to deadlines by giving each student the responsibility of being a team leader in turn. There are two strictly defined deadlines during the project cycle and performing tasks within given deadlines is one of the criteria for project evaluation.
- Combining participants' own interests and shared project through assessment process – the quality of the shared project and the responsibilities concerning deadlines are evaluated with a higher grade.

DP3: Emphasizing development and creativity through knowledge transformations and reflection

- Support versatile use of various kinds of knowledge: theoretical or literary sources; practical examples and cases; pictures, models. Discuss problems student faced during collective work on the shared object.
- Reflect on collective practices and knowledge.
- Make students comment on each other's work through the semester.
- Encourage students to independently and creatively initiate, lead and manage a development process.
- Require student to apply already obtained knowledge and skills in using professional CAD tools to resolve complex practical tasks.

DP4: Fostering long-term processes of knowledge advancement

- Continuous working process, iterative design process to improve circuit and system performance in order to fulfill technical specifications.
- Planning and start writing the documents, sharing the drafts, getting feedback from the teacher and other students, improving the project and project report, submitting relevant documents.
- The best final team projects and their shared reports are used during the course as good examples for other teams, as well as after the course by the students themselves or by others.

DP5: Promoting cross-fertilization of knowledge practices and artifacts across communities

- Students use professional tools for projects development. They are provided with professional work models and design flow cycle used in the software and electronic industry.
- Expert practices are modeled for students, via templates and tools.
- Students and teachers collaborate on solving a shared problem.
- Students use up-to-date cloud computing and communication tools for planning, organizing and writing shared reports.

DP6: Providing flexible tools for developing artifacts and practices

- Google Drive for collaborative authoring of the project reports, reviewing and commenting.
- Google calendar – to set deadlines and to monitor progress – assignments, intermediate stages reporting, deadline for submission of the project.
- For inter-team communications students can choose their preferred tools (chat, conferences, e-mail, forums).

- For student – teacher communications Google applications are used: Gmail, Calendar, Drive.

Applying these design principles during the last two years, several engineering courses were re-designed at the TU - Sofia and the Technology school "Electronic Systems" from traditional face-to-face to collaborative project oriented adopting and applying modern online learning platforms, cloud collaboration tools, and social software. In the next section are discussed two cases in which teachers promote students' knowledge work competences at university and secondary school education.

### III. ASIC DESIGN COURSE

The bachelor degree ASIC Design course was transformed to project oriented following trialogical design principles in order to give students opportunity to work collaboratively in group with clear role of each participant. The infrastructure for collaborative learning consists of public cloud based services, combined in a way that supports electronic design workflow (Fig. 1).

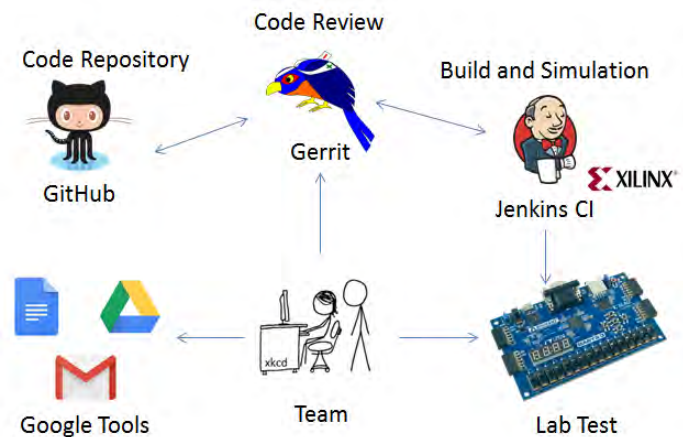


Fig.1. Collaborative workspace

Working in small teams, the students are required to design a digital Application Specific Integrated Circuit (ASIC). The design workflow is based on HDL (Hardware Design Language) modelling, verification and synthesis. The main design artefacts (HDL models and test-benches) are text files; therefore we are able to borrow many tools and workflows from the software development community. Projects are hosted on GitHub [5] – one repository per project.

In parallel with the code development, the teams are required to create and maintain a Google Docs document which is one of the major deliverables. Initially the document contains the technical specifications of the design. Later on, the students have to add description of the implemented algorithms and architectures, argumentation of the tradeoffs made and the results from the simulation, synthesis and physical design.

Most of development takes place outside the regular classes. For their intra-team communication, the students are free to choose whatever tools they prefer (chat, conferencing, email). For student - teacher communications we decided to



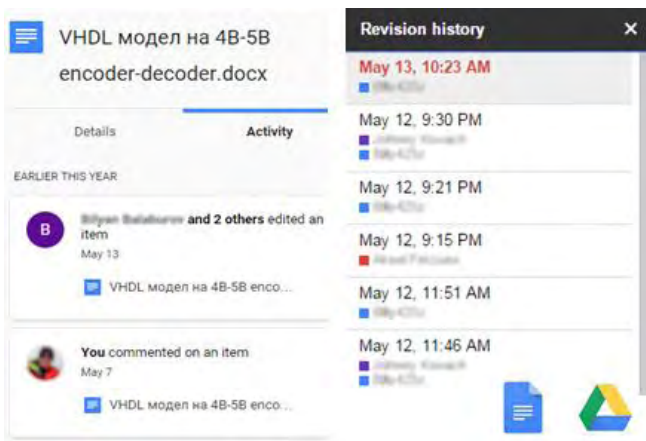


Fig. 2 Collaborative artefact evolution – Google Docs/Drive

use the Google tools. Students were encouraged to submit their questions as in-document comment or email.

Collective artefacts evolution is followed through GitHub Revision History and Google Docs Revision History (Fig.2). GitHub is optimized for hosting software projects. It provides a very detailed history of commits for each repository. In the majority of cases, the tools provided by GitHub are more than adequate for analyzing the evolution of the student's projects.

The functionality offered by Google Docs with respect to exploring documents history is rather limited. At a file level, there is an *Activity* view, that provides a good overview of when and who created or modified a particular document. At document level, we have a *Revision history*, which shows a timeline of the changes, but no information about the scope of each change.

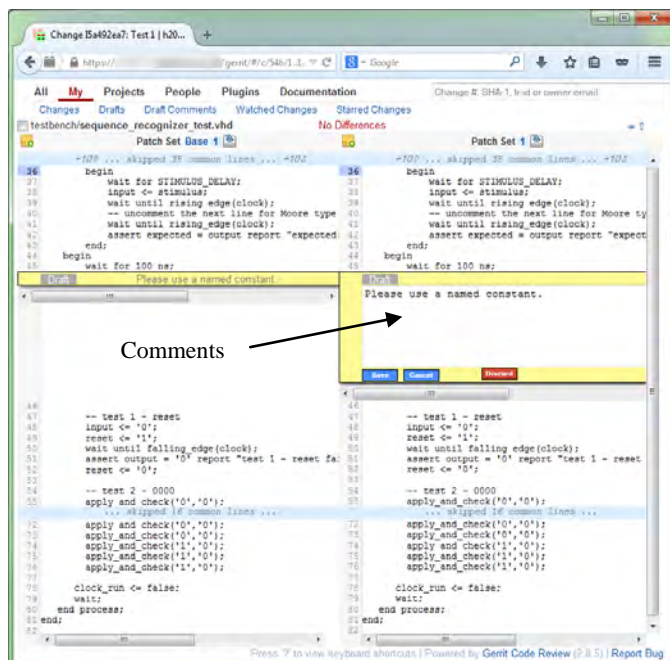


Fig. 3. Peer review

An effort is made to follow the test driven development process - first crate a test-bench then the model that makes all the tests pass. Students are aware that the comprehensive test

coverage will be one of the primary project evaluation and scoring criteria. All tests-benches should be self-checking - i.e. no "manual evaluation" of the simulation results should be required to determine the correctness of the model.

Team members have a collaborator rights for the respective repositories, but they were asked not to commit directly. Each change had to be peer reviewed before it can be committed to the project repository. When a team member submits a change for code review, the project is automatically built and the tests are executed. Another team member shall perform a code review and either approve the change or return it to the submitter for rework. System allows the reviewer to attach comments to a source code file or a particular line inside the file (Fig. 3).

The pilots were conducted with fourth year bachelor students. Each team had to choose a project subject from a list provided by the teacher. Two project milestones were set – intermediate report and final report. Each class had a Google calendar with all relevant milestones and class schedules. Project development in such practice permits for self-selected time and place allocation of the participants and teachers.

## IV. GLOBAL NETWORKS COURSE

This case is an example of introducing university type of education in secondary school. Instead of giving 12<sup>th</sup> grade students many separate or loosely connected tasks strictly developed by the teacher we provide them with a 31 weeks long project. All group activities are organized around shared objects – collaboratively development of common project, preparation of shared report, continuous working process, and final presentation in teams. Secondary school students have to perform collaborative network design (Fig.4). In addition to the collaborative project work, they were required to submit several individual homework assignments.

Guidance is provided through systematic instructions and group work rules. Assessment includes process and product assessments, and contribution evaluation of each participant to the collaborative project development.

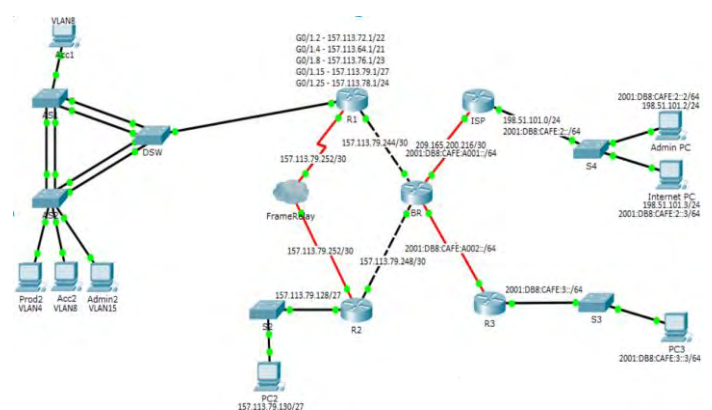


Fig. 4 Collaborative network design

Collaborative workspace for network design is shown in Fig. 5. Successful project development requires continues working process with iterative network configuration, performing numerous analysis of the designed network to

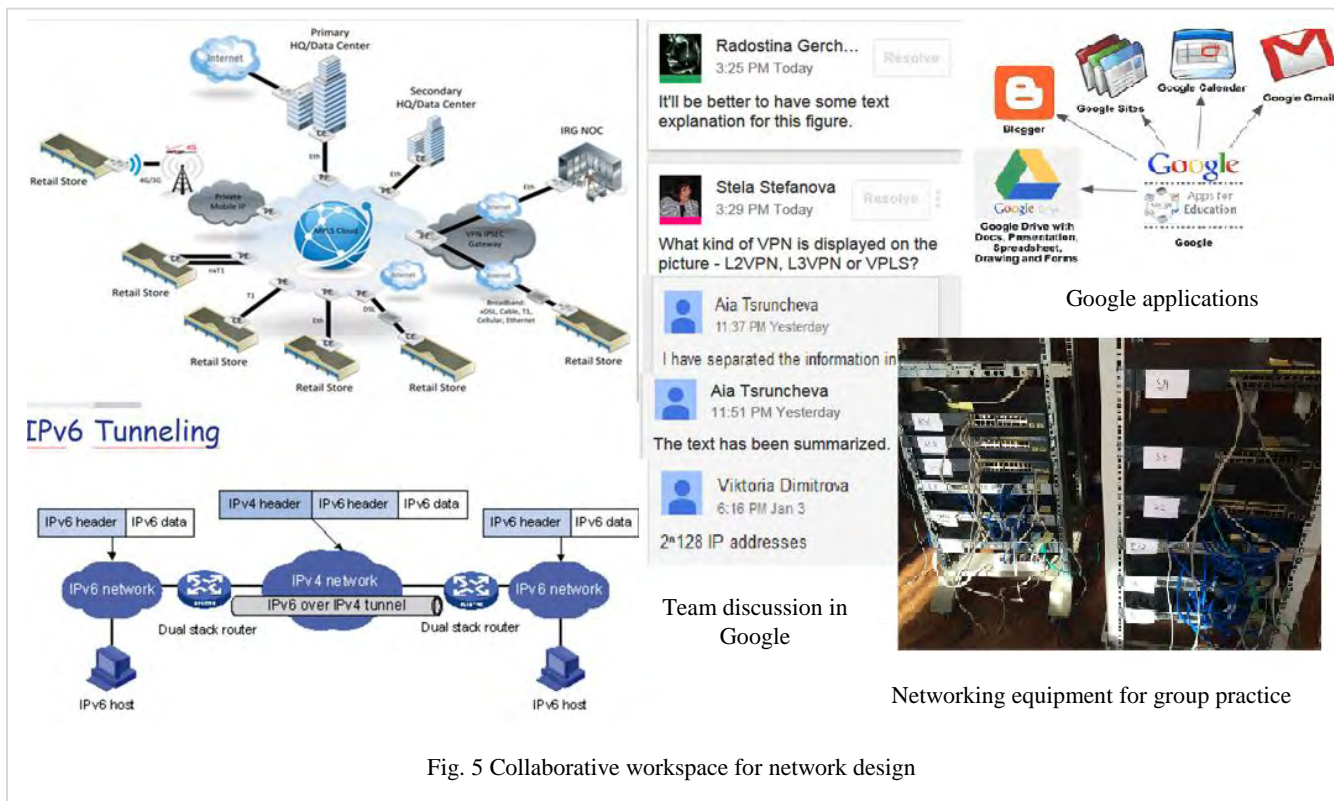


Fig. 5 Collaborative workspace for network design

refine the network parameters and characteristics, planning and writing the documentation, sharing the drafts, asking the teacher and other students for feedback, improving the project and project documentation, submitting respective report and presenting the obtained design and simulation results. All these require intensive use of modern communication tools - forums, blogs and social media for discussing problems. Students submit their questions as in-document comment, which are context related. Professor answers in the document. The discussion is part of the artefact (Fig.5).

Students are advised by specialists from the ICT industry. Industry professionals, teachers and students discuss and analyze collaborative experience. Already gained knowledge and skills are practiced in the laboratory to solve the tasks of the project by using dedicated networking equipment.

## V. CONCLUSION

The paper considers two case studies for introduction of collaborative learning approach at university and secondary school education to promote knowledge work competencies by implementing the trialogical design principle. The courses were re-designed from traditional face-to-face to collaborative project oriented adopting and applying modern online learning platforms, cloud collaboration tools, and social software.

Activities and environments used in the conducted pilots are discussed. In a whole, it has been a rewarding experience for both students and teachers. The trialogical approach was well accepted and considered as an appropriate path for transforming students' individual work into more collaborative activities. Students consider their experience to work in teams as very positive, challenging and useful to understand the benefits of collaborative working. Together

with learning subject matter they played (and learned) with new communication and collaborative technologies. The immediacy of the help provided via email and in-document, contextual comments, compared to the scheduled face to face meeting, was cited as a major plus in the post-courses surveys in both institutions – university and technology school.

## ACKNOWLEDGEMENT

The work reported in this paper is a part of the EU project “Promoting Knowledge Work Practices in Education – KNORK”, at the TU-Sofia and the Technology School “Electronic Systems”, and was supported by the Lifelong Learning Program of the European Commission.

## REFERENCES

- [1] H. Muukkonen, M. Lakkala, J. Kaistinen, G. Nyman, „Knowledge Creating Inquiry in a Distributed Project management course“, Research and Practice in Technology-Enhanced Learning, no. 5, pp. 73–96, 2010
- [2] Commission on Accountability in Higher Education, “Accountability for better results: A national imperative for higher education”, Boulder, CO: State Higher Education Executive Officers, 2005
- [3] S. Paavola, L. Lipponen, K. Hakkarainen, “Models of Innovative Knowledge Communities and Three Metaphors of Learning”, Review of Educational Research, vol. 74 no. 4, pp. 557-576, 2004
- [4] S. Paavola, M. Lakkala, H. Muukkonen, K. Kosonen, K. Karlgren, “The Roles and Uses of Design Principles for Developing the Trialogical Approach on Learning”, Research in Learning Technology, vol 19, no. 3, pp. 233-246, 2011
- [5] GitHub, <https://github.com/>

# Laboratory stand for experiments in real time with applications in training on control non-linear systems

Veselin Nachev<sup>1</sup>, Tanya Titova<sup>2</sup> and Plamen Nikovski<sup>3</sup>

**Abstract** – In the work are resumed theoretical and practical aspects in the study of oscillation in on nonlinear control systems. Addition to reviewing of analytical methods for determining the parameters of oscillation is described laboratory stand and methodology for carrying out the experimental research. The presented work in paper is applicable in the learning process of students in the course "Theory of automatic control", as is intended effect of stimulating the activity of students with the use of programming and physical model of object for control.

**Keywords** – nonlinear control, harmonic linearization, real time control, MATLAB instrumentation control

## I. INTRODUCTION

**Objectives of development:** Designing and realizing laboratory stand for demonstration and training in automation that integrates two basic methods for synthesis and analysis of control systems - simulation modeling and physical modeling.

**Application of laboratory stand:** Computer-based learning of basic principles of control, setting regulators, synthesis of control algorithms, studying effects such as non-linearity, saturation, "wind-up", noise, factors with adverse influence, etc., inherent to real objects for automation. Examines is the method of harmonic linearization (*describing function method*).

### Main advantages:

- Create a flexible environment, combining the advantages of simulation and physical modeling.
- Reducing the level of abstraction typical of simulation modeling.
- Opportunity for graphical visualization and data recording.
- Intensification of training related to the possibility of realization of a large number of experimental settings.

## II. MATERIALS AND METHODS

In the literature there are a significant number of

<sup>1</sup>Veselin Nachev, Department Automation and Control Systems University of Food Technologies, 26 Maritza Blvd, Plovdiv, 4002 Bulgaria, e-mail: v\_nachevbg@yahoo.com

<sup>2</sup>Tanya Titova, Department Automation and Control Systems University of Food Technologies, 26 Maritza Blvd, Plovdiv, 4002 Bulgaria, e-mail: t\_titova@abv.bg

<sup>3</sup>Plamen Nikovski, Department Electrical and Electronics University of Food Technologies, 26 Maritza Blvd, Plovdiv, 4002 Bulgaria, e-mail: plmnn@uft-plovdiv.bg

examples of the use of real-time control system in the automation training [1-3,6-8].

Oscillating behavior are typical of nonlinear systems. In these systems usually contain positional elements, such as controlling devices, secondary measuring means or actuators, which are characterized by simple realization, but reliability and low grade value. In most cases, non-linearity may be presented with straightforward types non-linearly characteristic - "dead zone", "saturation", positional elements and others. The auto-oscillations sufficiently objective is characterized by an amplitude ( $A$ ) and frequency of oscillations ( $\omega_a$ ). The value of  $A$  determines the accuracy of the regulation process and the frequency characterizes the intensity of the switching actuator. By considerations of reliability of the actuator in the synthesis of the system  $\omega_a$  is minimized. In many cases also, the parameters of oscillation allow to define the conditions for the stability of the system. The most common method for the study of oscillations of non-linear systems is the method of harmonic linearization. The essence of the method is presented in the following steps.

*1. Decoupled non-linear feedback system.* It is necessary the transformation of the structural scheme of the system, so that the one summarized non-linear element and the remainder a linear unit (Figure 1).

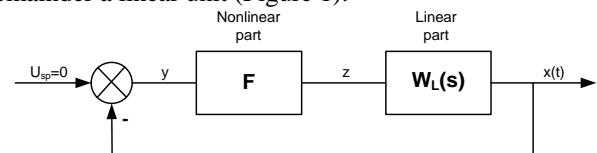


Fig.1. An equivalent circuit

In the transformation is not allowed:

- exchange of linear element with non-linear and conversely (without time delay). This is necessitated by the invalidity of the principle of commutative property.
- moving the non-linear unit through summing node. It is necessary failure to of the principle of superposition.

*2. Verification of the hypothesis of low-pass filter of the linear part.*

Perform a amplitude-frequency response of the linear part. Let there be the kind shown in Figure 2.



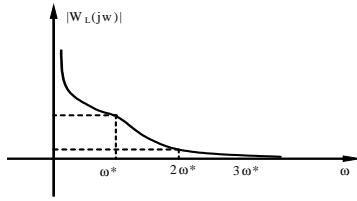


Fig.2. Magnitude of frequency response.

If  $\omega^*$  is the frequency of the first harmonic (fundamental frequency) of oscillation, higher harmonic multiples of it with  $2\omega^*$ ,  $3\omega^*$ , ... hypothesis for low-pass filter is confirmed in the case of satisfying the condition  $|W_L(j\omega^*)| \gg |W_L(jn\omega^*)|$ ,  $n = 2, 3, \dots$ .

### 3. Obtaining coefficient on harmonic linearization.

This is for structural scheme of the type at Fig.1

The linear part has the form (in autonomous system  $U_{sp} = 0$ ):

$$y = K_L(s)z,$$

When an oscillator process (if  $y$  is harmonic oscillation), passing through the non-linear element signal remains periodic not strictly harmonious, but with the same period and may be decomposed in order for Fourier harmonics with frequency components at integer multiples  $n\omega^*$ . Moreover, suppose that the linear dynamic system is a filter of low frequencies and missed only the first (hypothesis filter). If this assumption is not performing, errors of the method of harmonic linearization may be significant.

Let

$$z = F(y); \quad y = A \sin \omega t.$$

We present  $z$  in the order of Fourier:

$$z = C_0 + D_1 \sin \omega t + C_1 \cos \omega t + D_2 \sin 2\omega t + C_2 \cos 2\omega t + \dots$$

(1)

In symmetric non-linearity  $F(-y) = -F(y)$  is applied:

$$C_0 = \frac{1}{2\pi} \int_0^{2\pi} F(A \sin \omega t) d(\omega t) = 0.$$

When filtering harmonics of high order is necessary to determine only the coefficients  $D_1$  and  $C_1$  in equation (1).

After a substitution:

$$\begin{aligned} z &= F(y) = F(A \sin \omega t) = D_1 \sin \omega t + C_1 \cos \omega t = \\ &= D_1 \sin \omega t + C_1 \frac{d \sin \omega t}{dt} = \\ &= \left( D_1 \sin \omega t + \frac{C_1}{\omega} s \sin \omega t \right) \frac{A}{A} = \left( \frac{D_1}{A} + \frac{C_1}{A\omega} s \right) A \sin \omega t \end{aligned}$$

As a result, we obtain the equivalent transfer function of the non-linear element:

$$W_{NE} = \frac{F(A \sin \omega t)}{A \sin \omega t} = \frac{D_1}{A} + \frac{C_1}{A\omega} s = q_0 + q_1 \frac{s}{A\omega} \quad (2)$$

$$\text{or } (z = q_0 y + q_1 \frac{sy}{A\omega})$$

where:

$$q_0 = \frac{D_1}{A} = \frac{1}{\pi A} \int_0^{2\pi} F(A \sin \varphi) \sin \varphi \cdot d\varphi, \quad \varphi = \omega t; \quad (3)$$

$$q_1 = \frac{C_1}{A} = \frac{1}{\pi A} \int_0^{2\pi} F(A \sin \varphi) \cos \varphi \cdot d\varphi.$$

Thus the non-linear equation for  $z$  is replaced by the approximate equation (2) for the first harmonic. The coefficients  $q_0$  and  $q_1$  are called coefficient of harmonic linearization. They may depend on the amplitude and the frequency.

Displayed linearization coefficients for typical nonlinearities can be found in reference [4]. For the non-typical should be derived from (3).

### 4. Analysis.

The most commonly used methods for analyzing the method of harmonic linearization are [4]:

- Determining the parameters of oscillation using the criterion of Routh–Hurwitz.
- Determining the parameters of oscillation and stability using the criterion of Nyquist and derivatives - method of Goldfarb and Kochenburger.
- Determining the parameters of oscillation and stability using the criterion of Mihaylov.

## III. LABORATORY EXPERIMENTS

The principal scheme of laboratory stand which is realized is presented in Figure 3.

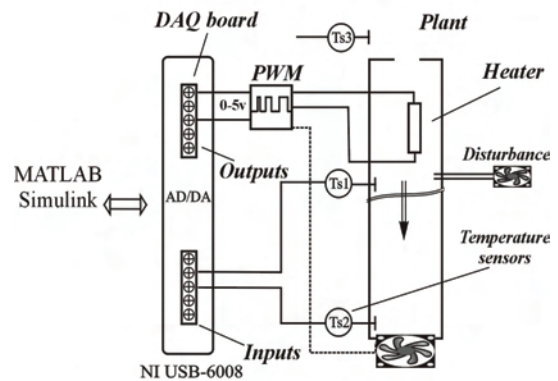


Fig.3



Figure 4 shows the existing model.



Fig. 4. Physical model and communication module.

The physical model communicates with the computer system and software, by means of a device for data collection (Data Acquisition, DAQ) - NI USB-6008 devices mainly include analog-to-digital (A/D) and digital-to-analog (D/A) converters, which are generally known as data converters.

Figure 5 shows scheme realizing the function of the a control device, in this case PID - regulator, which performed the necessary simulations and analyzes in the middle of MATLAB / SIMULINK. In particular - setting the parameters of the regulator.

Training example: a system of third order with two-position relay.

### 1. Object

$$G = \text{tf}(21.6, [187 \ 108 \ 147 \ 1], 'OutputDelay', 4)$$

Transfer function:

$$\frac{21.6 \exp(-4s)}{187s^3 + 108s^2 + 147s + 1}$$

### 2. Non-linear element - two-position polarized relay with hysteresis:

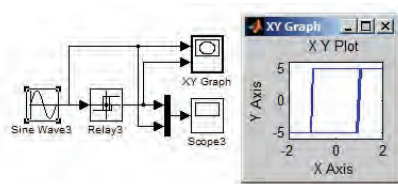


Fig.5

### 3. Receive coefficients of harmonic linearization.

$$A=1:0.0001:20;$$

$$W_{non} = ((4*5)/(pi.*A)).*sqrt(1-((1.^2)/(A.^2)))...$$

$$+j*(-4*5)/(pi.*(A.^2));$$

```
Z=-1./(Wnon);
Re=real(Z);
Im=imag(Z);
plot(Re,Im,'b*')
hold on
```

### 4. Plotting Nyquist on the linear part and Goldfarb $(-1/W_{NE}(A))$ .

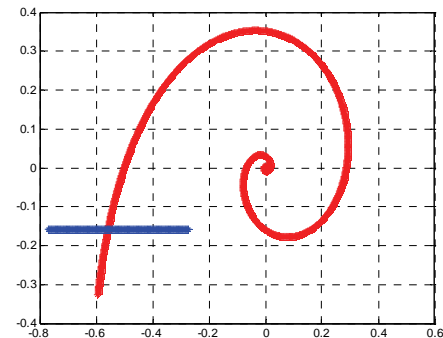


Fig.6

### 5. Defining the parameters of oscillation in the intersection point on hodographs.

### 6. Plotting model in Simulink, confirm the results of the presence of oscillator regime and evaluation of error for linearization.

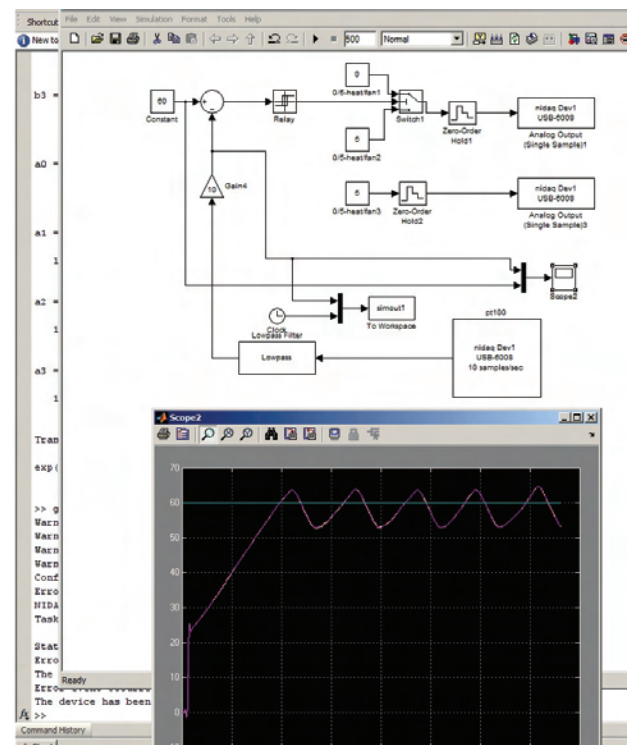


Fig.7 Real-time diagram with physical plant and SIMULINK controller.

Other schemes (tasks or "scenarios") that are made with laboratory stand are:

- Removing the transfer function on input and disturbance;
- Determination of the inertness on sensor;
- Definition of quality parameters of step response;
- Setting the controller on defined criteria of quality of step response;
- Control with two-position controller;
- Control with three-position controller;
- Synthesis of logical algorithms such as "if-then" eliminate the effects wind-up;
- Design of advanced control algorithms - neural, fuzzy, adaptive robust systems;
- Filtering, normalization and calibration signals from sensors, etc.

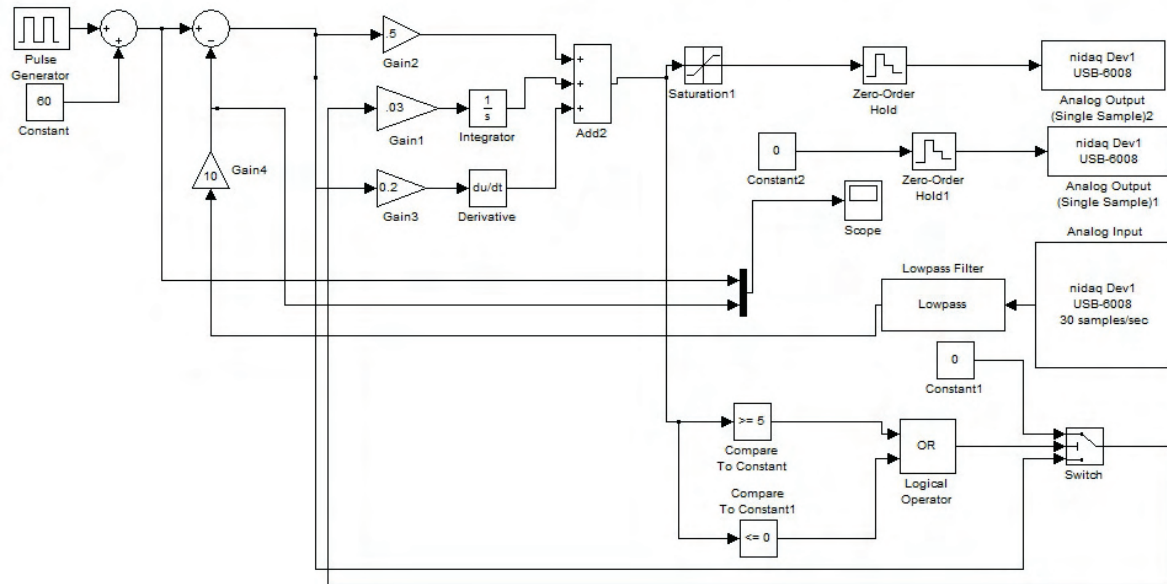


Fig. 10. Example simulation scheme for control of the object with PID and actuator saturation.

#### IV. CONCLUSION

The idea presented development is led from the need to implement innovative training methods - practical training in close to industrial environments, training base project development, problem-based learning and more. In this case this was achieved using a real physical object in combination with virtual tools for its control.

#### REFERENCES

- [1] F. Haugen, E. Fjeldalen, R. Dunia, T. Edgar, Demonstrating PID Control Principles using an Air Heater and LabVIEW, 2007
- [2] L. Smutny, R. Farana, The Consortial Approach to Advanced Control Laboratory Education, Proceedings of the 4th WSEAS/IASME International Conference on Engineering Education, Agios Nikolaos, Crete Island, Greece, July 24-26, 2007
- [3] K. Prekas, M. Rangoussi, S. Vassiliadis, G. Prekas, Performance of laboratory experiments over the Internet: towards an Intelligent Tutoring System on Automatic Control, Translations on Engineering, Computing and Technology, vol.2, 2004
- [4] C. Milosavljevic, Osnovi automatike: Nelinearni vremenski kontinualni sistemi automatskog upravljanja. Deo 2, Niš, Izdavač Elektronski fakultet, 2002
- [5] Z. Vukic and et. al., Nonlinear Control Systems, Marcel Dekker Inc., 2003
- [6] E. D. Bolat, Real Time Temperature Control of Oven Using Matlab-SIMULINK, Proceedings of the 11th WSEAS International Conference on SYSTEMS, Agios Nikolaos, Crete Island, Greece, July 23-25, 2007
- [7] <http://sine.ni.com/nips/cds/view/p/lang/en/nid/201986>
- [8] <http://homepages.lboro.ac.uk/~cgzkn/processcontrol/simulink4controlnewsinglepage.pdf>

# Using PLC for Control on Asynchronous Drives – Laboratory Simulator

Vasil Dimitrov<sup>1</sup> and Emiliya Dimitrova<sup>2</sup>

**Abstract** – In this paper some laboratory stands that offer opportunities for various asynchronous drives testing are described. A PLC has been involved into simulators for monitoring and control on the drive performance. A touch panel, connected to the PLC, gives a possibility of entering data of initialization and operation conditions. The simulators are useful in the students training and research.

**Keywords** – asynchronous drive, soft starter, programmable logic controller (PLC).

## I. INTRODUCTION

Programmable Logic Controllers (PLCs) are often used in the contemporary SCADA-systems and in many industrial sectors. PLCs are computer-based solid-state devices that control industrial equipment and processes. They have a user-programmable memory for storing instructions for the purpose of implementing specific functions such as I/O control, logic, timing, counting, three mode proportional-integral-derivative (PID) control, communication, arithmetic, and data and file processing. PLCs are used extensively in almost all industrial processes as the primary components in smaller control system configurations and provide operational control of discrete processes. PLCs are also used to accomplish the desired product and/or process tolerance around a specified set point. Then they can be employed in the control on localized processes and proportional, integral, and/or derivative settings on the PLC to be tuned to provide the desired tolerance as well as the rate of self-correction during process upsets (PID control). In SCADA systems, PLCs are often used at the field level as process controllers and RTUs because they are more economical, versatile, flexible, and configurable than special-purpose RTUs [1, 2, 6].

On the other hand, the fast industrial progress set up higher requirements of education quality. The training under the bachelor and master programmes in the Todor Kableshkov University of Transport prepares highly qualified experts in the fields of the Electrical Equipment of the transport, industry and power engineering. Training the skills for involving, programming, and setting-up the PLCs and human-machine interface is very important. Knowledge of the relevant software products is necessary. Evaluation of the PLC's possibilities for its using in different power drive systems for efficiency optimization is also very important.

<sup>1</sup> Assoc. Prof. Dr. Vasil Dimitrov is with the Faculty of Communications and Electrical Equipment in Transport at Todor Kableshkov University of Transport, E-mail: vdimitroff@abv.bg

<sup>2</sup> Dr. Emiliya Dimitrova is with the Faculty of Communications and Electrical Equipment in Transport at Todor Kableshkov University of Transport, 158 Geo Milev Str., Sofia, Bulgaria.

## II. LABORATORY SIMULATORS FOR ASYNCHRONOUS DRIVES EXAMINATIONS

The fast development of electrical drives and control systems has set up higher requirements to research and education quality. Several laboratory simulators have been designed and built on contemporary devices. They offer various opportunities to examine different asynchronous drives at inconstant load torque.

### A. Laboratory Simulators for testing of Frequency Controlled Asynchronous Drives

The main elements involved in the configuration of one of the simulators are developed by Siemens (Fig. 1): three phase energy-saving asynchronous motor with lightweight construction (AM – 2,2 kW, 1440 rpm); Sinamics G120 frequency converter; Simatic S7-200 Programmable Logic Controller (PLC); TP-177micro Touch Panel. A synchronous generator (SG – 30V, 60A) and variable resistor R simulate the load of the motor. Three sensors are mounted on the load system. They are connected to the PLC and form the system for assessing of the torque. The Differential Amplifier DA1 is used as voltage sensor. The current transducer HAS100S is a Hall Effect sensor measuring the generator current  $I_g$ . Its output voltage is amplified by DA2. A rotary encoder reports the generator speed  $n_g$  [2]. There is also a positioning system.

TP 177micro Touch panel connected to the PLC gives the possibility of monitoring and control on the drive system as well as of entering the initialization and parameterization data and work conditions.

A Windows-based computer system is used for configuring and data storage. The necessary commissioning software Starter is installed on this computer and is used for setting-up the Sinamics G120. A subprogram Trace is used for real-time tracing the inverter parameters. This is a great opportunity of an experimental research of transient modes of the frequency controlled asynchronous drives [3].

The simulator offers various possibilities of examination on a frequency controlled asynchronous drive at inconstant load torque. The influence of parameters and operating modes of the inverter on the energetic drive performance could be examined [2,3]. Using the subprogram Trace, the dynamic characteristics with starting, stopping and changing the load or the speed set point could be captured. All data are exported to Excel tables and graphs can be automatically plotted. Transient modes have a big practical importance, since the correct choice of motor power, the proper selection and sizing of the apparatus of control system, the reduction of energy consumption during starting and stopping are based on them.

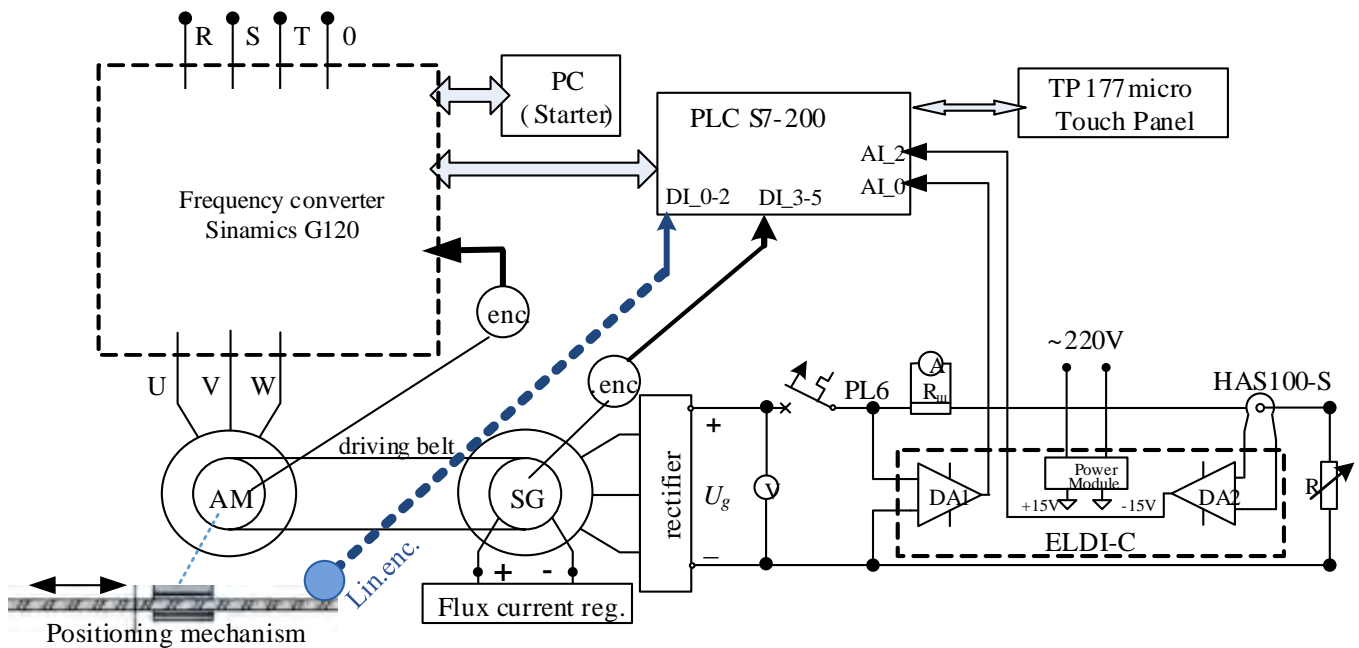


Fig. 1. Simulator for testing of an asynchronous drive with frequency converter Sinamics G120 (Siemens)

The existence of a PLC and touch panel TP177micro allows for simultaneous drive control and parameters tracing and capturing. Simatic S7-200 can be used for control and optimization tasks both for the inverter drive and the positioning system. A linear encoder reports the travelled distance of the positioning axis to the S7-200. A rotary encoder is used for closed loop speed control on the drive. A possibility of torque evaluation is created and optimal drive behaviour could be achieved. Adaptive control can be implemented. It is used when the automatic control drive system has to modify its own characteristics, depending on the operating conditions, i.e. the automatic control system adjusts (adapts) itself to variations in environment. Such control is embedded in many contemporary positioning systems. The requirements for operation with high precision, ensured and sustainable in all ranges of operating speeds, are achieved through control on the positioning, speed and torque of the motors. Adaptive control on a traction asynchronous drive could be also simulated – indirect identification of disturbances could be realized. It takes into consideration the change of controlled parameters as a result of a disturbance.

Another simulator is based on a single phase frequency converter ELDI developed by the Bulgarian company Electroinvent Ltd (Fig. 2). The load torque can be changed by

regulation of the flux of a synchronous generator. The influence of parameters and operating modes of the inverter on the energetic drive performance can be examined using the network analyzer Net Vision 2010. The true RMS values of the voltage and current are measured simultaneously. Active power, reactive power, power factor, sequence of rotation, harmonic analysis are computed and indicated. The vector diagram and the waveforms of the input signals can also be drawn on the graphic LCD display [4].

### B. Laboratory Simulators for testing of a directly plugged in the grid Asynchronous Drive at inconstant load torque

The laboratory stand for asynchronous drive examinations is shown in Fig. 3. This motor is made in Bulgaria; its power is 3 kW. A load of the motor can be changed by an electromagnetic brake (EMC). Possibility of reverse is provided. The necessary locks and protection and control equipment are also assured. Methods for laboratory examinations and tests by instrumentation with high accuracy are developed: determination of the static mechanical and electromechanical characteristics and characteristics that show the change of energetic parameters (power factor, efficiency, THD etc.) of the drive at the load change [5].

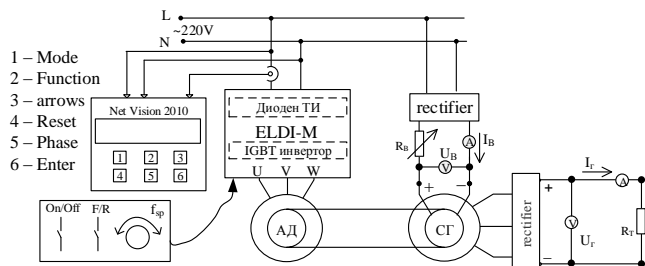


Fig. 2. Simulator for testing of an asynchronous drive with frequency converter ELDI (Electroinvent)

## III. UPGRADING THE LABORATORY SIMULATOR

The speed regulation of the asynchronous motors in many fields of industry is not necessary. Many drives need to be started and stopped smoothly but its speed has to be constant through the operating modes. Intelligent devices called “Soft starters” can be used in those cases.

A motor soft starter is a device used with AC induction motors to temporarily reduce the torque and current surge during start-up. This reduces the mechanical stress on the motor and shaft, as well as the electrodynamic stresses on the



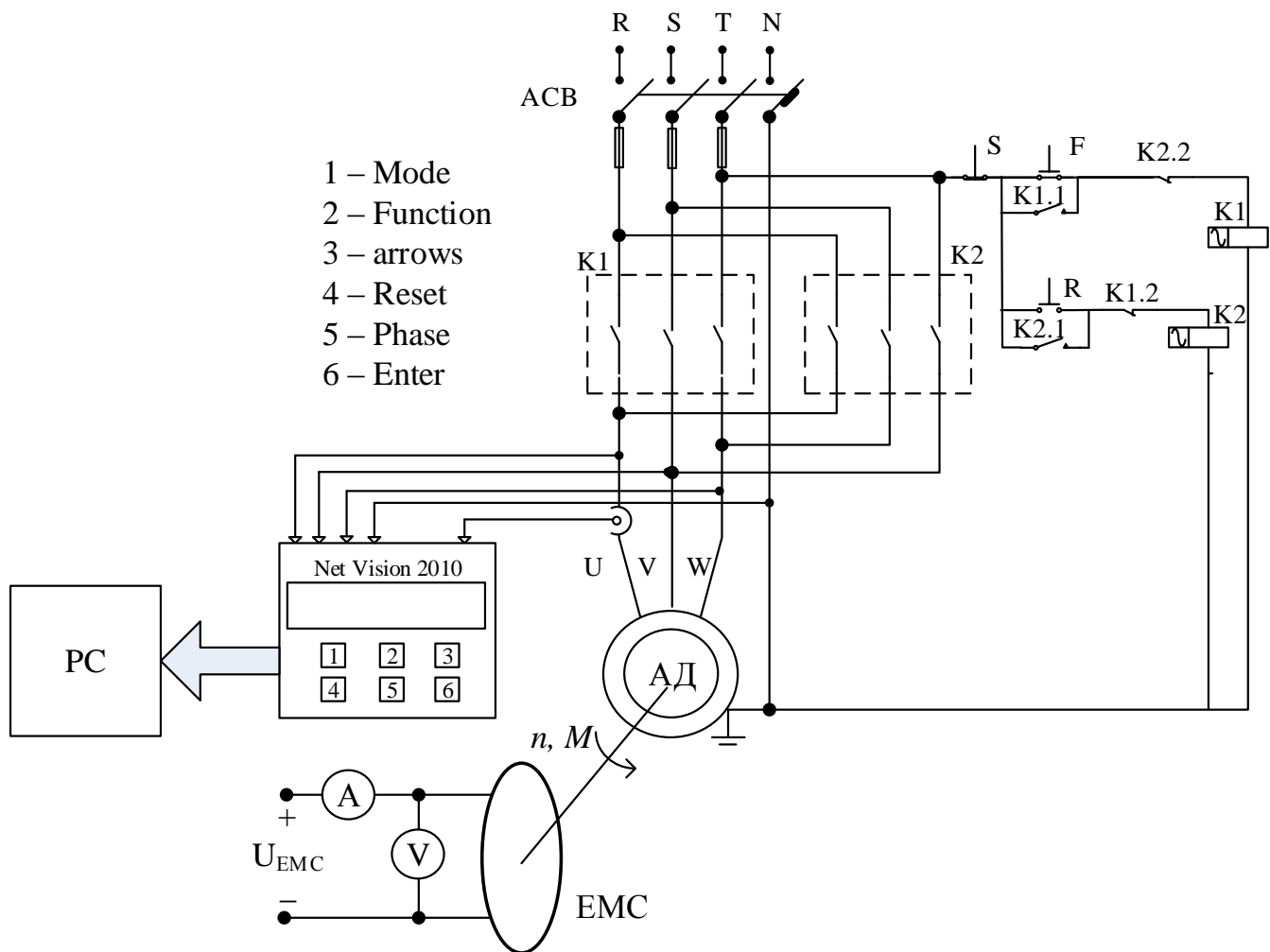


Fig. 3. Simulator for testing of a directly plugged in the grid asynchronous drive

electrical distribution network, extending the lifespan of the system [7]. Electrical soft starters reduce the torque by temporal reduction of the voltage or current input. Typically, the voltage is controlled by reverse-parallel-connected thyristors. The soft starter continuously controls the three-phase motor's voltage supply during the start-up phase. This way, the motor is adjusted to the machine's load behavior. Mechanical operating equipment is accelerated smoothly. Service life, operating behavior and work flows are positively influenced. The thyristors can be bypassed by a relay or a contactor in steady-state operation. Electrical soft starters can use solid state devices to control the current flow and therefore the voltage applied to the motor. They can be connected in series with the line voltage applied to the motor, or can be connected inside the delta ( $\Delta$ ) loop of a delta-connected motor, controlling the voltage applied to each winding. Solid state soft starters can control one or more phases of the voltage applied to the induction motor with the best results achieved by three-phase control.

Soft starters can be set up to the requirements of the individual application. In pump applications, a soft start can avoid pressure surges. Conveyor belt systems can be smoothly started, avoiding jerk and stress on drive components. Fans or other systems with belt drives can be started slowly to avoid

belt slipping. Soft starts are seen in electrical R/C helicopters, and allow the rotor blades to spool-up in a smooth, controlled manner rather than a sudden surge. In all systems, a soft start limits the inrush current and so improves stability of the power supply and reduces transient voltage drops that may affect other loads.

Therefore, the study and examination of the soft starters is very important. The laboratory stand shown in Fig. 3 can be upgraded with an appropriate soft starter. Altistart 01 soft start/soft stop units ATS01N2 developed by Schneider Electric Company could be used [7].

The Altistart 01 starters enhance the starting performance of asynchronous motors by allowing them to start gradually, smoothly, and in a controlled manner. It helps to prevent mechanical shocks, which cause wear and tear, and subsequent maintenance work and production downtime. The series includes several devices offering various features. Some of them operate only as a torque limiter on starting, but others can be used as a soft start/soft stop unit for asynchronous motors.

The Altistart U01 limits the starting torque and current peaks on starting on machines that do not require a high starting torque. It is designed for the following simple applications: conveyors, compressors, conveyor belts, pumps,

fans, automatic doors and gates, small cranes, belt-driven machinery, etc.

The Altistart 01 ATS01N1 soft starters control one phase of the motor power supply (single-phase or three-phase) to limit the starting torque. They feature an internal bypass relay. An external power supply is required for controlling the starter. A contactor is always required to shut off power to the motor.

The Altistart 01 ATS01N2 soft start/soft stop units control two phases of the motor power supply to limit the starting current and for deceleration. They also feature an internal bypass relay. The use of a line contactor is not necessary on machines where electrical isolation is not required. These devices are equipped with 3 potentiometers for setting the starting and deceleration time and for adjusting the starting voltage threshold according to the motor load. 2 logic inputs provide Run/Stop commands and 1 logic input can be used for the Boost function. 1 logic output indicates the end of starting and 1 relay output indicates that the motor has reached a standstill at the end of the deceleration stage. These inputs and outputs can be connected to the PLC Simatic S7-200. The scheme of the upgraded simulator is shown in Fig. 4. The line circuit breaker is mounted between the grid and soft starter, but is not shown in scheme. TP177micro Touch Panel can be configured for Start / Stop and Boost commands, which can be transmitted through the PLC to the soft starter. Digital inputs of the PLC can be used for receiving the information about drive condition.

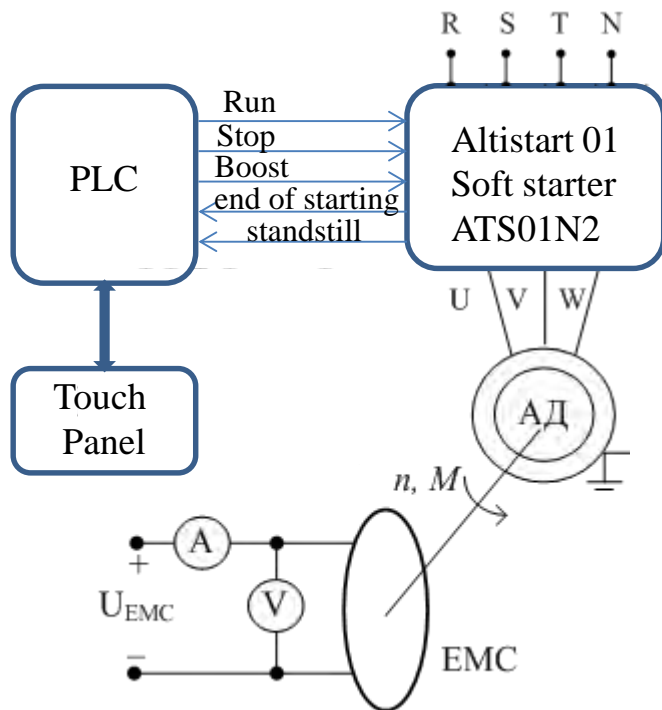


Fig. 4. Simulator for testing of an asynchronous drive with soft starter

## IV. CONCLUSION

Soft starters are widely used in industry for control on asynchronous drives when the speed regulation is not a mandatory requirement. They are cheaper than inverters and frequency converters and allow temporal reduction of the torque and current surge during start-up, as well as smooth deceleration.

PLCs are often implemented as primary components in small and medium control system configurations. The PLC Simatic S7-200 offers maximum automation at minimum cost. It is compact, fast and highly powerful, feature great communications capabilities and it is based on very user-friendly software and hardware. It can be used for complex automation tasks and finds application in many branches of industry, power engineering and transport.

These advantages cause to teaching the students how to use and program soft starters and PLCs. The upgraded simulator offers various possibilities of implementation into practice many laboratory exercises, for example:

- Examination of an asynchronous motor at a constant speed and variable load;
- Study of the influence of the start-up and ramp-down time on the drive performance.

The laboratory simulator gives possibilities of student's practical training in many terms of reference, for example:

- Training the programming skills for setting-up the soft starters, PLCs and touch screens – a knowledge of the relevant software products is necessary;
- Synthesis of algorithms for optimal control on the drive systems;
- Evaluation of the PLC's possibilities of its using in the drive systems for the efficiency optimization.

## REFERENCES

- [1] Dimitrov V., "Indirect Identification of the Disturbances by PLC Simatic S7-200", XLVI Int. Sc. Conf. "ICEST 2011", Serbia, Niš, June 29 - July 1, 2011, Proceedings of Papers, Vol. 3, ISBN: 978-86-6125-033-0, pp. 1018 – 1021.
- [2] Dimitrov V., "Adaptive control on traction asynchronous drives-laboratory simulator", Proceedings of Technical University of Sofia V<sup>th</sup> Conf. of Faculty of Electrical Engineering "EF 2013", ISSN 1311-0829, Vol. 63, Issue 5, 2013, September 2 - 5, 2013, Sozopol, Bulgaria, pp. 155-162
- [3] Dimitrov V., E. Dimitrova, "Research of PWM pulse frequency influence on the asynchronous drives characteristics – laboratory simulator", Acad. journal „E+E“, Vol. 48, № 9-10/ 2013, ISSN 0861-4717, pp. 37 – 44
- [4] Pavlov G., V. Dimitrov, "Frequency Converter ELDI-M Examinations", Sc. Seminar „KEIT 2014“, Acad. journal „Mechanics, Transport, Communications“, ISSN 1312-3823, Vol. 12, issue 3/2, 2014, pp. X-7 – X-12, Art.ID 1004
- [5] Dimitrov V., "Development of Laboratory System for Electrical Drives Examinations", Int. Sc. Conf. "Transport - 2015", Acad. journal „Mechanics, Transport, Communications“, ISSN 1312-3823, Vol. 13, issue 3/3, 2015, pp. X-24 – X-29, Art.ID 1246
- [6] *Simatic S7-200 – Programmable Controller*, System Manual, Siemens, 2008
- [7] *Soft starters Altistart 01*, Catalog, Schneider Electric, 2014

# Importance of Learning to Code, Coding to Learn in Higher Education

Anton J. Knierzinger<sup>1</sup> and Boyka Gradinarova<sup>2</sup>

**Abstract** – Computer and internet give essential features to our life. Should all students learn programming? The message of this paper is a plea to improve the quality of highest education via bringing forward the importance of algorithms by stating their role in different sectors of our society as well as the learning and teaching process. Algorithmic thinking should be part of all subjects at all levels of education, because it improves the chances of our students in many directions and reflects the responsibility of education for our society. But coding in that sense is different from traditional programming. Teaching coding has three aspects:

- a technical approach - what are we talking about
- a social approach - why are we talking about
- a didactic approach - how are we talking about.

**Keywords** – algorithm, digital society, coding, programming, education.

## I. INTRODUCTION

Let us spring from data. We commonly understand data as information, values or findings, which has been discovered by measurement or observation. Data are object of political discussion in form of data protection or data security. We affiliate their possession with economic power

A definition of an algorithms by Wikipedia is „a self-contained step-by-step set of operations to be performed” either in real life or virtually on a computer. „Algorithms exist that perform calculation, data processing, and automated reasoning”

Anyway algorithms are proposed actions. They occur everywhere in our life, cooking recipes, instructions for technical installations, turn-by-turn directions for a route, there are many other examples. So we all have internalized the concept of an algorithms. By defining them we structure a problem into pieces and doing so we prepare it to be operated in a machine.

Teaching algorithmic thinking conveys two important competences:

- problem analysis and
- problem solving.

Their importance for our life, for science, society and education is commonly underestimated. Algorithms allocate power and distribute chances. Especially because of that we propose that the preoccupation with this topic should be part of education at all age level

<sup>1</sup>Anton J. Knierzinger University of Education Linz, Austria  
anton.knierzinger@ph-linz.a.

<sup>2</sup>Boyka Gradinarova Technical University of Varna, Bulgaria  
BGradinarova@hotmail.com

Where in society do we meet algorithms. We identified four major areas, all of them with great influence on our life.

1. Science and technology They have a great relevance in IT studies. But there are numerous examples in many other disciplines where algorithms play an important role.

2. Society and economy Algorithms and their ownership provide clout and power in our society, but they also offer creativity and fun, they facilitate advancement and value.

3. Pedagogy and human development In general (full development of the human being via education) and in vocational training (the imparting and acquisition of specific skills) teaching algorithmic thinking and the knowledge how to deal with coding improves the chances of our students in different ways.

4. Didactics and education Using coding as a principle of teaching improves education by offering new ways of creativity, connection to real life and allow to have fun via learning. It is our experience since the introduction of IT into education that this can intensively change and improve the process of teaching and learning.

## II. SCIENCE AND TECHNOLOGY

Paradigmatically we want to have a look on different areas where coding is playing a major role in science and technology

### A. Automatisaton

Industry 4.0 is defined as the „informatization” of production technology and logistics by machine-to machine communication. It is also named as smart or networked production. Within industry 4.0 robots can take a lot of hard work from the shoulders of their human companions. They work on the basis of cyber-physical systems and what is called the internet of things.

But automatization will reach nearly all areas of our life. That leads to the question: Are we in danger to loose our jobs to a robot? BBC has published a web-site „Will a robot take your job? based on a study done by Oxford University and Deloitte (<http://www.bbc.com/news/technology-34066941>). Oxford University academics Michael Osborne and Carl Frey calculated how susceptible to automation each job is based on nine key skills required to perform it; social perceptiveness, negotiation, persuasion, assisting and caring for others, originality, fine arts, finger, manual dexterity and the need to work in a cramped work space. Students can investigate the future chances of their profession and find amazing facts.

### B. Robotics

Undoubtedly their are numerous fields for the use of robots. It can be inspiring and informative to present them in lesson.

Robots are fundamental of modern military strategies. We find them as drones, secret agents, fighters and helpers for soldiers. Associated with the technology we have to rise questions like:

- Is the inhibition threshold dropping?
- Who then decides on life or death?
- Will less or more people killed in an electronic war?

But robots will play other different roles in the future. They will act as servants, tour guides, play mates for kids, aid worker, lawnmowers and toys. At it we should discuss the following questions with our students:

- Under what conditions should robots replace humans?
- Can robots have human qualities like creativity, real emotion and spontaneity?

We have to discuss what Joseph Weizenbaum [3], famous for ELIZA, the first algorithm with human behavior, turned during the last years into an „old any man” asking: Are we in danger that when robots simulate man, humans are more an more simulating robots? Robots touch very much our self-conception

### *C. Autonomous Transportation Systems*

Autonomous logistics describes systems that offer unmanned, autonomous transfer of equipment, baggage, people, information or resources from one point to another without human intervention. This is a new area for research and development and therefor of great interest for higher education.

Visualization and coding of processes plays an ever-growing role in our society and therefor represent an emerging market. Examples are image editing, the revolution of music industry caused by the invention of the MP3 algorithm or the development of the digital presentation technique. An example of future trends in visualization can be seen in an advertisement for the upcoming video product HoloLense ([www.microsoft.com/microsoft-hololens/en-us?video-url=vdeHHP](http://www.microsoft.com/microsoft-hololens/en-us?video-url=vdeHHP)). Microsoft HoloLens is the first fully untethered, holographic computer, enabling high-definition holograms to integrate with your world. It offers the possibility to combine real and virtual objects to one perceived reality. A sentence quoted from Microsoft describes the impact of visualization: „If you can change the way you see the world, you can change the world you see!”. According to Marshall McLuhan a change in media technology implicates a didactic commission for higher education

### *D. Meteorology*

Meteorology is one of only two more examples of natural sciences that we show, where algorithms are involved in the process during the last decades. The success in the prediction of weather is mainly depending on the speed of the computer systems and the quality of the algorithms used. The improvement of both factors made it possible to increase the complexity of the models for computation.

IT is also used for the fast processing of the big amount of data produced by weather satellites, the prediction of natural disasters and the online visualization of weather forecasts.

### *E. Biology*

The importance of algorithms in biology is usually underestimated. Without adequate algorithms the decoding of the human genome would not have been possible in a short time. This has big influence in medicine, pharmacology, agriculture and forensic examination. But also IT is learning from biology. In artificial intelligence, an evolutionary algorithm (EA) is defined as a generic population-based optimization algorithm. It uses mechanisms inspired by biological evolution, such as reproduction and selection. Bionics is the use of methods taken from biological systems to the design of modern technology. In robotics the study of bionics has produced efficient designs for robots, artificial neurons, artificial neural networks and swarm intelligence. Algorithms are used for the simulation of biochemical processes and in the workflow management of clinical research. Therefore a lot of European universities offer special studies in bionics. They can be seen as an example for the improvement of higher education by an interdisciplinary approach.

## **III. SOCIETY**

There are many samples how algorithms influence our society and sequent our life. For illustration it is enough to demonstrate three of them.

### *A. Politics*

The participation of citizens in political decisions is based on the following mechanisms in the political process:

- Elections

It has been a long debate in Europe to what extent voting can and should be maintained by IT. But the fundament of free elections is the trust of the participating citizens in the system used to collect the votes. As people are always confronted with the misuse of their data in Austria there is no electronic voting at the moment. In Bulgaria there is a living public discussion on electronic voting.

#### Grassroots democracy

At the end IT supported ways of collecting the public opinion could lead to a stronger civic involvement. But the attempts of the Pirates Party to foster civic participation in Germany using a software called „Liquid Democracy” shows that there are limits and their violation obstruct the political process and the contribution of a political movement, because sometimes it takes too much time to come to an end.

- Open data and open codes

We need ways how responsible citizens come to the information which they need for participation. In the EU we see an ongoing debate on how the access to governmental data should be organized. The majority of the algorithms we are subject to is not in possession of governments but in the hand of a few global players in IT industry. We propose a discussion of their role and how the rights of the people on data and the algorithms used for their processing can be saved.

- Communication

For example the political processes in northern Africa in the last years have shown the importance of open communication for a progress in society



### B. Economy

Among the various possibilities of the use of algorithms in economy we want to present only two but very characteristic examples:

- User profiling

User profiling is used in different ways. One example is individually designed advertising. To do this you have to collect data about your clients, but the main factor for success is an algorithm for a suitable analysis. The most extreme use of client profiling we found in Switzerland. COOP tried individual pricing. The price of a specific good was not only determined by time and the location of the shop but also by the analysis of the buying behavior of the client detected by clandestine algorithms. This project has been terminated because it caused a storm of protest.

- Share market

A considerable part of the trading of shares is currently done by algorithms. This led to a competition concerning the quality of the algorithms used between the major players in stock market. The quality of the codes used has a big influence on the performance of trading companies. The CEO of an Austrian bank told us that he prefers mathematicians instead of economists in his treasury department.

### C. Arts

The influence of algorithms on arts is widely unknown, but there are interesting and numerous examples for the use of algorithms to generate artistic expression. Artists use algorithms since ancient times. For example you will find them by Leonardo da Vinci, Bach, Mozart, and Escher. Beside visual design algorithms are also used for computer generated music and poetry today.

Linz in Austria has a special and worldwide known center for electronic art, the AEC - Ars Electronica Center. It's FutureLab is a place for researching and trying out new cyberarts technologies. Once a year, Ars Electronica invites artists, scientists and researchers from all over the world to a conclave in Linz to confront a specific, interdisciplinary theme in the context of speeches, workshops, exhibitions and symposia. A yearly competition called u19 is dealing with questions like: Who decides, how the next generation will be in 20 years? Which technologies will be used then? How is our society developing? Which artistic ideas can help us?

## IV. PEDAGOGY AND HUMAN DEVELOPMENT

Beside all the societal, economic, creative or somehow other good reasons to talk about algorithms in university courses, there is one, and we think to be probably the most important for teachers, „Dealing with algorithms in education increases the chances in life of our students.“. And we can see this in various directions.

This affects at least four areas of competences: According to our experiences during the last decades and proven by

many studies, teaching coding and algorithmic thinking leads to an enhancement of the competences of our students.

This affects at least four areas of competences:

Professional Competences:

- cognitive competences, because they learn how a decision as a structure of choice has an influence on an algorithm

- analytic competences, because for the solution of a problem they have to isolate, analyze and structure the necessary information

- logical reasoning and faculty of abstraction

- problem solving • selection and application of suitable tools

Methodological competences

Students gain abilities in the use of IT-methods, the usage of digital media, acquisition of information and their evaluation and visualization.

Occupational outlook

The preoccupation with coding, the work on computers and the general competences in problem solving lead to proficiencies far wider than traditional IT-training. The excogitated dealing with algorithms in university courses lead to

- a widening of the spectrum of professions seen as possible and / or worthwhile

- improves the chances on the labor market. Students gain abilities in the use of IT-methods, the usage of digital media, acquisition of information and their evaluation and visualization.

Personal Competences

Beyond doubt we could observe an alteration and enhancement of the personal competences of our students and the kids participating in the computer camps in Bulgaria touches especially their self-confidence and the self-assessment of their own prospects [3].

The better and proper handling of digital media facilitates the most important competences in job and life in general. This can be demonstrated by the findings in the evaluation study of the „Power Girls Project 2013“ in Upper Austria done by Alfred Weinberger from University of Education Linz. Power Girl is an initiative of Education Group, a governmental agency aiming the integration of media and technology in schools. The results of the yearly evaluations since 2008 show clearly the positive effects of the engagement in technology of girls on occupational and personal competences.

Girls who participate in the „Power Girls Project“ boast

- a more emancipated and stronger non-traditional role model,

- value their own technical competences higher,

- are more interested in technical-oriented jobs and qualifications and

- choose more often higher technical education (schools, apprenticeships) than girls not taking part in this project.

## V. DIDACTICS

When we started to integrate IT into education at the beginning of the eighties of the last century, we could feel that we had received a medium for revolutionizing didactics at

school and university level. But during the years we unfortunately lost our feeling for this mission and a concept of IT as a medium for better employment gained ground. There are several reasons for that and it is worth to discuss them in order to redesign the whole process

We see that

- IT lost its special position
- persistence of the education system • a wrong selection of content

There are two main reasons why Informatics lost a large portion of her attractiveness and the number of beginners in IT studies in Germany and Austria is declining in the last years.

1. The abnegation of core topics of IT, like programming, and the concentration on application software might accommodate the interests of economists. But this doesn't fit to the expectations of our students. They want to work creatively and choose their tools by themselves.
2. We do not pay enough respect to McLuhan's subtext „The medium is the message“. The message of coding is that it needs its own didactics. When observed this can lead to fantastic results.

To support this view we want to show a model for creative learning presented by Mitchel Resnick. He was working with Seymour Papert and is the developer of the educational programming language LOGO. Resnick proposed a model with four columns all starting with the letter P, the 4'P Model for creative learning:

- Projects,
- Peers,
- Passion and
- Play.

Resnick developed a new version of LOGO called SCRATCH designed only for creative learning by offering a very different approach to coding. Coding is not any more a set of technical skills but a new way of thinking and personal expression.

Projects

„Scratch was designed with projects in mind.“ [5] Students learn better through active working on topics meaningful for them. Learning is more efficient via projects and when using proper tools. This diversity of projects is a reflection of the diversity of interests of young people. Therefore instructions should be:

- cross-disciplinary • project-oriented • targeted • have a practical attitude
- and first and foremost be creative.

Peers

„Encourage collaboration, sharing and help children to learn to build on the work of others. Coding shouldn't be a solitary activity“. [7]

Learning thrives as a social activity, when people exchange their ideas, work on common targets and share their results. We believe that interaction with peers should be a central element in the learning process. So much learning can be exciting when it is team-oriented and appropriate to age and background. When teaching coding we must also focus our attention to choose the right level of abstraction.

Passion

Teaching should raise enthusiasm, reason the doing and engender creativity. When working on projects which are meaningful, challenging and kidding students work harder, longer and more efficient. That leads to success and sustainability of the learning process.

Play

Playing is the way of teaching that will support the creativity of our students. Learning to code should support experiments, lead to personal limits, urge the students to rise and always allow to try new ways..

## VI. CONCLUSION

At the end we want to return to the starting point of the history of IT didactics and finally propose our conclusion and the idea which underlying this paper.

BASIC (an acronym for Beginner's All-purpose Symbolic Instruction Code) is a general-purpose programming languages whose design philosophy emphasizes ease of use. In 1964, Kemeny and Kurtz created the original BASIC language at Dartmouth College. They wanted to enable students in fields other than science and mathematics to use computers and to mediate them a feeling for algorithmic thinking, problem solving and the impact of algorithms on their professional and personal life. (see Wikipedia)

We see two reasons why we propose that we should again put a spotlight an coding and algorithmic thinking in tertiary education today:

- Algorithms have an underestimated power in many areas of science and society.
  - Coding offers new ways of creative learning
- Sometimes a step back can be a step forward.

## REFERENCES

- [1] Burton, A. (2013). Dolphins, dogs, and robot seals for the treatment of neurological disease., *Lancet Neurol.* 2013; 12(9):851-2 (ISSN: 1474-4465)
- [2] Knierzinger, A. (2015). learning to code,coding to learn - The World of Algorithms in School, E-Learning Fachtagung, presentation, Linz
- [3] Knierzinger, A. Gradinarova, B. Learning to code, coding to learn the world of algorithms in higher education
- [4] Marschmann, H. (2008). Joseph Weizenbaum: Der zornige alte Mann der Informatik. Artikel in „Der Spiegel
- [5] Maloney, J., Resnick, M., Rusk, N., Silverman, B., & Eastmond, E. (2010). The Scratch Programming Language and Environment. *ACM Transactions on Computing Education (TOCE)*, vol. 10, no. 4 (November 2010)
- [6] Repnik, R., Robic,D. Physics Learning in Primary and Secondary Schools with Computer Games- An Example-Sngry Birds, E-learning Instructional Design, Organizational Strategy and Management, InTech,2015.
- [7] Resnick, M., & Siegel, D. (2015). A Different Approach to Coding. Bright/Medium.

# Infrared investigation on the thermal field of the human face during the EEG session

Kalin Dimitrov<sup>1</sup>, Stanio Kolev<sup>2</sup>,  
Iglika Andonova<sup>3</sup>, Tsvetan Mitsev<sup>4</sup>

**Abstract** – This paper presents a study on human skin surface warming during EEG. The case is with EEG with Emotiv® brain computer interface. For the experiment was used the connection between the optical radiation of the objects themselves and their temperature. The study was conducted under laboratory conditions during a session with a wireless 14 channel EEG system. The measurements were done simultaneously at many points by using an infrared camera. The conditions in the rooms and the condition of the patients were taken into account. The emissivity coefficients of the objects studied were also considered. In this work is presented only the thermography analysis.

**Keywords** – Infrared thermography, medical applications.

## I. INTRODUCTION

Technology evolves every day and we face new perspectives in the field of health - diagnostic and therapeutic. What is observed in the last few years, is the significant application of infrared imaging, usually in the field of diagnostics. However, the potential of this technology in contemporary medicine is still to be explored in further detail. The current paper is an attempt to broaden the view and deepen the understanding of some of the existing diagnoses through the analysis of data provided by a thermal camera [1-13]. The primary analysis (EEG) is parallel with the infrared measuring [14-16].

## II. CURRENT STATE OF THE PROBLEM

Much research that is done now focuses on the measurement of one standard diagnostic procedure at a time. The study of the correlations between the fundamentally different physical nature of the random processes is relatively rare [17-20]. The study of EEG together with optical methods mainly is done in the near infrared spectrum called Functional Near-Infrared Spectroscopy (700-900nm) [19,20]. The study in the far infrared range (8-13um), based on the objects' own thermal radiation, is usually performed alone [2-4,13]. This is the reason to conduct this study.

<sup>1</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](mailto:kld@tu-sofia.bg).

<sup>2</sup>Stanio Kolev is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: [skolev@tu-sofia.bg](mailto:skolev@tu-sofia.bg).

<sup>3</sup>Iglika Andonova is with the Faculty of Automatics at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia, 1000, Bulgaria, E-mail: [iglikaa@tu-sofia.bg](mailto:iglikaa@tu-sofia.bg)

<sup>4</sup>Tsvetan Mitsev is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: [mitsev@tu-sofia.bg](mailto:mitsev@tu-sofia.bg)

## III. EXPERIMENTS

The infrared pictures do not influence the main diagnostic procedure, so there are no special requirements on the part of doctors apart from the general requirements for hospital hygiene and rules. As the procedures that we use are non-invasive, they can be conducted outside the hospital [21,22].

The thermal camera used for data collection was FLIR E40, with thermal sensitivity of  $< 0.07^{\circ}\text{C}$  and temperature range of  $(-20^{\circ}\text{C}$  to  $650^{\circ}\text{C})$  [23]. All examinations were performed in a sitting position in a quiet room at a constant room temperature of  $20 \pm 0.5^{\circ}\text{C}$  following an acclimatization period of 20 min keeping the hands free of any contact to the rest of the body or other objects [22]. For maximum accuracy, the camera was fixed on a stand and movement of the object was avoided. For the final analysis temperature values were determined and given in degrees Celsius ( $^{\circ}\text{C}$ ). Equally, the relative humidity showed stable values over time. All images were corrected using an emissivity factor of 0.98 for the human skin.

In this study, a series of images was made between certain intervals of time. The infrared pictures were taken at intervals of 20 sec. for 5 minutes.

During the experiment were taken 4 groups of pictures. The first one from the first group is shown in Fig. 1.

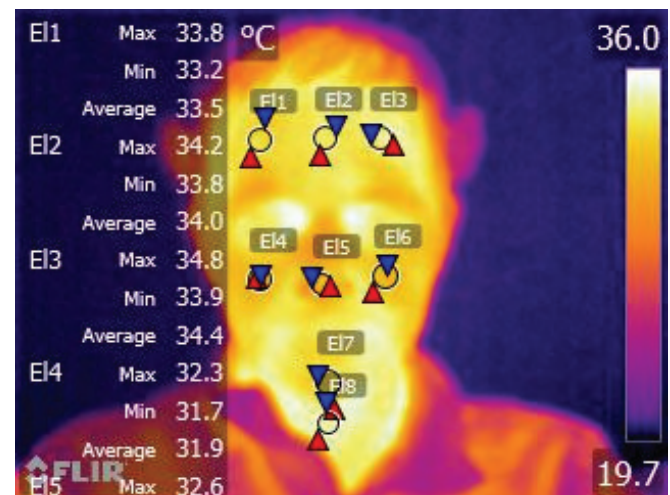


Fig.1. One of the infrared images taken during the experiment with areas of interest

They were organized and conducted in four states - relaxation, intense thinking, relaxation and intense thinking. For intensive thinking we used successive arithmetic

operations - removal of a number between 5 and 10 from 200. The aim was to make difficult but not impossible calculations.

Eight ellipses of interest were selected: EI1, EI2, EI3 – on the forehead; EI4, EI6 – in the area of the cheeks; EI5 – on the tip of the nose; EI7 – on the chin; EI8 – on the neck; . We selected ellipses, not points, to eliminate probable human little movements and noises. For the areas where the human skin is observed, is selected a radiation coefficient 0.98 [21].

The processing was made with the original software of the camera - version FLIR Tools+ 5.3.1.

#### IV. ANALYSIS

Figure 2 shows all the data obtained after the processing of data in areas EI1 to EI8. On the ordinate are combined numbers of the measurements of the four groups. The first picture is marked with 1 and the last one - with 60.

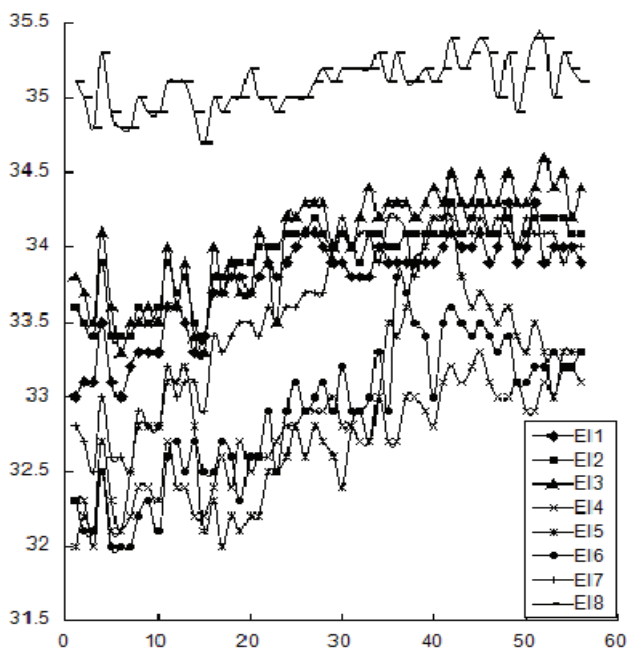


Fig.2. Results after averaging the data in the areas EI1 to EI2. On the x-axis is written the number of the measurement. It is a combination of four sets of measurements. The ordinate shows the temperature in degrees centigrade.

The results in Fig.2 are difficult to comment on. For this reason we made calculations for a variance [24] for each dataset. For a more detailed analysis we split each set into two equal parts. The results are shown in Fig.3, Fig.4, Fig.5 and Table 1.

After a careful study of the results of Fig.3, Fig.4, Fig.5, we noticed that the overall variance for the first two measurements (Series1 and Series2) is significantly greater than the second two measurements (Series3 and Series4). There is an exception for the results from area EI5. With the so selected times between the pictures, there is a clear correlation between variance and mental load for the

following areas of interest: EI2 - the first half of the data shows that the increase in mental load increases the variance; E4, E6 - in the first and in the second half of the data it can be seen that with the increase in mental load, variance decreases; E5 - in the first and in the second half of the data as well as the overall processing can be seen that when there is a reduction of mental load, variance is significantly increased.

After these calculations, we decided to do some of the 28 combinations of correlation [24] between the 8 areas of interest. The results are shown in table 2.

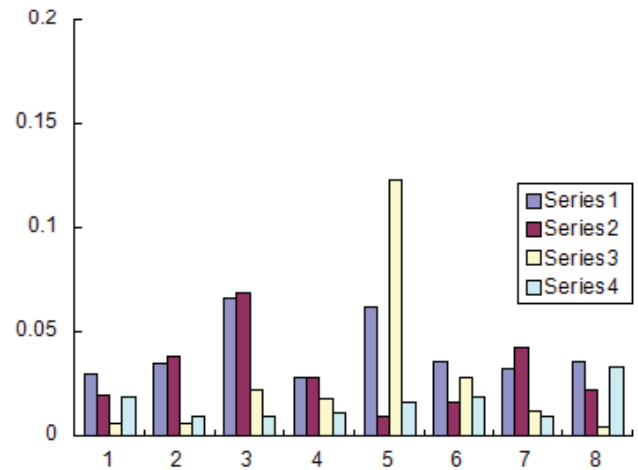


Fig.3. Results after calculating the variance based on the first half of the data set for fields of interest. The x-axis denotes the number of the respective region EI1 to EI8. The four groups of measurements are denoted as Series1 to Series4. The ordinate represents the values of calculated variance.

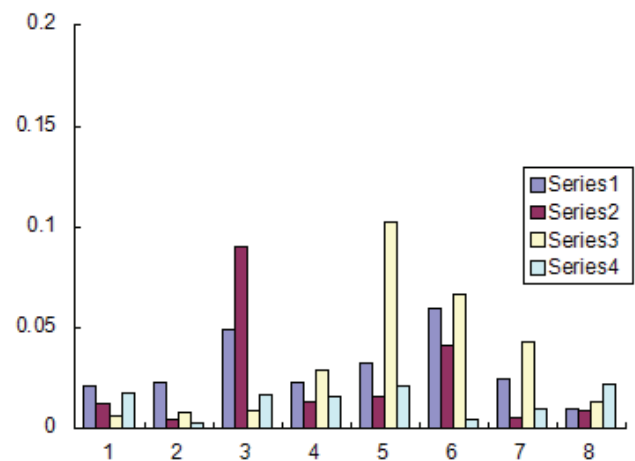


Fig.4. Results after calculating the variance based on the second half of the data set for fields of interest. The x-axis denotes the number of the respective region EI1 to EI8. The four groups of measurements are denoted as Series1 to Series4. The ordinate represents the values of calculated variance.



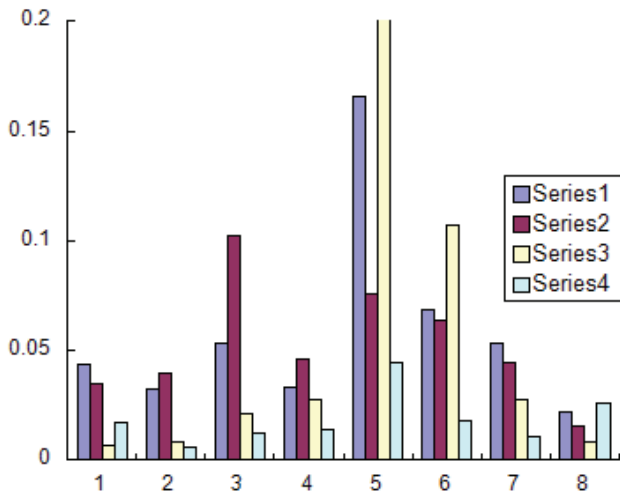


Fig.5. Results after calculating the dispersion based on the whole data set for fields of interest. The x-axis denotes the number of the respective region E11 to E18. The four groups of measurements are denoted as Series1 to Series4. The ordinate represents the values of calculated variance.

TABLE I

RESULTS AFTER CALCULATING THE DISPERSION BASED ON THE WHOLE DATA SET FOR FIELDS OF INTEREST (THE ROWS REPRESENT THE VALUES FOR EACH EXPERIMENT)

	E11	E12	E13	E14	E15	E16	E17	E18
1	0.04	0.03	0.05	0.03	0.17	0.07	0.05	0.02
2	0.03	0.04	0.10	0.05	0.08	0.06	0.04	0.02
3	0.01	0.01	0.02	0.03	0.42	0.11	0.03	0.01
4	0.02	0.01	0.01	0.01	0.04	0.02	0.01	0.03

TABLE II

CORRELATIONS BETWEEN RESULTS IN SOME SELECTED AREAS OF INTEREST

E11-E12	E11-E13	E12-E13	E12-E14	E12-E15	E12-E16
0.94	0.83	0.90	0.92	0.61	0.85
E12-E17	E12-E18	E15-E17	E15-E18	E15-E16	E15-E14
0.90	0.73	0.69	0.59	0.76	0.70

The results of the correlation analysis showed significantly poor correlation in combination with E1.5 than other areas. It is possible to think in the direction that the information we gain from this area is sufficiently different from the other areas of interest.

## V. CONCLUSION

The analyzed data provides another perspective on the myriad applications of thermal images for medical purposes. Such an analysis can be very useful in providing a new understanding of widely used technology in terms of safety and effectiveness in diagnostics. Since IR imaging is noninvasive and does not bear any risks, it can be used wherever it is useful and feasible. There is also great potential for diagnostic medicine to look for links between different points of temperature and completely different phenomena.

## REFERENCES

- [1] P. Childs, (et. all), *Practical Temperature Measurement*, Butterworth Heinemann, Elsevier, 2001.
- [2] I. Fernández-Cuevas, J. C. Bouzas Marins, J. A. Lastras, P. M. G. Carmona, S. P. Cano, M. Á. García-Concepción, M. Sillero-Quintana, "Classification of factors influencing the use of infrared thermography in humans: A review", *Elsevier, Infrared Physics & Technology*, Elsevier, 71, pp.28–55, 2015.
- [3] R. Fuksis, M. Greitans, O. Nikisins, M. Pudzs, "Infrared Imaging System for Analysis of Blood Vessel Structure", *Electronics and Electrical Engineering*, No.1(97), pp.45-48, 2010.
- [4] A. Szentkuti, H. S. Kavanagh, S. Grazio, "Infrared thermography and image analysis for biomedical use", *Periodicum Biologorum*, Vol. 113, No. 4, pp.385–392, 2011.
- [5] S. Pleshkova, A. Bekyarski, K. Peeva, "Testing Thermal Images Characteristics for Thermal Images Quality Estimation", *Latest Trends on Systems, 18th International Conference on Systems*, Vol.1, pp.251-256, 2014.
- [6] S. Pleshkova, A. Bekyarski, "Motion Detection in Thermal Images Sequence Using Wigner Distributions", *Latest Trends on Systems, 18th International Conference on Systems*, Vol.1, pp.153-156, 2014.
- [7] J. Serup, G. B. E. Jemec, G. L. Grove, *Handbook of Non-Invasive Methods and the Skin*, Second Edition, CRC Press, 2006.
- [8] B. Jones, "A reappraisal of the use of infrared thermal image analysis in medicine", 17(6), *IEEE Trans Med Imaging*, pp.1019-27, 1998.
- [9] F. Ring, A. Jung, J. Žuber, *Infrared Imaging*, IOP, 2015.
- [10] S. Nudelman, *Nuclear Medicine, Ultrasonics, and Thermography*, Springer Science & Business Media, 2013.
- [11] M. Diakides, J.D. Bronzino, D. R. Peterson, *Medical Infrared Imaging: Principles and Practices*, CRC Press, 2012.
- [12] Y. Houdas, E.F.J. Ring, *Human Body Temperature: Its Measurement and Regulation*, Springer Science & Business Media, 2013.
- [13] A. E. R. Young, T. J. Germon, N. J. Barnett, A. R. Manara and R. J. Nelson, "Behaviour of near-infrared light in the adult human head: implications for clinical near infrared spectroscopy", *British Journal of Anaesthesia*, vol. 84 (1), pp. 38–42, 2000.
- [14] F. L. da Silva, E. Niedermeyer, *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, Lippincott Williams & Wilkins, 2004.
- [15] R. Horlings, D. Datcu, L. Rothkrantz, "Emotion recognition using brain activity", In *Proceedings of the 9th International Conference on Computer Systems and Technologies and*

- Workshop for PhD Students in Computing, CompSysTech '08, pages 6:II.1–6:1, USA, 2008.
- [16] N. Schneiderman, G. Ironson, S.D. Siegel, "Stress and health: psychological, behavioral, and biological determinants", *Annual Review of Clinical Psychology*, vol. 1, pp. 607, 2005.
  - [17] E. Salazar-López, E. Domínguez, V. Juárez Ramos, J. de la Fuente, A. Meins, O. Iborra, G. Gálvez, M.A. Rodríguez-Artacho, E. Gómez-Milán, "The mental and subjective skin: Emotion, empathy, feelings and thermography", *Elsevier, Consciousness and Cognition*, 34, pp.149–162, 2015.
  - [18] C. D. Katsis, N. Katertsidis, G. Ganiatsas, and D. I. Fotiadis, "Toward Emotion Recognition in Car-Racing Drivers: A Biosignal Processing Approach", *IEEE Transactions on Systems Man and Cybernetics Part A Systems and Humans*, 38(3), May pp.502-512, 2008.
  - [19] E. Lareau, G. Simard, F. Lesage, D. Nguyen, M. Sawan, "Near infrared spectrometer combined with multichannel EEG for functional brain imaging", 2011 5th International Symposium on Medical Information and Communication Technology, IEEE Conference Publications, pp.122-126, 2011.
  - [20] F. Al-shargie, T. B. Tang, N. Badruddin, M. Kiguchi, Simultaneous measurement of EEG-fNIRS in classifying and localizing brain activation to mental stress, 2015 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), IEEE Conference Publications, pp.282-286, 2015.
  - [21] A. Guzman, M. Goryawala, M. Adjouadi, Generating thermal facial signatures using thermal infrared images, *Emerging Signal Processing Applications (ESPA)*, 2012 IEEE International Conference on, IEEE Conference Publications Pages: 21 – 24, 2012.
  - [22] K. Ammer, "The Glamorgan Protocol for recording and evaluation of thermal images of the human body", *Thermology International*, 10, 18(4), pp.125-129, 2008.
  - [23] <http://www.flir.com/cs/emea/en/view/?id=41372> - FLIR Systems, Inc. (last visit - 04.2016)
  - [24] H.C. Taneja, *Advanced Engineering Mathematics*, I. K. International Pvt Ltd, 2007.

# Contemporary Approach to Teaching GSM with USRP and OpenBTS

Slavica Marinković<sup>1</sup>, Amela Zeković<sup>2</sup>, Ivan Pavlović<sup>3</sup> and Milutin Nešić<sup>4</sup>

**Abstract** – The paper presents laboratory exercise implemented within third year Mobile Communications course in The School of Electrical and Computer Engineering of Applied Studies in Belgrade. The designed exercise follows the modern trend in teaching telecommunications based on software defined radio (SDR) platforms. The goal of the exercise is getting students relevant hands-on experience about organization and characteristics of Global System for Mobile Communication (GSM) system and motivating students to study wireless communication systems. GSM base station system is implemented with Universal Software Radio Peripheral (USRP) hardware and OpenBTS software. Students' experience with the new laboratory exercise was evaluated by questionery.

**Keywords** –Mobile communications teaching, Laboratory exercises, GSM, USRP, OpenBTS.

## I. INTRODUCTION

Software-defined radio is a system with characteristics that can be changed by software, that is, without the need to change system hardware components. In SDR the components that have typically been implemented in hardware in traditional radio systems are implemented in software, which allows for easy reconfigurability. Apart from being innovative wireless technology emerging in commercial market, SDR platforms offer great opportunities in enhancing practical teaching in the field of wireless and mobile communications [1]. The costs of SDR systems are minimal compared to conventional radio and network equipment and therefore enable academic institutions to deliver relevant, real world, practical exercises. The other important impact on the telecommunication engineering education of the current industry trends in SDR development and applications is that these changes inspire integrated SDR curriculum design [2]. The main advantages of SDR technology is flexibility in creating different communication technologies with little hardware components and usage of free, open source, publicly available software, which together makes this a very cost effective solution for various communications system implementation. This allows more rapid communications

system development and testing. Some of the examples of the implemented communications systems are: FM radio transmission and reception, digital video signal transmission, radio frequency identification, cognitive radio systems.

One of the very popular hardware platforms used for research and teaching in academia is universal software radio peripheral (USRP) [3,4]. This platform is commonly combined with LabView [4], or GNU radio software [5] to create software defined radio. Block diagram of the SDR platform is shown in Figure 1. Radio Frequency (RF) front end is used for signal transmission and reception. Intermediate frequency part of the USRP is used for analog to digital and digital to analog conversion (ADC/DAC) and signal processing by the Field Programmable Gate Array (FPGA) that can include filtering, modulation and demodulation as well as digital up conversion (DUC) or digital down conversion (DDC). The last block is used for baseband signal processing in software run at personal computer.



Figure 1 Block diagram of the SDR platform

In our laboratory exercise the software run on PC is OpenBTS application [6], open source software for GSM base transceiver station (BTS) implementation. Open BTS together with USRP implements GSM access point which allows GSM compatible phones to communicate. Open BTS is the first free open source implementation of the standardized GSM protocol stack. This application implements air interface of the GSM standard (Um interface) and mobile stations (MS) see identical interface as in the case of conventional GSM systems. The rest of the network differs from the conventional GSM in that, that it uses Internet Telephony protocols [7]. The Session Initiation Protocol (SIP) and Real-time Transport Protocol (RTP) are the two protocols that OpenBTS uses to convert GSM traffic into voice over Internet Protocol (VoIP) traffic [7]. Open source software Asterisk Private Branch Exchange (PBX) [8] is used to connect calls. OpenBTS GSM system components are shown in Figure 2 [7].

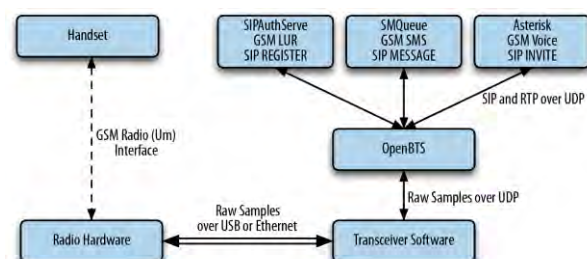


Figure 2 OpenBTS GSM system components

<sup>1</sup>Slavica Marinković is with the School of Electrical and Computer Engineering of Applied Studies, 283 Vojvode Stepe, 11000 Belgrade, Serbia (e-mail: [slavica.marinkovic@viser.edu.rs](mailto:slavica.marinkovic@viser.edu.rs)).

<sup>2</sup>Amela Zeković is with the School of Electrical and Computer Engineering of Applied Studies, 283 Vojvode Stepe, 11000 Belgrade, Serbia (e-mail: [amela.zekovic@viser.edu.rs](mailto:amela.zekovic@viser.edu.rs)).

<sup>3</sup>Ivan Pavlović is with the School of Electrical and Computer Engineering of Applied Studies, 283 Vojvode Stepe, 11000 Belgrade, Serbia (e-mail: [ivan.pavlovic@viser.edu.rs](mailto:ivan.pavlovic@viser.edu.rs)).

<sup>4</sup>Milutin Nešić is with the School of Electrical and Computer Engineering of Applied Studies, 283 Vojvode Stepe, 11000 Belgrade, Serbia (e-mail: [milutin.nesic@viser.edu.rs](mailto:milutin.nesic@viser.edu.rs)).



Figure 3 Laboratory set up picture

GSM system architecture and Um interface are studied in detail theoretically in lectures. What was lacking was a practical experience with the GSM Um interface parameters that would help students to faster adopt knowledge in the field. The implemented exercise is based on reference [7].

Hardware that we used consists of USRP N210 with SBX daughterboard, GSM quad band antenna, PC and mobile phone. The installed software is Ubuntu 14.04 LTS, OpenBTS 5.0, ASTERISK 11.7.0.4. The laboratory set up is shown in Fig 3.

## II. OVERVIEW OF LABORATORY EXERCISE

Laboratory exercise is divided in four sections: GSM configuration, identification of GSM users, testing GSM services, and measuring signal quality.

### A. GSM Configuration

Within “GSM Configuration” part of the laboratory exercise students gain experience with GSM system parameters, revise knowledge of frequency bands, downlink and uplink frequency separation, received signal strength indicator (RSSI) and examine relationship between noise levels and signal strength in respect to changes in transmit output power and receive and transmit antenna orientation.

Students set frequency band of GSM to 900MHz and select uplink/downlink channel pair by setting ARFCN (Absolute Radio Frequency Channel Number) to 124. This channel is not used by mobile operators in Serbia. The other parameters are set to default values.

Apart from examining the frequency settings, students get acquainted with the other characteristics of mobile communication systems by inspection of other parameters such parameters for setting maximum expected delay spread which is used for equalization, parameters for minimum and maximum output power used for output power control and parameters for target RSSI and signal to noise ratio (SNR) used for mobile station (MS) power control loop. Figure 4 shows an example of configuration parameters.

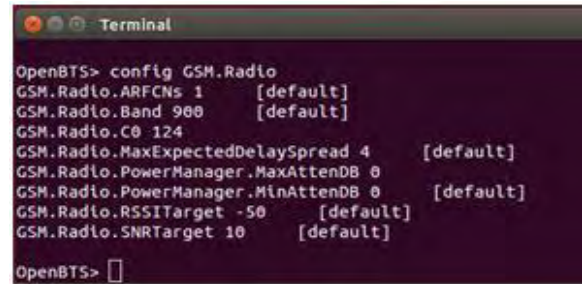


Figure 4 Configuration parameters

Students check the noise level and check whether uplink RSSI for MS is higher than noise level for at least 10 dB which was set by target SNR parameter.

It is further demonstrated that exists mutual influence of transmit and receive signal paths. This done by observing how noise levels change when transmit signal power is reduced and when antennas positions are changed.

### B. Identification of GSM Users

Within “Identification of GSM Users” part of the laboratory exercise students get familiarized with numbers that are used for identification of mobile subscribers in GSM: IMSI (International Mobile Subscriber Identity), TMSI (Temporary Mobile Subscriber Identity), MSISDN (Mobile station ISDN Number), and IMEI (International Mobile Equipment Identity).

Concrete tasks comprise of checking the list of users that have requested interaction with openBTS. This list displays IMSI, TMSI and IMEI numbers as well as parameter AUTH that denotes whether the users have been authenticated. Figure 5 shows an example of displayed parameters for users connected to openBTS system.

Students check IMSI and IMEI numbers by mobile phone application on the phone and compare it to the one reported by openBTS.

In order to be able to authenticate users and register them with the openBTS GSM network, it is necessary to create entries in the database of subscriber information that replaces GSM Home Location Register (HLR) found in a conventional GSM. Students put the following entries in the database: user name, IMSI and MSISDN, and check whether the authentication has been successful.

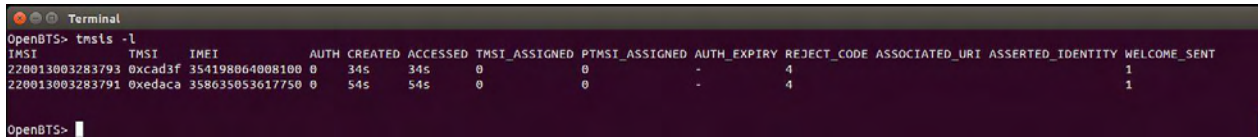
### C. Testing GSM Services

Within “Testing GSM Services” part of the laboratory exercise, students get to test voice and SMS service.

After users are registered with the network, the services can be checked by dialing test numbers predefined in the Asterisk PBX. Students send test message to 411 (predefined test number) and get, in a reply, IMSI and MSISDN numbers as parts of SMS, Figure 6 (on the next page).

By calling the test number 2602 students check the reception of the test tone. It is clearly heard in the case of good quality of communication channel. The loss of information can be heard as tone interruptions.



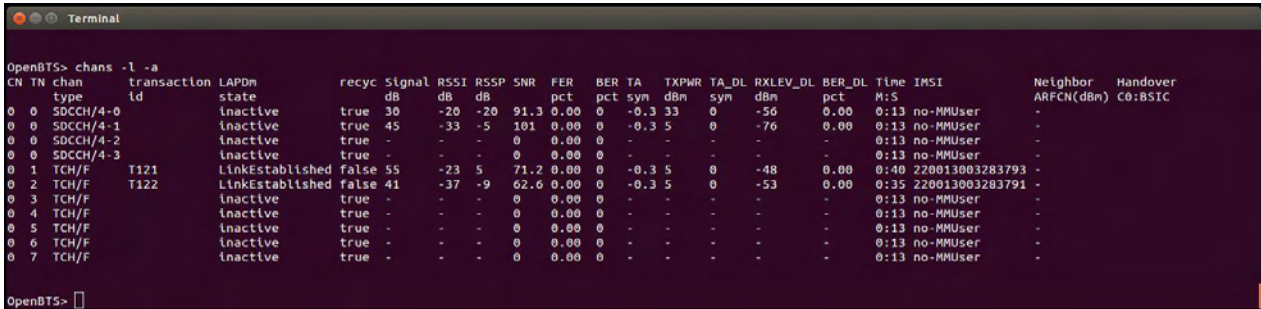


```

OpenBTS> tmsis -l
TMSI      TMSI      IMEI      AUTH  CREATED  ACCESSED  TMSI_ASSIGNED  PTMSI_ASSIGNED  AUTH_EXPIRY  REJECT_CODE  ASSOCIATED_URI  ASSERTED_IDENTITY  WELCOME_SENT
220013003283793 0xcad3f 354198064008100 0 34s 34s 0 0 - 4 1
220013003283791 0xedaca 358635053617750 0 54s 54s 0 0 - 4 1
OpenBTS>

```

Figure 5 An example of display of user's requests

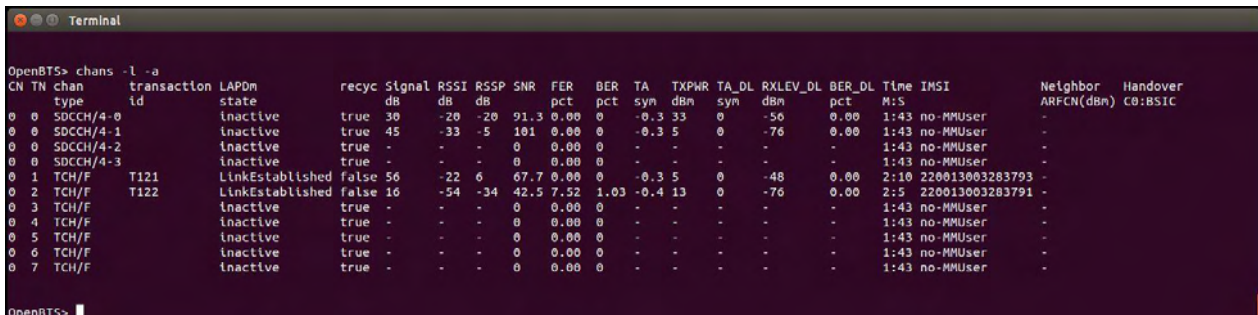


```

OpenBTS> chans -l -a
CN TN chan transaction LAPDM recyc Signal RSSI RSCP SNR FER BER TA TXPWR TA_DL RXLEV_DL BER_DL Time IMSI Neighbor Handover
type id state dB dB dB dB pct pct sym dBm sym dBm pct M:S ARFCN(dBm) C0:BSIC
0 0 SDCCH/4-0 inactive true 30 -20 -20 91.3 0.00 0 -0.3 33 0 -56 0.00 0:13 no-MMUser -
0 0 SDCCH/4-1 inactive true 45 -33 -5 101 0.00 0 -0.3 5 0 -76 0.00 0:13 no-MMUser -
0 0 SDCCH/4-2 inactive true - - - 0 0.00 0 - - - - - 0:13 no-MMUser -
0 0 SDCCH/4-3 inactive true - - - 0 0.00 0 - - - - - 0:13 no-MMUser -
0 1 TCH/F T121 LinkEstablished false 55 -23 5 71.2 0.00 0 -0.3 5 0 -48 0.00 0:40 220013003283793 -
0 2 TCH/F T122 LinkEstablished false 41 -37 -9 62.6 0.00 0 -0.3 5 0 -53 0.00 0:35 220013003283791 -
0 3 TCH/F inactive true - - - 0 0.00 0 - - - - - 0:13 no-MMUser -
0 4 TCH/F inactive true - - - 0 0.00 0 - - - - - 0:13 no-MMUser -
0 5 TCH/F inactive true - - - 0 0.00 0 - - - - - 0:13 no-MMUser -
0 6 TCH/F inactive true - - - 0 0.00 0 - - - - - 0:13 no-MMUser -
0 7 TCH/F inactive true - - - 0 0.00 0 - - - - - 0:13 no-MMUser -
OpenBTS>

```

Figure 7 Results when users are close to BS

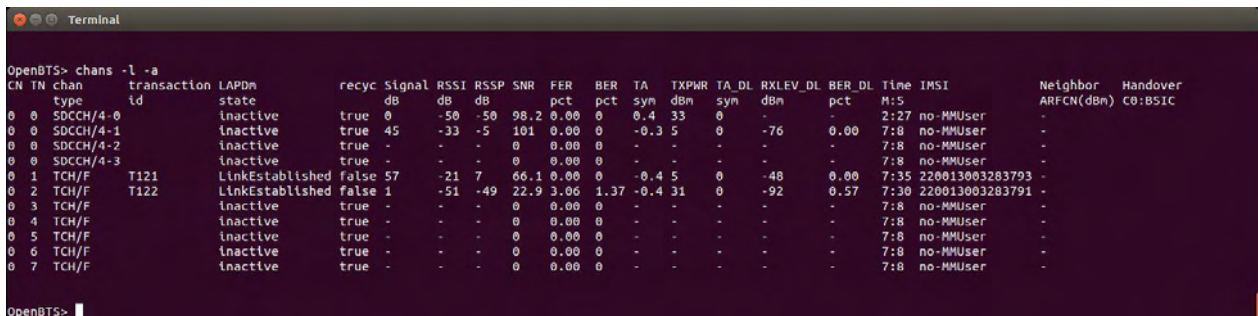


```

OpenBTS> chans -l -a
CN TN chan transaction LAPDM recyc Signal RSSI RSCP SNR FER BER TA TXPWR TA_DL RXLEV_DL BER_DL Time IMSI Neighbor Handover
type id state dB dB dB dB pct pct sym dBm sym dBm pct M:S ARFCN(dBm) C0:BSIC
0 0 SDCCH/4-0 inactive true 30 -20 -20 91.3 0.00 0 -0.3 33 0 -56 0.00 1:43 no-MMUser -
0 0 SDCCH/4-1 inactive true 45 -33 -5 101 0.00 0 -0.3 5 0 -76 0.00 1:43 no-MMUser -
0 0 SDCCH/4-2 inactive true - - - 0 0.00 0 - - - - - 1:43 no-MMUser -
0 0 SDCCH/4-3 inactive true - - - 0 0.00 0 - - - - - 1:43 no-MMUser -
0 1 TCH/F T121 LinkEstablished false 56 -22 6 67.7 0.00 0 -0.3 5 0 -48 0.00 2:10 220013003283793 -
0 2 TCH/F T122 LinkEstablished false 16 -54 -34 42.5 7.52 1.03 -0.4 13 0 -76 0.00 2:5 220013003283791 -
0 3 TCH/F inactive true - - - 0 0.00 0 - - - - - 1:43 no-MMUser -
0 4 TCH/F inactive true - - - 0 0.00 0 - - - - - 1:43 no-MMUser -
0 5 TCH/F inactive true - - - 0 0.00 0 - - - - - 1:43 no-MMUser -
0 6 TCH/F inactive true - - - 0 0.00 0 - - - - - 1:43 no-MMUser -
0 7 TCH/F inactive true - - - 0 0.00 0 - - - - - 1:43 no-MMUser -
OpenBTS>

```

Figure 8 Results when users are distant to BS



```

OpenBTS> chans -l -a
CN TN chan transaction LAPDM recyc Signal RSSI RSCP SNR FER BER TA TXPWR TA_DL RXLEV_DL BER_DL Time IMSI Neighbor Handover
type id state dB dB dB dB pct pct sym dBm sym dBm pct M:S ARFCN(dBm) C0:BSIC
0 0 SDCCH/4-0 inactive true 0 -50 -50 98.2 0.00 0 0.4 33 0 - - 2:27 no-MMUser -
0 0 SDCCH/4-1 inactive true 45 -33 -5 101 0.00 0 -0.3 5 0 -76 0.00 7:8 no-MMUser -
0 0 SDCCH/4-2 inactive true - - - 0 0.00 0 - - - - - 7:8 no-MMUser -
0 0 SDCCH/4-3 inactive true - - - 0 0.00 0 - - - - - 7:8 no-MMUser -
0 1 TCH/F T121 LinkEstablished false 57 -21 7 66.1 0.00 0 -0.4 5 0 -48 0.00 7:35 220013003283793 -
0 2 TCH/F T122 LinkEstablished false 1 -51 -49 22.9 3.06 1.37 -0.4 31 0 -92 0.57 7:30 220013003283791 -
0 3 TCH/F inactive true - - - 0 0.00 0 - - - - - 7:8 no-MMUser -
0 4 TCH/F inactive true - - - 0 0.00 0 - - - - - 7:8 no-MMUser -
0 5 TCH/F inactive true - - - 0 0.00 0 - - - - - 7:8 no-MMUser -
0 6 TCH/F inactive true - - - 0 0.00 0 - - - - - 7:8 no-MMUser -
0 7 TCH/F inactive true - - - 0 0.00 0 - - - - - 7:8 no-MMUser -
OpenBTS>

```

Figure 9 Results when users are one floor apart

Students further perform echo call test in order to check both uplink and downlink connection. In an echo call test the voice message that has been sent by MS in the uplink channel is sent back to MS on the downlink channel.

#### D. Measuring Signal Quality

Within "Measuring Signal Quality" part of the laboratory exercise, students get familiarized with signal quality measures such as SNR and Bit Error Rate (BER) in GSM.

In this part of the exercise students make a two party call and examine what channels are used and what are their characteristics. They also examine the change of the characteristics of the links as the users are increasing distance from the base station.

Figure 7 shows the results for the call established between the two users when both users and base station are in the same room. The parameters that are identified are the following: the users use two channels (TN1 and TN2) which are full rate

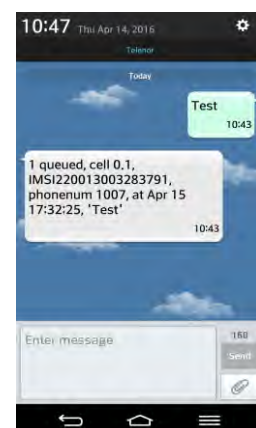


Figure 6 SMS reply message to dialling test number 411

traffic channels on time slots 1 and 2 of the C0 ARFCN. Transmit power (TXPWR) for both users is 5dBm. RSSI for user with IMSI 22013003283793 is -23 dB and for user with IMSI 22013003283791 is -37 dB. BER is zero percent.

Figure 8 shows the results when the user with IMSI number 22013003283793 is in another room. It can be seen from Figure 8 that by increasing distance of the MS from the base station, the MS transmission power is increasing from 5 to 13 dBm, in order to maintain target RSSI at -50dB. At the same time SNR has decreased and the BER is increased to 1.03 percent.

Figure 9 Shows the results when the user with IMSI 22013003283793 is one floor below base station. Further degradation of SNR and BER can be observed.

### III RESULTS OF STUDENT SURVEY

One of the main reasons for introducing this laboratory exercise was to motivate students to learn about complex technologies used in contemporary mobile communications. In order to examine students' overall experience with the newly introduced laboratory exercise questionnaire has been created. The questionnaire has been filled in by 33 students. The questionnaire consists of ten questions grouped in three categories.

The first category of questions ask students to evaluate their knowledge of SDR and GSM before and after GSM laboratory class. Majority of students have answered that their understanding of GSM and SDR systems has improved after attending laboratory class.

The second group of questions refers to evaluating complexity of the examined system and the complexity of the laboratory exercise. Majority of students have evaluated complexity of laboratory exercise as moderate.

The third group of questions evaluates student opinion on how the exercise contributes to understanding of SDR and GSM concepts, and whether it inspires further interest in studying wireless/mobile communication systems.

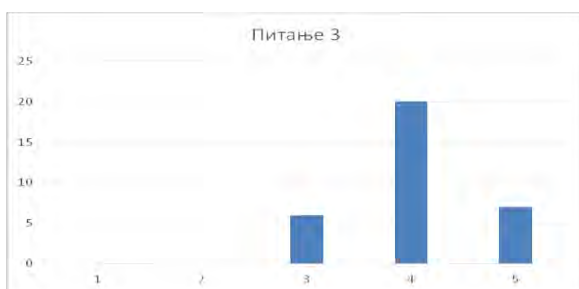


Figure 10 Students' answers to question how much does this laboratory exercise contribute to your understanding of GSM

Figure 10 shows students' answers to question: "How much does this laboratory exercise contribute to your understanding of GSM?". Majority of students (27 out of 33) have answered with scores 4 and 5 on the five point scale. Answers show that

students have evaluated that this exercise has satisfactorily contributed to their understanding of GSM system.

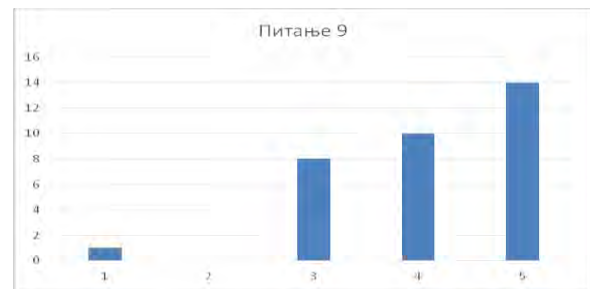


Figure 11 Students' answers to question at what degree does this laboratory exercise inspires interest in studying GSM

Figure 11 shows students' answers to question: "At what degree does this laboratory exercise inspires interest in studying GSM?". Majority of students (24 out of 33) have answered with scores 4 and 5 on the 5 point scale. Answers show that students have evaluated that laboratory exercise has motivated them to further study of GSM.

### IV CONCLUSION

Laboratory exercise "GSM mobile technology with SDR and openBTS" significantly improves teaching of GSM by integrating the teaching of theory with the practical experience. Students get exposed to modern teaching tools used at world's top universities. Student survey results show that introduced laboratory class is well accepted. We plan to further advance the study of GSM in the future by using protocol analyzer software Wireshark and to introduce laboratory on UMTS.

### REFERENCES

- [1] E. Luther1, J. Dinolfo, and S. Katti, "Software Defined Radio Provides New Opportunities for Hands-on RF Education". Proc. 2012 Canadian Engineering Education Association (CEE12) Conf. Paper 100, Winnipeg, 2012.
- [2] Dietrich, B. Farhang-Boroujeny, J. V. Urbina, S. H. Edwards, and J. H. Reed, "Software-Defined Radio: A New Paradigm for Integrated Curriculum Delivery", IEEE Communications Magazine, pp 184-193, May 2014.
- [3] <https://www.ettus.com/product/category/USRP-Networked-Series>
- [4] <http://www.ni.com/sdr/education/>
- [5] <https://gnuradio.org/redmine/projects/gnuradio>
- [6] <http://openbts.org/>
- [7] M. Iedema, *Getting Started with OpenBTS*, 2015 Range Networks. O'Reilly Media, 2015
- [8] <http://www.asterisk.org/>

**AUTHOR INDEX**

- Acevski N., 335  
 Adamović M., 247  
 Aleksandrov A., 289, 293  
 Alrahman Y. A., 17  
 Anastas M., 209  
 Anastasov J., 131  
 Andonov A., 103  
 Andonova I., 479  
 Angelov K., 33  
 Antonov S., 195  
 Antoska V., 389, 405  
 Arapinoski B., 335, 355, 359  
 Arnaudov R., 397  
 Atanasković A., 73  
 Atanasov I., 201, 205  
 Atanasov N., 385  
 Atanasovski M., 335, 355  
  
 Bankov N., 273, 277  
 Bashev V., 427, 431  
 Belovski I., 289  
 Bobev H., 225  
 Bonev B., 29  
 Bozhikova V., 259, 263  
 Bozinovski G., 343  
 Bozinovski S., 3  
 Brodić D., 161, 179, 267, 415  
 Brusev T., 281  
  
 Čabarkapa D., 53  
 Ceselkoska V., 355  
 Chiflidzhanova Z., 103  
 Chikov O., 107, 135, 213  
 Christoff N., 171  
 Ćirić T., 45  
 Cundeva M., 423  
 Cvetkovic A., 131  
 Cvetković S., 167  
  
 Damyanov C., 381  
 Denić D., 419  
 Deskovski S., 389, 405  
 Dimitrov D., 331  
 Dimitrov K., 37, 479  
 Dimitrov V., 115, 471  
 Dimitrova E., 111, 115, 471  
 Dimitrova V., 141  
 Djukic S., 149  
 Doncov N., 99  
 Doneva M., 451  
 Drača D., 83  
 Draganov D., 297  
 Draganov I., 179, 267, 415  
 Draganov N., 297  
 Draganova T., 297  
  
 Đorđević V., 65  
 Đorić A., 73  
  
 Eftimov T., 61  
  
 Filiposka S., 209  
  
 Gacovski Z., 389  
 Gadjeva E., 281  
 Gaydazhiev D., 305  
 Georgiev A., 317, 321  
 Georgiev G., 347, 455  
 Georgieva T., 111  
  
 Georgieva V., 187  
 Gerganova S., 441  
 Goranov G., 285, 293, 309  
 Gormond M. S., 171  
 Gospić N., 95  
 Gradinarova B., 475  
 Grancharova N., 183  
 Greef H., 171  
 Grigorov I., 385  
 Guliaschi V., 231, 339  
  
 Hristov M., 301  
 Hubenov P., 293  
 Hubenov E., 103  
  
 Iliev A., 343  
 Iliev B., 423  
 Iliev I., 17, 25  
 Iontchev E., 427, 431  
 Ivanova I., 411  
 Ivanova R., 157  
 Ivetić D., 247  
  
 Jagodić D., 445  
 Janevski T., 79, 91  
 Janković D., 243  
 Jankovic M., 251  
 Jevtić M., 415  
 Jocić A., 419  
 Jorda L., 171  
 Jovkovic S., 251  
  
 Kalcheva N., 225  
 Kalinovska K., 21, 33  
 Karova M., 225  
 Kireva D., 401  
 Kirilov B., 255  
 Kitanov S., 91  
 Knierzinger A., 475  
 Kokić A. A., 123  
 Kolev S., 479  
 Kolev A., 13, 25  
 Kosanovic M., 119  
 Kostadinov S., 305  
 Kostić D., 153  
 Kostov M., 335  
 Kotevska E., 335  
 Kunov G., 281  
 Kupusinac A., 247  
  
 Letskovska S., 347, 455  
 Lichev A., 273, 277  
  
 Madankov Y., 273, 277  
 Madzharov N., 285, 309  
 Maleš-Ilić N., 73  
 Mančić Ž., 49  
 Manolova A., 171  
 Mari J., 171  
 Marinković Z., 45, 65  
 Marinković S., 483  
 Marinova G., 213, 339  
 Marković V., 45, 65  
 Maslinkov I., 411  
 Miletiev R., 427, 431  
 Milijić M., 45  
 Milivojević Z., 153, 161, 191, 267  
 Miljković G., 419  
 Milosevic B., 251  
  
 Milosevic D., 459  
 Milovanovic B., 99  
 Milovanovic I., 99  
 Mironov R. P., 175  
 Mishev K., 209  
 Mitsev Ts., 41, 479  
  
 Nachev V., 381, 467  
 Nedelchev M., 13, 25  
 Nenov H., 217  
 Neshkovska R., 371, 375  
 Nešić M., 483  
 Nikolaeva D., 263  
 Nikolić S., 167  
 Nikolov A., 205  
 Nikolov N., 317, 321  
 Nikolov A., 201  
 Nikolova K., 141, 411  
 Nikovski P., 467  
  
 Ouzounov S., 301  
  
 Pacheco Cl., 435  
 Pacheco C., 435  
 Panagiev O., 57, 69  
 Panajotović A., 83  
 Panić S., 123  
 Panov E., 451  
 Papanchev T., 317, 321  
 Patil D. P., 195  
 Pavlović I., 483  
 Pavlović P., 53  
 Pavlović V., 145  
 Pencheva E., 201, 205  
 Penev I., 225  
 Petkov P., 21, 29  
 Petrov I., 79  
 Petrovic V., 49  
 Petrovska O., 405  
 Popnikolova R. M., 355, 359  
 Prlinčević B., 123  
 Pronić-Rančić O., 65  
 Pronić-Rančić L., 45  
  
 Radak Z., 95  
 Radić M., 363, 367  
 Radmanović M., 239  
 Radosavljević J., 415  
 Rahnev P., 289, 347  
 Randić S., 445  
 Rankovska V., 325  
 Rašić M., 363, 367  
 Reis A., 435  
 Ristova M., 371, 375  
  
 Savić N., 161  
 Sekulović N., 83  
 Sestrimska M., 381  
 Seymenliyski K., 347, 455  
 Shikoski J., 61  
 Simić M., 419  
 Smiljakovic V., 313  
 Sonar P. S., 195  
 Soyakov S., 183  
 Spalević P., 123  
 Spassova N., 29  
 Spirov R., 183  
 Spirovski M., 355, 359  
 Stajić Z., 363  
 Stajić Z., 367  
  
 Stanchev D., 351  
 Stanchev G., 61, 393  
 Stančić G., 167  
 Staneva Li., 289  
 Stanković R., 239  
 Stankovic Z., 99  
 Stefanova S., 463  
 Stefanovic C., 127  
 Stoeva M., 259, 263  
 Stojanović M., 167  
 Stojanović V., 153  
 Stošić B., 145  
  
 Talevski N., 359  
 Tanev A., 301  
 Tatić D., 243  
 Tchoumatchenko V., 463  
 Titova T., 381, 467  
 Tochev A., 231  
 Todorova M., 157, 225  
 Toneva D., 441  
 Tronkov I., 235  
 Trpezanovski Lj., 331  
  
 Uzunov I., 301, 305  
  
 Valkova-J. Z., 141  
 Vasilev V., 87  
 Vasileva M., 351  
 Vasileva T., 463  
 Vassilev S., 187  
 Veiga H., 435  
 Veinović M., 123  
 Velevska J., 371, 375  
 Veličković Z., 161, 191  
 Veličković M., 191  
 Velikova N., 351  
 Veskovic M., 149  
 Vietzorreck L., 45  
 Vitanov R., 393, 397  
 Vuchev A., 273, 277  
 Vujičić D., 445  
  
 Yanev Y., 441  
 Yordanov R., 427, 431  
 Yordanova S., 111  
  
 Zaharieva S. E., 221  
 Zdravkovic N., 131  
 Zeković A., 483  
 Zivanovic Z., 313  
 Zlatev J., 61

