53rd INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION, COMMUNICATION AND ENERGY SYSTEMS AND TECHNOLOGIES

(ICEST 2018)

Sozopol, Bulgaria, June 28-30, 2018

Technical University of Sofia,
Faculty of Telecommunications,
Bulgaria

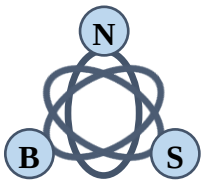
University St. Kliment Ohridski,
Faculty of Technical Sciences,
Bitola, Macedonia

University of Niš,
Faculty of Electronic Engineering,
Serbia

Proceedings of Papers

ISSN: 2603-3259 Issue: 1

53rd INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION, COMMUNICATION AND ENERGY SYSTEMS AND TECHNOLOGIES



iCEST 2018

Proceedings of Papers

Published by Publishing House, Technical University of Sofia

ICEST 2018 – 53rd INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION, COMMUNICATION AND ENERGY SYSTEMS AND TECHNOLOGIES

Sozopol, Bulgaria, June 28 - 30, 2018

Proceedings of Papers

Issue: 1

ISSN: 2603-3259 (Print)
ISSN: 2603-3267 (Online)

Number of copies printed: 30

Printed by: Publishing House, Technical University of Sofia, 2018

ICEST 2018 - LIII INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION, COMMUNICATION AND ENERGY SYSTEMS AND TECHNOLOGIES, Bulgaria, Sozopol, June 28 - 30, 2018

Proceedings of Papers

Editors: Assoc. Prof. Dr. Kalin Dimitrov
Prof. Dr. Nebojša S. Dončov
Prof. Dr. Cvetko Mitrovski

Technical Support: Assoc. Prof. Dr. Ivo R. Draganov

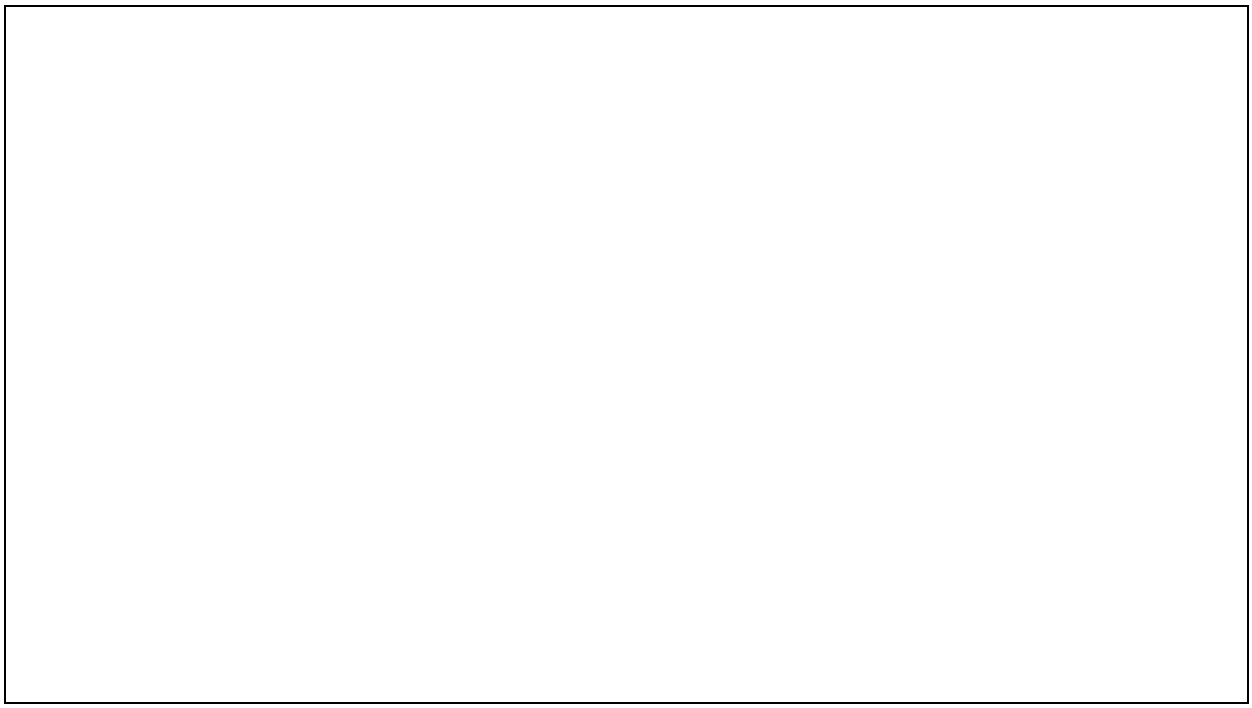
Published by: Faculty of Telecommunications, Technical University of Sofia, Bulgaria
Faculty of Electronic Engineering, University of Niš, Serbia
Faculty of Technical Sciences, St. Kliment Ohridski University, Bitola, Macedonia

Printed by: Publishing House of Technical University of Sofia, Sofia, Bulgaria

Number of copies printed: 30

Publishing of this edition has been financially supported by
Faculty of Telecommunications, Technical University of Sofia, Bulgaria

ISSN: 2603-3259 (Print)
ISSN: 2603-3267 (Online)



The LIII International Scientific Conference on Information Communication and Energy Systems and Technologies – ICEST 2018 was held from June 28 to 30 in Sozopol, Bulgaria. The Conference is for seventeenth time jointly organized by the Faculty of Telecommunications, Sofia, Bulgaria, the Faculty of Electronic Engineering, Niš, Serbia and by the Faculty of Technical Sciences, Bitola, Macedonia. As to the earlier ICEST conferences many authors from all over the world submitted their papers. This year 86 papers have been accepted for oral or poster presentation.

Two plenary lectures were given - by Prof. Vincenzo Piuri and by Prof. Nebojsa Doncov. Furthermore, the Conference also included Workshop "Reasoning-based Intelligent Systems" with chair Prof. Kazumi Nakamatsu from University of Hyogo, Japan. I am glad that all participants took the opportunity to exchange their knowledge, experiences and ideas, and also made contacts and established further collaboration. A social program, rich in events, provided more relaxing atmosphere during the meetings among colleagues.

I hope that we will meet at the next ICEST conference.

Assoc. Prof. PhD Kalin Dimitrov,

ICEST 2018 Conference Chairman

ICEST 2018

organized by

Faculty of Telecommunications,
Technical University of Sofia, Bulgaria

Faculty of Technical Sciences,
St. Kliment Ohridski University, Bitola, Macedonia

Faculty of Electronic Engineering,
University of Niš, Serbia

in co-operation with

- IEEE Bulgaria Section
- IEEE Macedonia Section
- IEEE Serbia Section

TECHNICAL PROGRAM COMMITTEE

Chairman:

K. Dimitrov	Technical University of Sofia, Bulgaria

Vice Chairmen:

B. Milovanović	Singidunum University, Niš, Serbia
C. Mitrovski	St. Kliment Ohridski University, Bitola, Macedonia

Local Coordinators:

I. Iliev	Dean of the Faculty of Telecommunications, Technical University of Sofia, Bulgaria
R. Arnaudov	Technical University of Sofia, Bulgaria

Members:

A. Jevremović	Singidunum University, Belgrade, Serbia
A. Manolova	Technical University of Sofia, Bulgaria
A. Markovski	St. Kliment Ohridski University, Bitola, Macedonia
A. Tsenov	Technical University of Sofia, Bulgaria
B. Bonev	Technical University of Sofia, Bulgaria
B. Dokić	University of Banja Luka, Bosnia and Herzegovina
B. Nikolova	Technical University of Sofia, Bulgaria
B. Stefanovski	St. Kliment Ohridski University, Bitola, Macedonia
B. Stošić	University of Niš, Serbia
B. Veselić	University of Niš, Serbia
D. Denić	University of Niš, Serbia
D. Janković	University of Niš, Serbia
D. Mančić	University of Niš, Serbia
D. Pantić	University of Niš, Serbia
D. Stojanović	University of Niš, Serbia
E. Pencheva	Technical University of Sofia, Bulgaria
G. Iliev	Technical University of Sofia, Bulgaria
G. Marinova	Technical University of Sofia, Bulgaria
G. S. Djordjević	University of Niš, Serbia
I. Atanasov	Technical University of Sofia, Bulgaria
I. Dochev	Technical University of Sofia, Bulgaria
I. Draganov	Technical University of Sofia, Bulgaria
I. Milentijević	University of Niš, Serbia
I. Jolevski	St. Kliment Ohridski University Bitola, Macedonia
I. Nedelkovski	St. Kliment Ohridski University, Bitola, Macedonia
J. Makal	Tech. University of Byalistok, Poland
K. Kasev	Technical University of Sofia, Bulgaria
K. Nakamatsu	University of Hyogo, Japan
K. Nikolova	Technical University of Sofia, Bulgaria
L. Docheva	Technical University of Sofia, Bulgaria
L. Stoimenov	University of Niš, Serbia
L. Zieleznik	Brookes University of Oxford, UK
Lj. Trpezanovski	St. Kliment Ohridski University, Bitola, Macedonia
M. Atanasovski	St. Kliment Ohridski University, Bitola, Macedonia
M. Ivković	Technical Faculty “Mihajlo Pupin” Zrenjanin, Serbia
M. Kostov	St. Kliment Ohridski University, Bitola, Macedonia
M. Lutovac	Singidunum University, Belgrade, Serbia
M. Milanova	University of Arkansas at Little Rock, USA
M. Nenova	Technical University of Sofia, Bulgaria

M. P. Radevska	St. Kliment Ohridski University, Bitola, Macedonia
M. Stojčev	University of Niš, Serbia
M. Veinović	Singidunum University, Belgrade, Serbia
N. Acevski	St. Kliment Ohridski University, Bitola, Macedonia
N. Dončov	University of Niš, Serbia
O. Panagiev	Technical University of Sofia, Bulgaria
P. Mitrevski	St. Kliment Ohridski University, Bitola, Macedonia
P. Petkov	Technical University of Sofia, Bulgaria
P. Petković	University of Niš, Serbia
P. Spalević	Faculty of Tehnical Sciences, K. Mitrovica, Serbia
R. Mironov	Technical University of Sofia, Bulgaria
R. Stanković	University of Niš, Serbia
S. Mirchev	Technical University of Sofia, Bulgaria
S. Pleshkova-Bekiarska	Technical University of Sofia, Bulgaria
S. Valtchev	NOVA-University, Lisbon, Portugal
T. Dimovski	St. Kliment Ohridski University, Bitola, Macedonia
T. Eftimov	Plovdiv University “Paisii Hilendarski”, Bulgaria
T. Mitsev	Technical University of Sofia, Bulgaria
V. Češelkoska	St. Kliment Ohridski University, Bitola, Macedonia
V. Georgieva	Technical University of Sofia, Bulgaria
V. Marković	University of Niš, Serbia
V. Poulkov	Technical University of Sofia, Bulgaria
V. Radoičić	Universitety of Belgrade, Serbia
W. Bock	University of Ottawa, Canada
Z. Jovanović	University of Niš, Serbia
Z. Milivojevic	College of Applied Technical Sciences, Niš, Serbia
Z. Stanković	University of Niš, Serbia
Z. Valkova-Djarvis	Technical University of Sofia, Bulgaria

CONFERENCE ORGANIZING COMMITTEE

Chairman:

K. Dimitrov	Technical University of Sofia, Bulgaria

International Coordinators:

B. Milovanović	University of Niš, Serbia
C. Mitrovski	St. Kliment Ohridski University, Bitola, Macedonia

Technical editor:

I. Draganov	Technical University of Sofia, Bulgaria

Members:

A. Atanasković	University of Niš, Serbia
A. Mihaylova	Technical University of Sofia, Bulgaria
B. Arapinovski	St. Kliment Ohridski University, Bitola, Macedonia
B. Stošić	University of Niš, Serbia
B. Veselić	University of Niš, Serbia
D. Janković	University of Niš, Serbia
D. Kireva-Mihova	Technical University of Sofia, Bulgaria
D. Mančić	University of Niš, Serbia
D. Mihaylova	Technical University of Sofia, Bulgaria
J. Joković	University of Niš, Serbia
J. Pargovski	St. Kliment Ohridski University, Bitola, Macedonia
K. Raynova	Technical University of Sofia, Bulgaria

K. Stoyanova	Technical University of Sofia, Bulgaria
K. Valkov	Technical University of Sofia, Bulgaria
L. Laskov	Technical University of Sofia, Bulgaria
M. Milijić	University of Niš, Serbia
M. Nedyalkova	Technical University of Sofia, Bulgaria
M. Spirovski	St. Kliment Ohridski University, Bitola, Macedonia
M. Stoyanova	Technical University of Sofia, Bulgaria
M. Todorov	Technical University of Sofia, Bulgaria
N. Hristova	Technical University of Sofia, Bulgaria
N. Maleš-Ilić	University of Niš, Serbia
N. Mojsovska	St. Kliment Ohridski University, Bitola, Macedonia
O. Pronić-Rančić	University of Niš, Serbia
S. Antonov	Technical University of Sofia, Bulgaria
T. Brusev	Technical University of Sofia, Bulgaria
T. Dimitrijević	University of Niš, Serbia
T. Dimovski	St. Kliment Ohridski University, Bitola, Macedonia
V. Dolapchieva	Technical University of Sofia, Bulgaria
V. Jović	University of Niš, Serbia
V. Marković	University of Niš, Serbia
V. Stoynov	Technical University of Sofia, Bulgaria
Z. Marinković	University of Niš, Serbia

CONFERENCE SECRETARIAT

K. Stoyanova Technical University of Sofia, Bulgaria

Address:

ICEST 2018 Conference
Technical University of Sofia
Faculty of Telecommunications
8, Kl. Ohridski, Blvd, 1000 Sofia, Bulgaria
1 bldg, 2 fl, rooms 1254, 1257
phone: +359 2 965 3145, +359 2 965 2272
E-mail: icest@tu-sofia.bg
Web: http://www.icestconf.org

LIST OF ICEST 2018 REVIEWERS

Venera Andonova
Technical University of Sofia, Bulgaria

Kliment Angelov
Technical University of Sofia, Bulgaria

Dimitar Atamyan
Technical University of Sofia, Bulgaria

Aleksandar Atanasković
University of Niš, Serbia

Ivaylo Atanasov
Technical University of Sofia, Bulgaria

Metodija Atanasovski
St. Kliment Ohridski University, Bitola, Macedonia

Georgi Balabanov
Technical University of Sofia, Bulgaria

Tihomir Brusev
Technical University of Sofia, Bulgaria

Dragan Denić
University of Niš, Serbia

Stojce Deskovski
St. Kliment Ohridski University, Bitola, Macedonia

Tijana Dimitrijević
University of Niš, Serbia

Milan Dinčić
University of Niš, Serbia

Goran Đorđević
University of Niš, Serbia

Ivo Dochev
Technical University of Sofia, Bulgaria

Lilyana Docheva
Technical University of Sofia, Bulgaria

Nebojša Dončov
University of Niš, Serbia

Predrag Eferica
University of Niš, Serbia

Elissaveta Gadjeva
Technical University of Sofia, Bulgaria

Veska Georgieva
Technical University of Sofia, Bulgaria

Vasil Guliashki
Bulgarian Academy of Sciences

Edmond Hajrizi
University for Business and Technology, Kosovo

Georgi Iliev
Technical University of Sofia, Bulgaria

Aleksandar Janjić
University of Niš, Serbia

Jugoslav Joković
University of Niš, Serbia

Ilija Jolevski
St. Kliment Ohridski University, Bitola, Macedonia

Aleksandra Jovanović
University of Niš, Serbia

Goran Jovanović
University of Niš, Serbia

Jelena Jovanović
University of Niš, Serbia

Zoran Jovanović
University of Niš, Serbia

Kiril Kasev
Technical University of Sofia, Bulgaria

Zivko Kokolanski
Ss. Cyril and Methodius University, Skopje

Stanyo Kolev
Technical University of Sofia, Bulgaria

Pavlina Koleva
Technical University of Sofia, Bulgaria

Lidija Korunović
University of Niš, Serbia

Mitko Kostov
St. Kliment Ohridski University, Bitola, Macedonia

Georgi Kunov
Technical University of Sofia, Bulgaria

Lyubomir Laskov
Technical University of Sofia, Bulgaria

Ludvig Lubih
Technical University of Sofia, Bulgaria

Nataša Maleš-Ilić
University of Niš, Serbia

Dragan Mančić
University of Niš, Serbia

Violeta Manevska
St. Kliment Ohridski University, Bitola, Macedonia

Agata Manolova
Technical University of Sofia, Bulgaria

Zlatica Marinković
University of Niš, Serbia

Galya Marinova
Technical University of Sofia, Bulgaria

Aleksandar Markoski
St. Kliment Ohridski University, Bitola, Macedonia

Vera Marković
University of Niš, Serbia

Rosen Miletiev
Technical University of Sofia, Bulgaria

Dejan Milić
University of Niš, Serbia

Marija Milijić
University of Niš, Serbia

Aleksandar Milosavljević
University of Niš, Serbia

Nenad Milosević
University of Niš, Serbia

Seferin Mirchev
Technical University of Sofia, Bulgaria

Rumen Mironov
Technical University of Sofia, Bulgaria

Cvetko Mitrovski
St. Kliment Ohridski University, Bitola, Macedonia

Milica Naumović
University of Niš, Serbia

Marin Nedelchev
Technical University of Sofia, Bulgaria

Maria Nenova
Technical University of Sofia, Bulgaria

Nikolay Neshov
Technical University of Sofia, Bulgaria

Jelena Nikolić
University of Niš, Serbia

Tatjana Nikolić
University of Niš, Serbia

Georgi Nikolov
Technical University of Sofia, Bulgaria

Boyanka Nikolova
Technical University of Sofia, Bulgaria

Kamelia Nikolova
Technical University of Sofia, Bulgaria

Oleg Panagiev
Technical University of Sofia, Bulgaria

Aleksandra Panajotović
University of Niš, Serbia

Ivailo Pandiev
Technical University of Sofia, Bulgaria

Roberto Pasic
St. Kliment Ohridski University, Bitola, Macedonia

Evelina Pencheva
Technical University of Sofia, Bulgaria

Zoran Perić
University of Niš, Serbia

Peter Petkov
Technical University of Sofia, Bulgaria

Milutin Petronijević
University of Niš, Serbia

Snejana Pleshkova-Bekiarska
Technical University of Sofia, Bulgaria

Olivera Pronić-Rančić
University of Niš, Serbia

Božidar Radenković
University of Belgrade, Serbia

Nikola Rajaković
University of Belgrade, Serbia

Ralitza Raynova
Technical University of Sofia, Bulgaria

Blagoja Ristevski
St. Kliment Ohridski University, Bitola, Macedonia

Blaž Rodič
Faculty of Information Studies, Slovenia

TABLE OF CONTENTS

ORAL SESSIONS

PLENARY SESSION

WORKSHOP REASONING-BASED INTELLIGENT SYSTEMS (WRIS'2018)

CEEPUS WORKSHOP "MODELLING, SIMULATION AND COMPUTER-AIDED DESIGN IN ENGINEERING AND MANAGEMENT"

RADIO COMMUNICATIONS, MICROWAVES, ANTENNAS

DIGITAL IMAGE AND SIGNAL PROCESSING, AND ELECTRONICS

COMPUTER SYSTEMS AND INTERNET TECHNOLOGIES

TELECOMMUNICATION SYSTEMS AND TECHNOLOGY, AND REMOTE ECOLOGICAL MONITORING

INFORMATICS AND COMPUTER SCIENCE, AND ENGINEERING EDUCATION

ENERGY SYSTEMS AND EFFICIENCY, AND MEASUREMENT SCIENCE AND TECHNOLOGY

POSTER SESSIONS

POSTER SESSION 1 - RADIO COMMUNICATIONS, MICROWAVES, ANTENNAS AND MEASUREMENTS SCIENCE AND TECHNOLOGY

POSTER SESSION 2 - CONTROL SYSTEMS AND REMOTE ECOLOGICAL MONITORING

POSTER SESSION 3 – ENERGY SYSTEMS AND EFFICIENCY, AND ELECTRONICS

POSTER SESSION 4 – TELECOMMUNICATION SYSTEMS AND TECHNOLOGY, COMPUTER SYSTEMS AND INTERNET TECHNOLOGIES, AND INFORMATICS AND COMPUTER SCIENCE

POSTER SESSION 5 – SIGNAL PROCESSING AND ENGINEERING EDUCATION

ORAL SESSIONS

PLENARY SESSION

Integrating Technologies for Ambient Intelligence: Convergence of Artificial Intelligence, Machine Learning, Biometrics, Cloud Computing, Internet of Things, and Cyber-physical Systems

Vincenzo Piuri[1]

Abstract – Adaptability and advanced services for ambient intelligence require an intelligent technological support for understanding the current needs and the desires of users in the interactions with the environment for their daily use, as well as for understanding the current status of the environment also in complex situations. This infrastructure constitutes an essential base for smart living. Various technologies are nowadays converging to support the creation of efficient and effective infrastructures for ambient intelligence.

Artificial intelligence can provide flexible techniques for designing and implementing monitoring and control systems, which can be configured from behavioral examples or by mimicking approximate reasoning processes to achieve adaptable systems. Machine learning can be effective in extracting knowledge form data and learn the actual and desired behaviors and needs of individuals as well as the environment to support informed decisions in managing the environment itself and its adaptation to the people's needs.

Biometrics can help in identifying individuals or groups: their profiles can be used for adjusting the behavior of the environment. Machine learning can be exploited for dynamically learning the preferences and needs of individuals and enrich/update the profile associated either to such individual or to the group. Biometrics can also be used to create advanced human-computer interaction frameworks.

Cloud computing environments will be instrumental in allowing for world-wide availability of knowledge about the preferences and needs of individuals as well as services for ambient intelligence to build applications easily.

This talk analyzes the opportunities offered by these technologies to support the realization of adaptable operations and intelligent services for smart living in an ambient intelligent infrastructures.

[1] Prof. Vincenzo Piuri, FIEEE, is with the Department of Computer Science, Università degli Studi di Milano, Italy,
E-mail: vincenzo.piuri@unimi.it ,
Web: http://www.di.unimi.it/piuri

Overview of Neural Network Approaches in the Localization of Stochastic EM Sources

Zoran Stankovic[1], Nebojsa Doncov[1], Ivan Milovanovic[2], Bratislav Milovanovic[2]

Abstract – **In this paper we present a brief overview of the results achieved in the field of application of artificial neural networks for a spatial localization of stochastic EM radiation sources. The theory of stochastic sources near-field representation in the form of antenna arrays with dipole elements and correlation matrix calculation in the far-field scan area as well as the architecture of developed neural models for the 1D and 2D DoA estimation are brefly presented here. Scenarios when sources are moving along one direction and where their position is determined by only one protrusion angle in relation to the linear antenna array and when stochastic springs move in a plane that is parallel to the rectangular antenna array so that their position is determined with two spatial angles are both considered in the paper. In addition, a few example are given to ilustrate the presented neural model capability to efficiently perform the DoA estimation for either uncorrelated or partially correlated sources.**

Keywords – **Source localization, Stochastic radiation, Moveable sources, Correlation matrix, Neural networks.**

I. INTRODUCTION

Today, there are many techniques aiming to improve the performances of different wireless communication systems. One of them is a technique based on antenna arrays with adaptive beamforming capabilities [1,2]. It offers possibilities to minimize or even eliminate interferences and to enhance information signals at the reception by using antennas with radiation pattern optimized from the knowledge of interference sources positions. Location of unwanted signal sources are usually determined by employing some of the available DoA (Direction of Arrival) estimation algorithms, e.g. MUSIC [1-3]. These algorithms are usually very accurate but majority of them require complex and hardware- and time-consuming mathematical operations which limit their applications in real working environment. One possible alternative to these algorithms is based on using artificial neural networks (ANNs) [4-16]. They provide the DoA estimation of almost same precision as e.g. MUSIC but with much faster run-time due to avoidance of exhausting calculations.

Authors of this paper have considered in the past the application of ANNs for the one dimensional (1D) and two-dimensional (2D) DoA estimation of deterministic electromagnetic (EM) radiation sources [4-6]. Stochastic EM sources have to be also considered from the interference point of view [17,18] and their angular positions have been determined by using the ANN models developed by the authors of this paper in [7-14]. In [7-12], authors have proposed the neural model for the 1D DoA estimation of moving stochastic sources whose locations were characterized only by one angular coordinate (azimuth). In [7-9] neural models for the DoA estimation of either single or multiple uncorrelated stochastic sources were developed while in [10] an enhancement of these models was done in order to be able to first determine the number of sources present in the monitoring space and then perform the DoA estimation. In [11-12] neural models for 1D DoA estimation of stochastic sources, which are mutually partially correlated, were proposed. Also in [13,14] neural models for 2D DoA estimation were developed. In [13], the model was applied to determine two angular spatial coordinates of one stochastic source, and in [14] for determining angular coordinates of two uncorrelated stochastic sources moving in 2D space (plane).

[1]Zoran Stankovic and Nebojsa Doncov are with the Faculty of Electronic Engineering, University of Nis, A. Medvedeva 14, 18000 Nis, E-mail: [zoran.stankovic, nebojsa.doncov]@elfak.ni.ac.rs.

[2]Ivan Milovanovic and Bratislav Milovanovic are with the Singidunum University, DLS center Nis, 18000 Nis, ivanshix@gmail.com, batam@pogled.net.

II. STOCHASTIC SOURCE LOCATION BASED ON 1D DOA ANN MODELING

II.A. Correlation matrix calculation for 1D source movement

In [17,18], a stochastic source was modelled by a uniform antenna array with N dipole elements at mutual distance equal to the half of the operating wavelength (Fig.1). The level of correlation between dipoles feed currents $\mathbf{I}=[I_1, I_2, \ldots, I_N]$), $\mathbf{c}^I(\omega)$, describing the noisy source radiation, is of the form [17,18]:

$$\mathbf{c}^I(\omega) = \lim_{T\to\infty} \frac{1}{2T}\left[I(\omega)I(\omega)^H\right] \tag{1}$$

To find the total electric field in the far-field sampling point, vector $\mathbf{M}$, comprising the contribution of each antenna array element with radiation pattern $F(\theta)$, is needed:

$$\mathbf{M}(\theta) = jz_0 \frac{F(\theta)}{2\pi r}\left[\frac{e^{jkr_1}}{r_1} \quad \frac{e^{jkr_2}}{r_2} \ldots \frac{e^{jkr_N}}{r_N}\right] \tag{2}$$

$$E(\theta) = \mathbf{M}(\theta)\mathbf{I} \tag{3}$$

where z_0 and k are free-space impedance and phase constant, respectively, $r_1, r_2, \ldots, r_N$ are the distances between the far-

field sampling point and antenna array elements and $E(\theta)$ is the total electric field intensity in sampling point in far-field positioned at angle θ in azimuth plane with the respect to the first element of antenna array. If there are M sampling points in the far-field, then notation $r_{i,m}$, representing the distance between i-th element ($1 \le i \le N$) in the antenna array and m-th sampling point in the far-field ($1 \le m \le M$) (see Fig. 1), can be used. Position of these points (Y_1 Y_2, ..., Y_M) in the far-field, is described by the azimuth plane angles (θ_1), (θ), ... , (θ_M), determined as previously said with the respect to the first element of antenna array. The correlation matrix of signals sampled at these points can be calculated as:

$$\mathbf{C}_E[i,j] = \mathbf{M}(\theta_i)\mathbf{c}^I\mathbf{M}(\theta_j)^H \quad i=1,...,M \;\; j=1,...,M \tag{4}$$

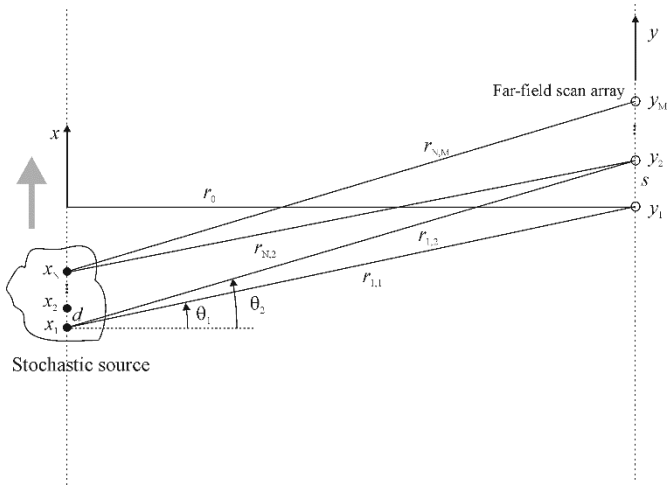


Figure 1: The position of stochastic source in azimuth plane with respect to the location of EM field sampling points in the far-field scan area [8].

For the multiple stochastic sources in the azimuth plane, the level of EM field and the elements of correlation matrix $\mathbf{C}_E$ in particular far-field sampling point, are calculated by the superposition of radiation field from all sources. If there are S stochastic sorces, then the vector M has a form [8]:

$$\mathbf{M}(\theta) = jz_0 \frac{F(\theta)}{2\pi} \cdot \left[\frac{e^{jkr_1^{(1)}}}{r_1^{(1)}} \cdots \frac{e^{jkr_N^{(1)}}}{r_N^{(1)}} \frac{e^{jkr_1^{(2)}}}{r_1^{(2)}} \cdots \frac{e^{jkr_N^{(2)}}}{r_N^{(2)}} \cdots \frac{e^{jkr_1^{(S)}}}{r_1^{(S)}} \cdots \frac{e^{jkr_N^{(S)}}}{r_N^{(S)}} \right] \tag{5}$$

where $r_i^{(j)}$ is the distance between i-th element in antenna array, representing j-th stochastic source, and the sampling point in far-field, while the feed currents vector is:

$$\mathbf{I} = \left[I_1^{(1)} \ldots I_N^{(1)} \; I_1^{(2)} \ldots I_N^{(2)} \ldots I_1^{(S)} \ldots I_N^{(S)} \right] \tag{6}$$

where $I_i^{(j)}$ is the feed current of i-th element in antenna array representing j-th stochastic source [8].

By using Eqs.(3-6), the EM field in the sampling point in the far-field, as well as the elements of correlation matrix $\mathbf{C}_E$ can be obtained. In addition, if unknown, matrix $\mathbf{c}^I(\omega)$ can be determined by measuring the electric field in the near-field of stochastic source [8].

II.B. Neural model for 1D DoA estimation

The neural model, based on MLP ANN, is trained to extract information about position of S stochastic sources in azimuth plane from the correlation matrix $\mathbf{C}_E$ [8]:

$$\boldsymbol{\theta} = f(\mathbf{C}_E) \tag{7}$$

In Eq.(7), vector $\boldsymbol{\theta}$ is azimuthal angle vector of stochastic sources, $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_S]$, while the elevation coordinates of radiation sources position are in this case neglected. The architecture of developed neural model is shown in Fig.2 and it can be represented as:

$$\mathbf{y}_l = F(\mathbf{w}_l \mathbf{y}_{l-1} + \mathbf{b}_l) \quad l = 1,2 \tag{8}$$

where $\mathbf{y}_{l-1}$ vector describes 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 [8].. F is the activation function of neurons in hidden layers and in this case it is a hyperbolic tangent sigmoid transfer function:

$$F(u) = \frac{e^u - e^{-u}}{e^u + e^{-u}} \tag{9}$$

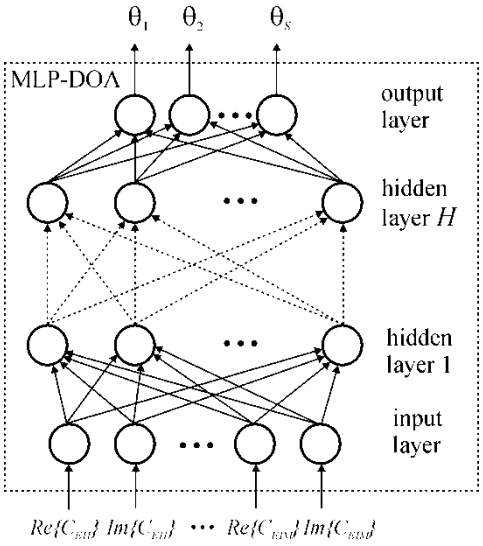


Figure 2: Architecture of MLP neural model for 1D DOA estimation of stochastic EM source signal in azimuth plane [8]

In order to perform mapping it is sufficient to take only the first column of correlation matrix and therefore $\mathbf{y}_0$ = [$Re\{\mathbf{C}_E[1,1]\}$ $Im\{\mathbf{C}_E[1,1]\}$,..., $Re\{\mathbf{C}_E[1,M]\}$ $Im\{\mathbf{C}_E[1,M]\}$]. Also, $\boldsymbol{\theta}$ is given as $\boldsymbol{\theta} = \mathbf{w}_3\mathbf{y}_2$ where $\mathbf{w}_3$ is a connection weight matrix between neurons of last hidden layer and neurons in output layer [8].. The optimization of weight matrices and biases values during the training allows ANN to

approximate the mapping with the wanted accuracy. The general designation for the MLP neural model is MLPH-N_1-…-N_i-…-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.

II.C. 1D DoA estimation - modelling results for uncorrelated sources

The application of neural model for the 1D DoA estimation of uncorrelated stochastic sources will be illustrated here on the example of two moveable sources (S=2). The training of ANN is conducted when there are four equidistant sampling points (M=4) in the far-field scan area, at the mutual distance $d = \lambda/2$, located 100 m from stochastic sources (Table I) [8]. Angular position in the azimuth plane of two sources at arbitrary angle distance will be determined from the correlation matrix information in the far field scan area. Each of two stochastic sources is described by antenna array with four dipole elements (N=4). The feed currents of four dipole elements are mutually uncorrelated so that $\mathbf{c}^I$ is the unit diagonal matrix. By using Eqs.(4-5) for N=4 and M=4, 861 and 276 uniformly distributed samples are generated for training and testing, respectively, in the range [-80° 80°] at the working frequency of 7.5 GHz. Levenberg-Marquartd method with prescribed accuracy of 10^{-4} is used as a training algorithm. The testing results for six MLP models with the lowest average case error are shown in Table II, and MLP2-16-8 is chosen as representative neural model.

TABLE I
THE VALUES OF PARAMETERS USED IN SAMPLING PROCESS [8].

Frequency	f = 7.5 GHz
Number of sources	S = 2
Number of antenna array elements per one source	N = 4
Sampling points distance from source trajectory	r_0 = 100 m
Number of sampling points	M = 4
Mutual distance of the sampling points	$s = \lambda/2$ (0.02 m)

TABLE II
TESTING RESULTS FOR SIX MLP NEURAL MODELS WITH THE BEST AVERAGE ERRORS STATISTICS [8].

MLP model	WCE [%]	ACE [%]
MLP2-16-8	1.81	0.42
MLP2-14-14	2.26	0.42
MLP2-18-14	2.67	0.39
MLP2-20-10	2.71	0.38
MLP2-16-16	3.27	0.39
MLP2-16-11	3.76	0.41

The neural model simulation of testing samples set shows a very good agreement between the output values of neural model and referent azimuth values for two sources (Fig.3 and Fig.4) [8]. In addition, a good agreement with results obtained by MUSIC algorithm can be observed.

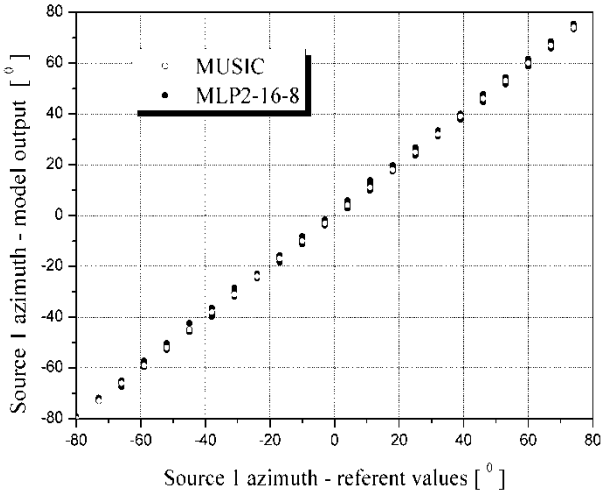


Figure 3: Comparison of MLP2-16-8 model output 1 (azimuth of source 1) with MUSIC and referent azimuth values [8].

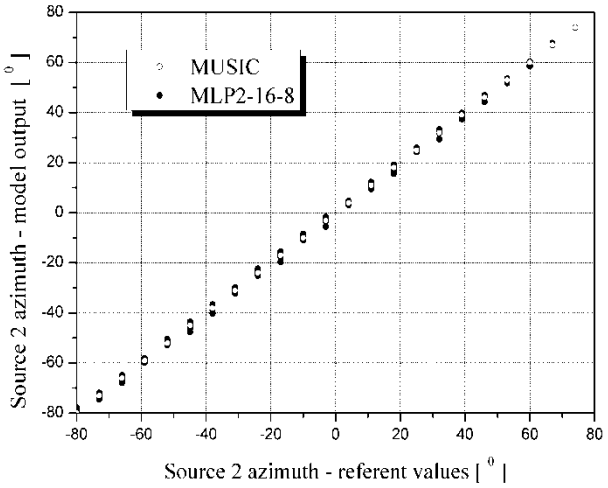


Figure 4: Comparison of MLP2-16-8 model output 2 (azimuth of source 2) with MUSIC and referent azimuth values [8].

II.D. 1D DoA estimation - modelling results for partially correlated sources

MLP neural model, whose architecture was presented in II.A section, can be also applied for 1D DoA estimation of partially correlated stochastic sources. This capability will be illustrated for two EM sources (S=2) that independently move along linear trajectory distant r_0 = 100 m from the antenna array, with whom the signal is sampled at frequency f = 28 GHz [11]. Each stochastic source is represented by a dipole (N=1) whose axis is normal on the azimuthal plane.

The scanning width of antenna array in the azimuth is [-30° 30°]. Partial correlation between sources is:

$$\mathbf{c}^I = \begin{bmatrix} 1 & c_{12}^I \\ c_{21}^I & 1 \end{bmatrix} \tag{10}$$

where $c = c_{12}{}^I = c_{21}{}^I$ is in the range $0.05 \le c \le 0.8$ [11. If sources are uncorrelated then c is equal to zero. For partially correlated and uncorrelated cases we generated separate sample sets for network training and testing, while using common parameters of sampling that are given in Table III [11].

For training and testing samples generation we use relations (4) and (5) that establish inverse mapping from that of the MLP DoA model

$$\mathbf{C}_E^t = f_{DoA}^{-1}\left(\theta_1^t, \theta_1^t, c\right) \tag{11}$$

Samples for neural network training and testing are given in the format $\{(\mathbf{x}^t(\theta^t_1, \theta^t_2, c^t), \theta^t_1, \theta^t_2)\}$, where $\mathbf{x}^t$ is vector of input model values $\mathbf{x}^t = [Re\{\mathbf{C}_E{}^t[1,1]\},\ldots, Re\{\mathbf{C}_E{}^t[1,M]\}, Im\{\mathbf{C}_E{}^t[1,1]\},\ldots, Im\{\mathbf{C}_E{}^t[1,M]\}]^T$. For each element of vector $\mathbf{x}^t$ we used uniform distribution of samples for azimuth angles of radiation source location and correlation of the form $[\theta^t_{min} : \theta^t_{step} : \theta^t_{max}]$ and $[c^t_{min} : c^t_{step} : c^t_{max}]$, where θ^t_{min} [°] and c^t_{min} represent the lowest limit of distribution, θ^t_{max} [°] and c^t_{max} represent the highest limit of distribution, while θ^t_{step} and c^t_{step} represent uniform sampling steps [11].

TABLE III
THE VALUES OF PARAMETERS USED IN SAMPLING PROCESS [11]

Frequency	f = 28 GHz
Number of sources	S = 2
Number of antenna array elements per one source	N = 1
Sampling points distance from linear source trajectory	r_0 = 100 m
Number of sampling antenna array sensors	M = 6
Mutual distance of the antenna sensors	$s = \lambda/2$ (5.4 mm)

For neural model training and testing in the uncorrelated case the following sets were generated [11]:

TRAINING_A set (28920 samples):

$$\left\{ \begin{array}{l} (\mathbf{x}^t(\theta_1^t, \theta_2^t, 0), \theta_1^t, \theta_2^t) \mid \\ \theta_1^t \in [-30 : 0.25 : 30],\ \theta_2^t \in [-30 : 0.25 : 30],\ \theta_1^t > \theta_2^t \end{array} \right\} \tag{12}$$

TEST_A set (5050 samples):

$$\left\{ \begin{array}{l} (\mathbf{x}^t(\theta_1^t, \theta_2^t, 0), \theta_1^t, \theta_2^t) \mid \\ \theta_1^t \in [-30 : 0.6 : 30],\ \theta_2^t \in [-30 : 0.6 : 30],\ \theta_1^t > \theta_2^t \end{array} \right\} \tag{13}$$

For neural model training and testing in the partially correlated case the following sets were generated:

TRAINING_B set (29280 samples):

$$\left\{ \begin{array}{l} (\mathbf{x}^t(\theta_1^t, \theta_2^t, c^t), \theta_1^t, \theta_2^t) \mid \\ \theta_1^t \in [-30 : 1 : 30],\ \theta_2^t \in [-30 : 1 : 30],\ \theta_1^t > \theta_2^t, \\ c^t \in [0.05 : 0.05 : 0.8] \end{array} \right\} \tag{14}$$

TEST_B set (3036 samples):

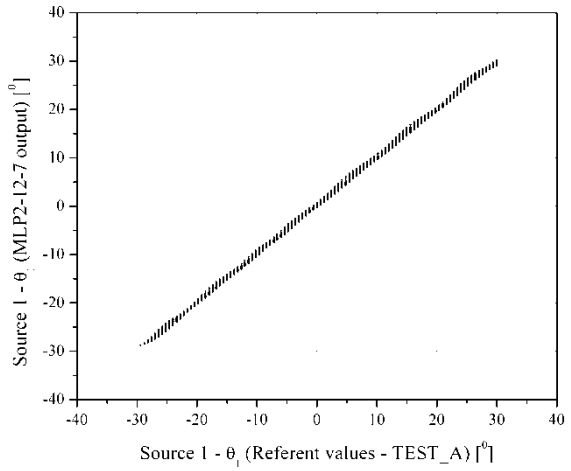
$$\left\{ \begin{array}{l} (\mathbf{x}^t(\theta_1^t, \theta_2^t, c^t), \theta_1^t, \theta_2^t) \mid \\ \theta_1^t \in [-30 : 2.7 : 30],\ \theta_2^t \in [-30 : 2.7 : 30],\ \theta_1^t > \theta_2^t, \\ c^t \in [0 : 0.07 : 0.8] \end{array} \right\} \tag{15}$$

Samples in the TRAINING_A and TEST_A sets were generated under conditions of uncorrelated radiation of two stochastic sources while the samples in the TRAINING_B set are generated under conditions of partial correlation in the range $0.05 \le c \le 0.8$. Set TEST_B contains samples that are generated also under variable radiation correlation in above range but also it has samples that are generated under no correlation. Different number of MLP ANNs with two hidden layers (H = 2) and different number of neurons in them (MLP2-N_1-N_2 , $4 \le N_1, N_2 \le 23$) were considered during the training, For the testing the values of worst case error (WCE) and average error (ACE) [4,5] were taken into account for model accuracy estimation.

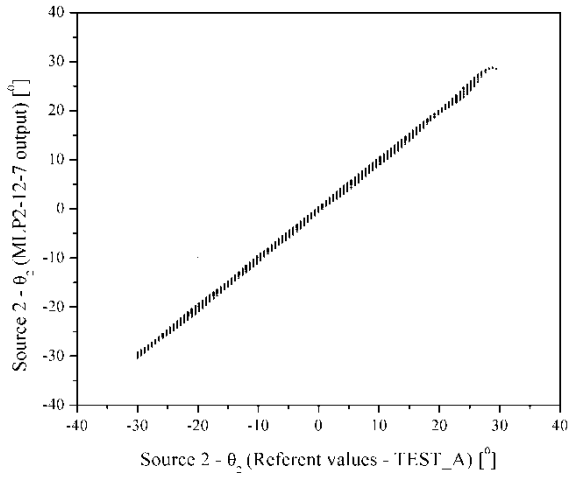
After training the MLP neural models with the set TRAINING_A, the results of the testing of the six models with lowest value of WCE on the set TEST_A are shown in Table IV [11]. Figs. 5(a) and 5(b) [11] present a scattering diagram of MLP2-12-7 model on TEST_A set (this model has shown lowest WCE on that set). It can be seen that all six models show high accuracy in source location estimation. But if in the network input we deliver samples that are generated with some source radiation correlation (set TEST_B) then the models show a high WCE value or high imprecision in source location estimation. That may be seen also in Table IV [11] also from scattering diagram of MLP2-12-7 model on TEST_B set (Figs 6(a) and 6(b)) [11].

TABLE IV
TESTING RESULTS FOR SIX MLP NEURAL MODELS TRAINED ON TRAINING_A SET WITH THE BEST AVERAGE ERRORS STATISTICS [11]

MLP model	TEST_A set		TEST_B set	
	WCE [%]	ACE [%]	WCE [%]	ACE [%]
MLP2-12-7	2.25	0.36	60.63	6.76
MLP2-13-13	2.26	0.38	87.81	6.01
MLP2-11-4	2.38	0.37	87.31	7.16
MLP2-16-16	2.43	0.37	59.02	6.07
MLP2-12-12	2.51	0.37	89.37	7.17
MLP2-12-5	2.55	0.37	83.99	6.91



(a)



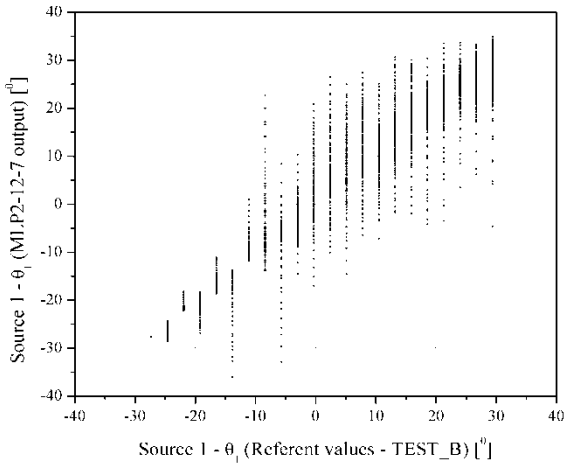
(b)

Figure 5: Scattering diagram of MLP2-12-7 model θ_1 output (a), and θ_2 output (b), on TEST_A set [11]

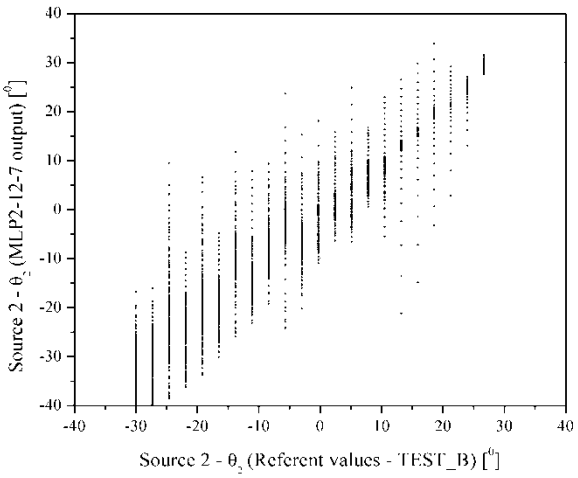
After training MLP neural models on the set TRAINING_B, the results from testing of six models with lowest value for WCE on set TEST_B are shown in Table V [11]. In Figs. 7(a) and 7(b) [11] we may see scattering diagram of the MLP2-22-22 model on the set TEST_B (this model has shown lowest value of WCE on that set). It can be seen that the model MLP2-22-22 has a satisfactory accuracy in angle position determination when we use samples for different partial correlations of sources.

TABLE V
TESTING RESULTS FOR SIX MLP NEURAL MODELS TRAINED ON TRAINING_B SET WITH THE BEST AVERAGE ERRORS STATISTICS [11]

MLP model	TEST_B set	
	WCE [%]	ACE [%]
MLP2-22-22	4.53	0.35
MLP2-23-23	5.58	0.33
MLP2-22-20	5.70	0.36
MLP2-18-16	5.74	0.36
MLP2-18-18	6.33	0.34
MLP2-16-16	8.27	0.35



(a)



(b)

Figure 6: Scattering diagram of MLP2-12-7 model θ_1 output (a), and θ_2 output (b), on TEST_B set [11]

III. STOCHASTIC SOURCE LOCATION BASED ON 2D DOA ANN MODELING

III.A. Correlation matrix calculation for 2D source movement

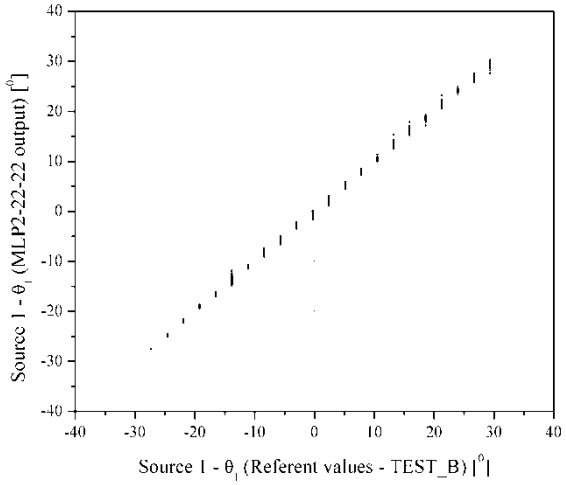
For 2D movement scenarios, Eqs.(2-3) have to be modified [14]:

$$\mathbf{M}(\theta,\varphi) = jz_0 \frac{F(\theta,\varphi)}{2\pi} \cdot \left[\frac{e^{jkr_1^{(1)}}}{r_1^{(1)}} \cdots \frac{e^{jkr_N^{(1)}}}{r_N^{(1)}} \frac{e^{jkr_1^{(2)}}}{r_1^{(2)}} \cdots \frac{e^{jkr_N^{(2)}}}{r_N^{(2)}} \cdots \frac{e^{jkr_1^{(S)}}}{r_1^{(S)}} \cdots \frac{e^{jkr_N^{(S)}}}{r_N^{(S)}} \right] \quad (16)$$

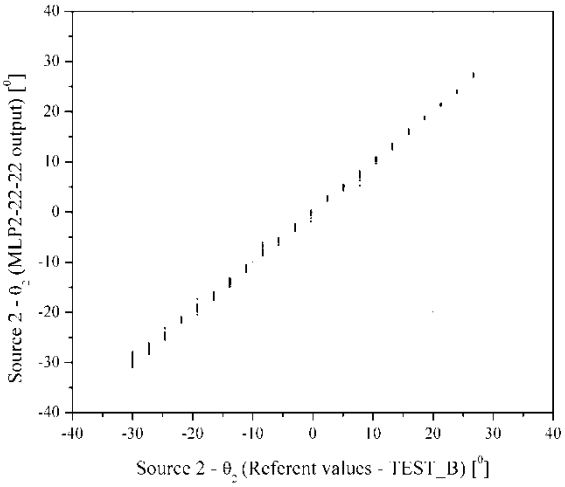
$$E(\theta,\varphi) = \mathbf{M}(\theta,\varphi)\mathbf{I} \quad (17)$$

where $r_i^{(j)}$ is the distance between i-th element in antenna array, representing j-th stochastic source, and the sampling point in far-field. The correlation matrix of signals received

in these sampling points can be obtained from the correlation matrix of antenna elements feed currents as [14]:



(a)



(b)

Figure 7: Scattering diagram of MLP2-22-22 model θ_1 output (a), and θ_2 output (b), on TEST_B set [11]

$$\mathbf{C}_E[i,j] = \mathbf{M}(\theta_i,\varphi_i)\mathbf{c}^I\mathbf{M}(\theta_j,\varphi_j)^H \qquad i=1,\ldots,M \quad j=1,\ldots,M \tag{18}$$

Spatial angles θ and φ in Eqs.(16) and (17) θ describe stochastic source location with respect to the selected sampling point in far-field, $F(\theta,\varphi)$ is the radiation characteristic of the antenna array element and r_s the distance between s-th stochastic EM source and selected sampling point. When a rectangular antenna array of dimensions $M\times P$ is used at the reception, the far-field sampling points correspond to the positions of elements of this array and their total number will be $K=M\cdot P$. In our scenario shown in Fig.8, S stochastic sources are moving independently from each other in a plane parallel to the rectangular antenna array [14]. Axes of dipoles representing stochastic sources are lying in the plane in which sources are moving and they are parallel to the x-axis so that their radiation in the direction of sampling points can be approximated with isotropic radiation. The distance between elements of receiving antenna array along the x-axis is dx,

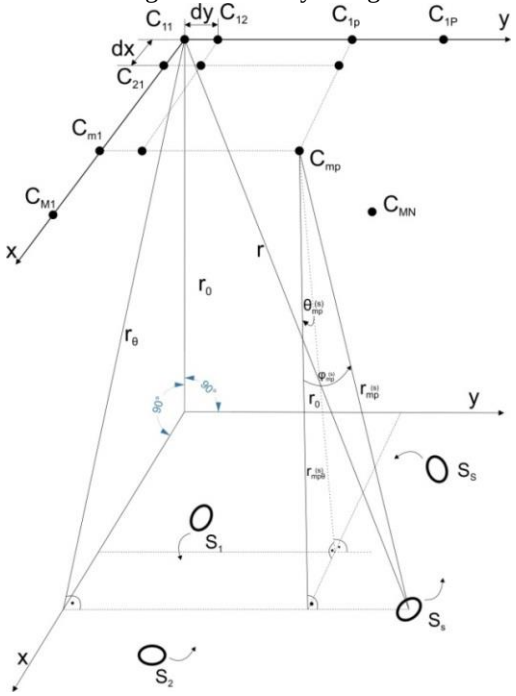


Figure 8: The position of stochastic EM sources in x-y plane with respect to the location of sampling points in the far-field scan area [14].

while the distance between elements along the y-axis is dy. The distance between the plane of planar receiving antenna array and the plane in which stochastic sources are moving is r_0. When the mapping from Eq.(18) is applied separately on each of the sampling points, the appropriate distance between s-th stochastic source and the sampling point at the position of (m,p) sensor of planar receiving array is [14]:

$$r_{mp}^{(s)} = \frac{r_0}{\cos\varphi_{mp}^{(s)}\cos\theta_{mp}^{(s)}} \tag{19}$$

where r_0 represents the distance between the plane of planar receiving antenna array and the plane is which S stochastic sources are moving, while $\theta_{mp}^{(s)}$ and $\varphi_{mp}^{(s)}$ are spatial angles related to the position of s-th source with respect to the (m,p) sensor position and they are [14]:

$$\theta_{mp}^{(s)} = \arctan\left[\tan\theta_{11}^{(s)} - \frac{(m-1)\cdot dx}{r_0}\right] \tag{20}$$

$$\varphi_{mp}^{(s)} = \arctan\left[\tan\varphi_{11}^{(s)} - \frac{(p-1)\cdot dy}{r_0}\right] \tag{21}$$

where $\theta_{11}^{(s)}$ and $\varphi_{11}^{(s)}$ are spatial angles related to the position of s-th source with respect to the referent $(1,1)$ position of antenna array (referent sensor of array). Angles $\theta_{11}^{(s)}$ and $\varphi_{11}^{(s)}$ represent at the same time the angular (θ,φ)

position of *s*-th stochastic source with respect to the planar receiving antenna array so that $\theta_{11}^{(s)} = \theta$ and $\varphi_{11}^{(s)} = \varphi$.

Correlation matrix of signals in sampling points is defined as [14]:

$$\tilde{\mathbf{C}}_E[i,j] = \mathbf{M}(\theta_{\left(\left[\frac{i}{P}\right]+1\right)\left(i-\left[\frac{i}{P}\right]\right)}, \varphi_{\left(\left[\frac{i}{P}\right]+1\right)\left(i-\left[\frac{i}{P}\right]\right)}) \mathbf{c}^I \mathbf{M}(\theta_{\left(\left[\frac{j}{P}\right]+1\right)\left(j-\left[\frac{j}{P}\right]\right)}, \varphi_{\left(\left[\frac{j}{P}\right]+1\right)\left(j-\left[\frac{j}{P}\right]\right)})^H,$$

$$i = 1,\ldots, K \quad j = 1,\ldots, K \quad K = M \cdot P \qquad (22)$$

From Eq.(16) vector **M** is determined, and then according to the angular position of stochastic EM source with respect to the sampling points, the elements of correlation matrix can be calculated using Eq.(18). Matrix elements are normalized with respect to the first matrix element [14]

$$\mathbf{C}_E = \frac{1}{\tilde{C}_{E11}} \cdot \tilde{\mathbf{C}}_E \qquad (23)$$

in order to obtain in our considered scenario the correlation matrix that does not depend on value r_0. For the training of neural network it is sufficient to use only the first row of matrix C_E ([C_{E11}, C_{E12}, … , C_{E1K}]) due to the fact that this row contains sufficient information for determination of angular coordinates of stochastic sources [14].

III.B. Neural model for 2D DoA estimation

The neural model, based on MLP ANN, is trained to extract information about position of *S* stochastic sources in azimuth and elevation planes from the correlation matrix $\mathbf{C}_E$ [14]:

$$[\varphi_1\ \theta_1\ \varphi_2\ \theta_2 \ldots\ \varphi_s\ \theta_s\ \ldots\ \varphi_S\ \theta_S]^T = f(\mathbf{C}_E) \qquad (24)$$

where $[\varphi_1\ \theta_1\ \varphi_2\ \theta_2 \ldots\ \varphi_s\ \theta_s \ldots\ \varphi_S\ \theta_S]^T$ is vector of spatial angles of arrival of the stochastic sources radiation. The architecture of developed neural model is shown in Fig.9, while Eqs.(8-9) from II.C section are also used here.

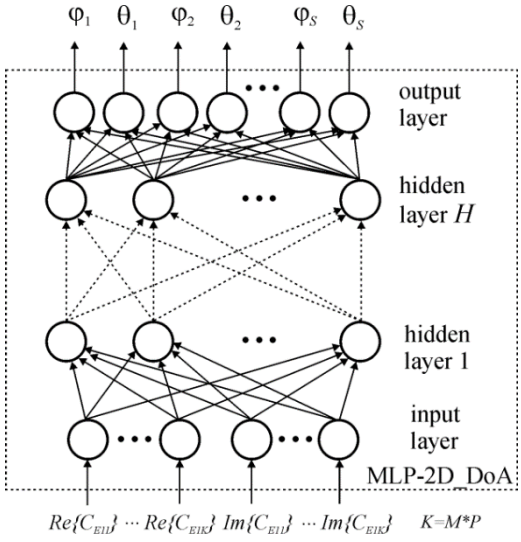


Figure 9: Architecture of MLP neural model for 2D DOA estimation of stochastic EM source signals in *x-y* plane [14]

Again as in II.B section, in order to perform mapping it is sufficient to take only the first column of correlation matrix and therefore $\mathbf{y}_0$ = [Re{$\mathbf{C}_E$[1,1]},…, Re{$\mathbf{C}_E$[1,K]}, Im{$\mathbf{C}_E$[1,1]},…, Im{$\mathbf{C}_E$[1,K]}]. Also, output of the neural network model is given as $[\varphi_1\ \theta_1\ \varphi_2\ \theta_2 \ldots\ \varphi_s\ \theta_s \ldots\ \varphi_S\ \theta_S]^T = \mathbf{w}_{H+1}\mathbf{y}_H$ where $\mathbf{w}_{H+1}$ is a connection weight matrix between neurons of last hidden layer and neurons in output layer. The optimization of weight matrices $\mathbf{w}_1$, $\mathbf{w}_2$,…, $\mathbf{w}_H$, $\mathbf{w}_{H+1}$ and biases values during the training allows ANN to approximate the mapping with the desired accuracy.

III.C. 2D DoA estimation - modelling results for uncorrelated sources

The application of neural model for the 2D DoA estimation of uncorrelated stochastic sources will be illustrated here on the example of one moveable sources (*S*=1). Stochastic source is represented with an array of two elements (*N*=2) with isotropic characteristics and with uncorrelated currents supply so that $\mathbf{c}^I$ is the unit diagonal matrix. Signal is sampled in the far-field in nine points that correspond to the rectangular *M*×*P*=3×3 antenna array (*K*=*M*·*P* =9) [13]. Table VI provides the values of parameters for the the scenarios used to generate samples for training the neural models [13]. Sets of samples for training and testing of the MLP models were generated by using Eqs.(18) and (22). Any combination of angles θ and φ 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 VI
THE VALUES OF PARAMETERS USED IN SAMPLING PROCESS [13]

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 VII
TESTING RESULTS FOR SIX MLP NEURAL MODELS WITH THE BEST AVERAGE ERRORS STATISTICS [13]

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

A set of 14641 samples was created for training applying uniform distribution of θ and φ with 0.5° step in the range [-30° 30°] and with a 0.5° step. Quazi-Newton method with prescribed accuracy of 10^{-4} is used for the training algorithm. Testing set with 7396 samples is providing with uniform distribution of θ and φ angles in the range [-30° 30°] with a 0.7° step. The testing results for six MLP models with the lowest average (ACE) and worst case error (WCE) are shown in Table VII, and MLP2-15-11 is chosen as representative neural model. A close agreement between the output values of neural model and referent θ and φ values can be observed thorough the scattering diagram of testing samples set shown in Figs.10 and 11 [13].

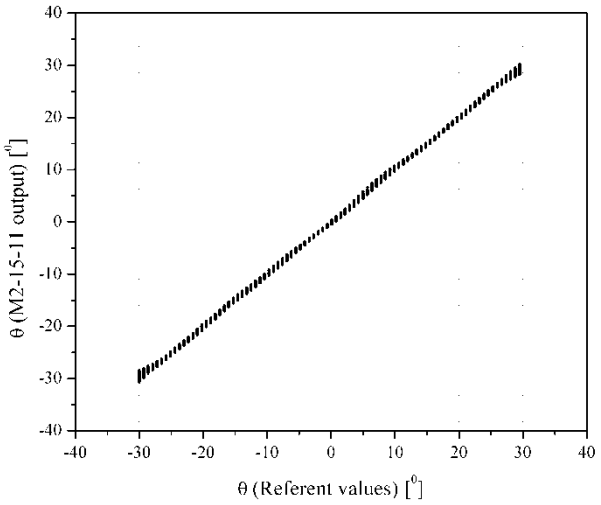


Figure 10: Scattering diagram of MLP2-15-11 model θ output [13]

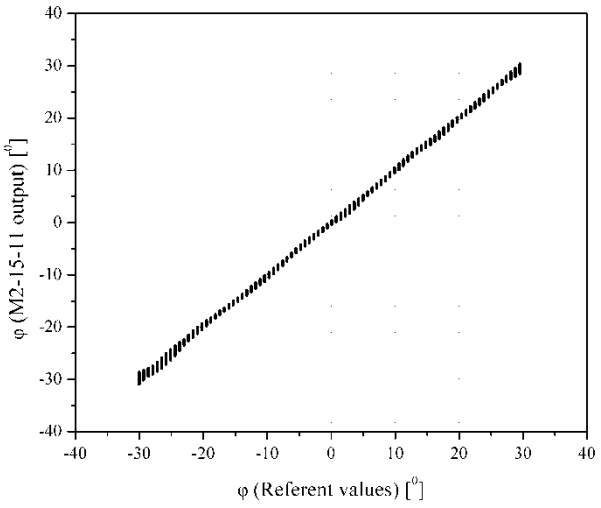


Figure 11: Scattering diagram of MLP2-15-11 model φ output [13]

MLP2-15-11 model was employed to track the movement of the hypothetic stochastic source on earth's surface in a square area in size 800×800 km. The source position was varied along the test trajectory described by the function $y = 3 \cdot 10^{-6} \cdot (x - 10^5)^2 - 3 \cdot 10^5$ where x and y are relative latitude and longitude expressed in meters. Evaluation paths of origin was carried out by sampling in time the correlation matrix of the 69 points shown in Figure 12 [13]. A satisfactory agreement can be observed between the values of the source positions which war estimated by the neuron model and the referent source trajectory.

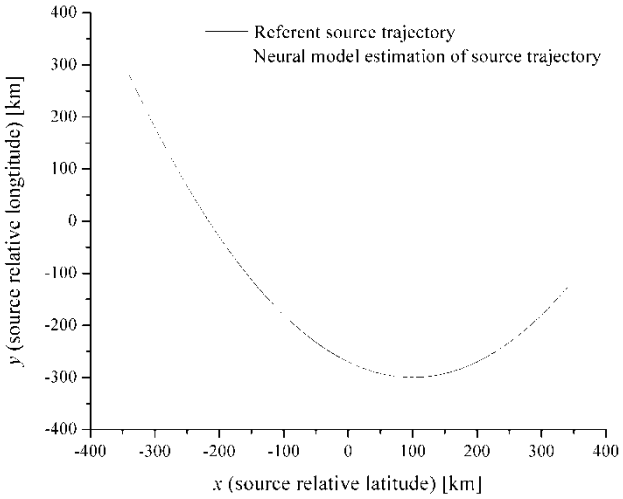


Figure 12: A simulation of localization and tracking the movement of the hypothetic source of stochastic radiation on earth's surface with the MLP2-15-11 model [13]

IV. Conclusion

Overview of using MLP ANNs for the localization of stochastic EM sources was given in the paper. It was illustrated that the develop neural models can be used for efficient and accurate 1D and 2D estimation of electromagnetic radiation originating from either uncorrelated or partially correlated stochastic sources. Future research will be focused to the more general 2D DOA estimation of multiple stochastic sources with different signal level and at arbitrary angular and radial distances from the reception point.

Acknowledgement

This work has been supported by the Ministry for Education, Science and Technological Development of Serbia, projects number TR32052 and TR32054.

References

[1] B. Allen, M. Ghavami, *Adaptive Array Systems: fundamentals and applications*, Wiley, 2005

[2] L. C. Godara, *Smart Antennas*, CRC Press LLC, 2004

[3] R. Schmidt, "Multiple emitter location and signal parameter estimation", *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276-28, 1986.

[4] C. G. Christodoulou, M. Georgiopoulos, *Application of neural networks in electromagnetics*, Artech House, December 2000.

[5] 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.

[6] 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.

[7] 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", *1st IEEE Int. Conf. on Numerical Electromagnetic Modeling and Optimization for RF, Microwave, and Terahertz Applications*, NEMO 2014.

[8] 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.

[9] 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.

[10] 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.

[11] Z. Stanković, N. Doncov, I. Milovanović, M. Sarevska, B. Milovanović, "Neural Model for Far-Field 1D Localization of Mobile Stochastic EM Sources with Partially Correlated Radiation", *Proceedings of the International Scientific Conference on Information Technology and Data Related Research - SINTEZA 2017*, Belgrade, Serbia, pp. 169-175, 2017.

[12] Z. Stanković, N. Doncov, I. Milovanović, B. Milovanović, "1D Localization of Highly Correlated Mobile Stochastic EM Sources using Neural Model", *Proceedings of the 13th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Services - TELSIKS 2017*, Niš, Serbia, October 18-20, pp. 33-37, 2017.

[13] Z. Stankovic, I. Milovanovic, N. Doncov, B. Milovanovic, „2D Localization of Source of Stochastic EM Radiation by using Neural Networks", *LI International Scientific Conference on Information, Communication, and Energy Systems and Technologies – ICEST 2016 - Proceedings of Papers*, pp. 99-102, 2016.

[14] Z. Stanković, N. Doncov, B. Milovanović, I. Milovanović," Efficient 2D Localization of a Number of Mutually Arbitrary Positioned Stochastic EM Sources in Far-Field using Neural Model", accepted paper, *2017 International Conference on Electromagnetics in Advanced Applications* (*ICEAA*), Italy, Verona, September 11 – 15, 2017, pp. 1391-1394. DOI: 10.1109/ICEAA.2017.8065537, ISBN: 978-1-5090-4450-4.

[15] S. Haykin, *Neural Networks*, New York, IEEE, 1994.

[16] Q. J. Zhang, K. C. Gupta, *Neural networks for RF and microwave design*, Artech House, Boston, MA, 2000.

[17] 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.

[18] 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

WORKSHOP ON REASONING-BASED INTELLIGENT SYSTEMS – WRIS'18

A Fuzzy Decision Maker to Determine Optimal Starting Time of Shiftable Loads in the Smart Grids

İsmail H. Altaş[1] and Recep Çakmak[2]

Abstract – **Smart grid studies have been increased tremendously for past ten years *in* order to modernize and solve problems of current electrical grids. One of the aim of the smart grids to react autonomously by means of artificial intelligence and decision maker. Fuzzy logic based embedded control systems simulates human decision and thinking process. So, fuzzy logic and fuzzy decision makers can be utilized in smart grids for automated management and decision facilities. In this paper, a fuzzy decision maker has been proposed to manage of time-shiftable loads in the residences. The proposed fuzzy decision maker determines optimal starting time of time-shiftable loads in the residential area in order to provide balanced power curve and decrease peak load consumptions by scheduling of these loads. Design stage of the proposed fuzzy decision maker have been introduced and presented clearly. Finally, a design example has been given and simulated to show the decision result of the proposed sytem.**

Keywords – ***Fuzzy Logic, Fuzzy Decision Maker, Demand Side Management, Load Scheduling***

I. INTRODUCTION

Electricity generation methods and electricity grid infrastructure has been evolving for last several decades to compensate increased electricity demand and to decrease climate change issues. Renewable energy sources have been utilized more in the electricity generation and energy efficiency has been attracted great attention in the developed countries. Smart grids promise that smart management of the evolved electrical grid to retain sustainability, to avoid collapse and to increase efficiency of the electrical grid. Demand side management (DSM) is the one of the main component of the smart grids to control electricity usage of the consumers to increase efficiency and to reduce greenhouse effects. DSM is also allow to the consumers to make more informed about their electricity consumptions and to collaborate with them to shape power consumption curves of them [1].

Electricity consumption of residences in European countries (EU-25) have been increased by 10.8% during the period of 1999 to 2004 [2]. Residential electricity consumption has share of 29% in total electricity consumption in EU-28 countries by 2014 [3] and in the OECD countries residential sector has share 32% in final electricity consumption by the year of 2015 [4]. There are many devices in the residences such as washing machine and tumble demand high power from the electrical grid. If these appliances engaged at the similar hours of the day, they can occur overload in the electrical distribution system. The overload of the distribution network might decrease the efficiency and might cause instability issues on the distribution network. Therefore, demand side management is significant part of the smart grids [5, 6] and DSM studies can solve these problems mentioned above. In a distribution network, power companies desire to design a decision maker so that the operating times of some loads can be shifted from heavy loading periods of a day to light loading periods of the day.

Many DSM techniques have been proposed in the literature such as price based, incentive based and direct load control based [7]. Heuristic algorithms are applied in demand side management studies to get optimal scheduling according to desired objective functions [8-10]. Direct load control is an important DSM technique and it is proposed in the literature by controlling different loads and considering different objectives [11-14]. Mortaji et al.[11] proposed a smart direct load control by using internet of things (IoT) to control sudden load changes and to decrease peak to average ration (PAR) of the power consumption in the distribution network. They simulated their proposed system by hundred customers and they showed that the peak load can be decrease by 30%. Basnet et al. [12] showed that effect of direct load control by scheduling of HVAC(heating, ventilation and air conditioner). Chen et al. [13] proposed a novel two-layer communication-based distributed direct load control for residential area by allocating control tasks into each home's energy management controller unit, distributively. Martin et al. [14] proposed a decision model to implement direct load control (DLC) on charging of electrical vehicle batteries.

Fuzzy logic (FL) [15] is described by Zadeh as computing technique with words are used in place of numbers for computing and reasoning. FL is utilized on numerous studies about power systems in the literature [16-18]. The main advantage of the FL is to modify and design the system by linguistic variables which are utilized in human decisions. FL based DSM studies are also studied and proposed in the literature such that examples [19-23]. Chandran et al.[19] compared fuzzy logic control(FLC) based demand response with classical Boolean logic based direct load control. They showed that advantages of FLC based DLC over classical DLC and they used forecasted load curves instead of price curve to avoid distributing the customer comfort level. However, they utilize fuzzy logic approach as a controller unit based on customer priorities. Keshtkar et al. [20] proposed a FL based autonomous system to manage residential Heating, Ventilation, and Air Conditioning (HVAC) units by considering outdoor temperature, electricity price, occupant presence and electricity demand of the house.

[1]İsmail H. Altaş is with the Faculty of Engineering at Karadeniz Technical University, 61080 Trabzon, Turkey,
E-mail: ihaltas@ktu.edu.tr

[2]Recep Çakmak is with the Faculty of Engineering and Natural Sciences at Gümüşhane University, Gümüşhane, Turkey,
E-mail: rcakmak@gumushane.edu.tr

A fuzzy logic based domestic load management which use comfort, cost and demand response parameters has been proposed by Rahman et al.[22]. However, their proposed system does not consider the consumptions of other residences, it operates for the just one residence by considering the price signal. So, it can be get financial benefits for the one residences but if the all residences act same way considering the same price signal it might be lead adverse results instead of desired demand response effects.

In this paper a fuzzy decision maker (FDM) has been proposed such that determine optimal starting time when a demand is entered to it. The proposed systems assumed that the customer will get financial benefits is they allowed the control of their time-shiftable loads and it operates direct load control approach. Main objective of the proposed system is considering the residential area load curve to inhibit high power consumptions by utilizing the demand power and forecasted load powers. To give an example, if the predicted load in a time ahead is lower than the current load power, then the active demands should be postponed to be operated during these lower demand period hours of the day. The proposed FDM based demand scheduling system has been generated on MATLAB by modifying the generalized fuzzy logic controller design study [24]. So, it can be easily modified and/or improved for the future studies.

The rest of the paper is organized as follows: Section II introduces designing of the proposed fuzzy decision maker. Section III includes case study and results. Section IV. presents the conclusion.

II. Designing of a Fuzzy Decision Maker to Determine Optimal Starting Time of the Shiftable Loads

The proposed FDM processes demand power and the difference between demand power and predicted future power as two inputs. The overall demand management mechanism of the proposed system is illustrated in Fig. 1 and explained below.

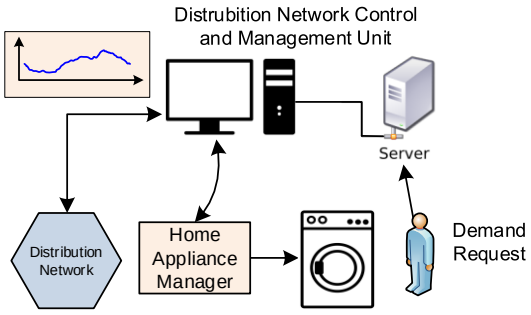


Fig. 1. Overall demand management mechanism of the proposed system

Overall process steps of the proposed demand management mechanism:

1. Customer send the demand request to the server of the distribution network control and management unit(DNC&MU).
2. DNC&MU which includes the proposed FDM processes the forecasted demand loads and the demand request based on the pre-determined fuzzy rules.
3. The FDM generate an optimal operating time of the load which is requested by the customer.
4. Then the optimal operating time information sends to the home appliance manager to execute the operation of the requested load when the time is equal to the optimal operating time.

The flowchart of the proposed system is presented in Fig. 2. It is searched that is there any demand request in every control interval loop the system. If there is a demand, the system gets the demand power and evaluates it on FDM to determine optimal starting time for it. After accomplishing the above process current and future load data is updated. The customer notified by the system about the delay time of him scheduled load and the system operate the shifted load automatically at the scheduled time.

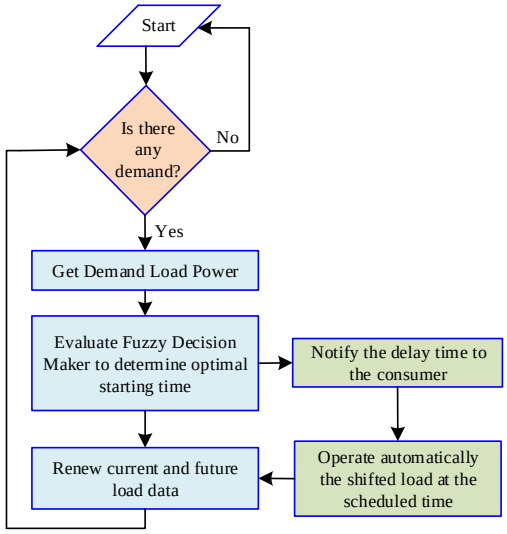


Fig. 2. The flowchart of the proposed system

Demand power may be categorized as low (LW), average (AV) and high (HI). Since the difference may be positive, negative and zero depending on demand and predicted power values, the universe of power difference can be portioned into five fuzzy subsets as negative high (NH), negative low (NL), zero (ZE), positive low (PL) and positive high (PH). The output space representing the delay time is divided into three fuzzy subsets as short (SH), medium (ME) and long (LN). A block diagram of the FDM is given in Fig.3. The FDM block includes the fuzzy rule table given by Table 1, which includes 15 fuzzy rules for decision-making.

The FDM given in Fig.3 uses Mamdani fuzzy reasoning algorithm with two crisp inputs and one crisp output. The current power demand (P_D) is represented by variable x and the difference (ΔP) between current load power and predicted future depended power is represented by variable y. The output is the delay time t_D represented by variable z.

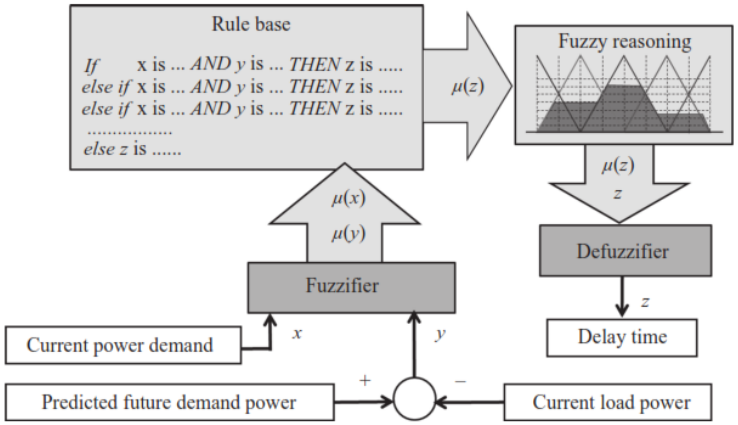


Fig. 3. Block diagram of the proposed fuzzy decision maker

If the current demand power is LW and the difference power ΔP is NH, then this load demand can be delayed for LN because negative ΔP means that the future power demand will be smaller than the current power dissipation so that the operating time of this power demand can be shifted ahead with a long delay. If current power demand is HI and difference power ΔP is PH, then this load demand should not be delayed because the predicted future power demand will already be high. If this high-power demand is postponed to a later time, the load demand will be much higher such that the generated power might be insufficient as the network faces overloading operating conditions, which cause overheating and higher losses. The other fuzzy rules are generated with a similar point of view to represent the experience of a power network operator.

TABLE I
RULE TABLE OF THE PROPOSED FDM

	ΔP				
P_D	**NH**	**NL**	**ZE**	**PL**	**PH**
LW	LN	ME	SH	ZE	ZE
AV	LN	ME	SH	ZE	ZE
HI	LN	SH	ZE	ZE	ZE

Let us assume that the current power demand is 2300 W and the difference power ΔP is negative 400 W and try to find the decision about how long the load demand can be delayed. The fuzzy decision process for this example is depicted in Fig. 4. Defuzzification process is depicted in Fig. 5.

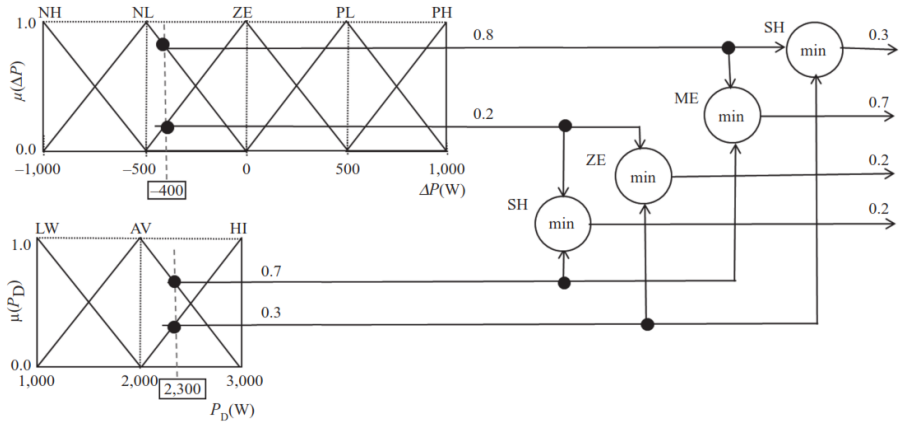


Fig. 4. Example of fuzzy decision maker process.

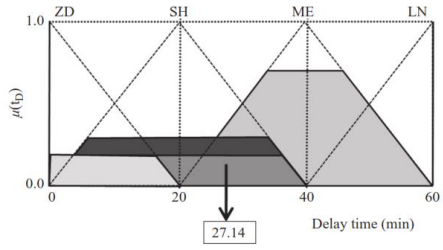


Fig. 5. Defuzzification of the fuzzy decision for the example

III. CASE STUDY AND RESULTS

In this chapter, a case study has been simulated in order to show operation of the proposed system. Three different demand requests (DR1, DR2 and DR3) have been applied and the response of the proposed system has been observed. The results with and without the proposed FDM based scheduling have been compared.

It is assumed that predicted daily load curve for twenty customers is given for 30 min. time intervals for 24 hours as Fig. 6. It is also assumed that there are three demand requests at the specified times as shown in the figure. Let say that the average of demand powers is 1500W and duration time on operation of them 90 minutes.

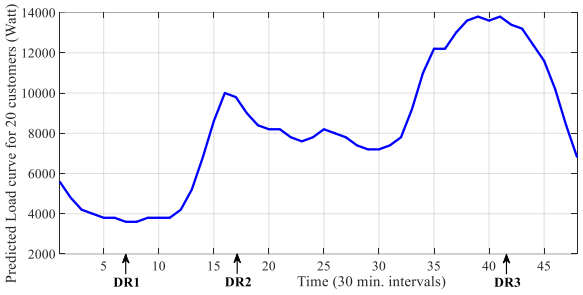


Fig. 6. Predicted daily load curve for twenty customers and demand request times in the test case.

According to the above predicted daily load curve membership functions of the ΔP have been designed as triangle fuzzy membership functions as shown in Fig. 7.

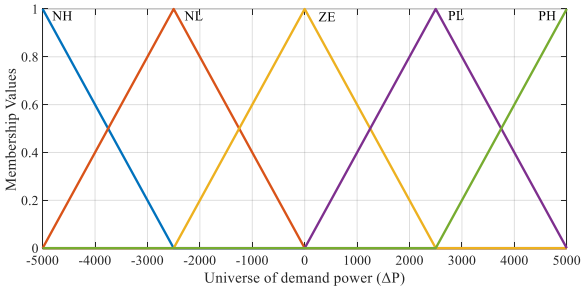


Fig. 7. Designed membership functions for ΔP.

It is assumed that the means of demand power are change between 1000 W and 3000 W. So, membership functions of the demand power have been created as triangle membership functions as depicted in Fig. 8.

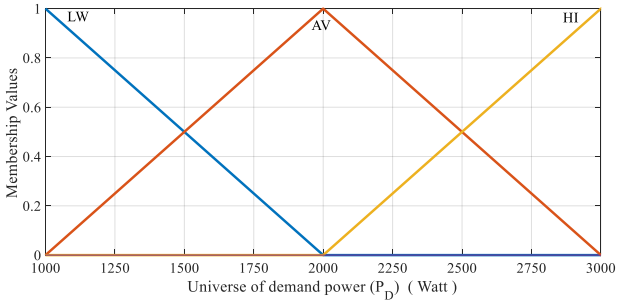


Fig. 8. Designed membership functions for demand power(PD).

The range of the delay time (t_D) which will be determined by the proposed FDM is designed between 0 and 90 minutes then fuzzy membership functions of it has been created as illustrated in Fig 9.

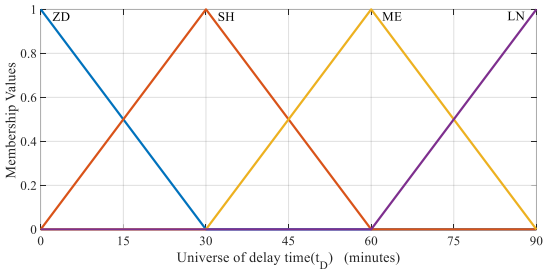


Fig. 9. Designed membership functions for delay time (t_D).

In the simulation Case 1, it is assumed that there is no DSM action for the customers and the customers operate their time-shiftable loads at the specified times which is shown in Fig.6. In the simulation Case 2, it is assumed that the proposed FDM is active and schedule the demand requests based on the rules on Table I. According to the both Case 1 and Case 2 simulations, the aggregated power curves are obtained as in Fig. 10.

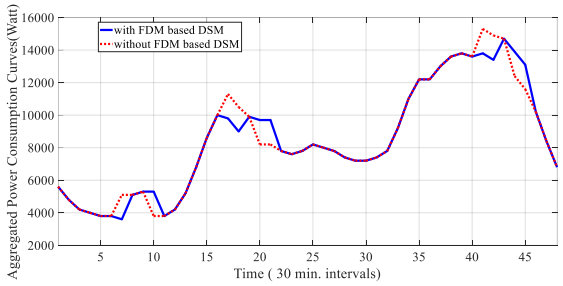


Fig. 10. Designed membership functions for delay time (t_D).

Comparison of the results are presented in Table II. As shown is clearly, PAR and peak power can be reduced the FDM even with three time-shiftable loads. If the participation of the customers to the DSM, the prediction and scheduling will be improved, of course.

TABLE II
COMPARISON OF THE RESULTS

	Without DSM	With Proposed FDM based DSM
Average power (W)	8527.1	8527.1
Peak power (W)	15300	14700
Peak to Average Ratio (PAR)	1.7943	1.7239

IV. CONCLUSION

Demand side management is one of the key components of the smart grids. Distribution networks in residential regions can be faced high power demand at some periods of a day because of high power requested devices of the residences. The drawn high-power causes to engage extra generation units and overloaded the distribution network. However, some high-power demand household devices can be shift to another

time periods. DSM studies want to shift these shiftable appliances to lower demand regions of a day by financial incentives. Instead of manual or human controlled management of these appliances fuzzy logic can be good alternative to create automated DSM system which act like human decisions.

In this study a fuzzy decision maker has been developed to determine optimal starting times of time shiftable loads in the residences. The proposed system act as dynamically and can be modified according to the objectives by changing of the fuzzy rules. Bu the case study a mini simulation has been studied to show effects and response of the proposed FDM based dynamic scheduling unit. The results show that the benefits of the scheduling and decisions of the proposed system. This proposed FDM will be good reference for evolving smart grids studies in the future. Of course, the proposed FDM based scheduling system will be improved in the next studies by comprehensive approaches.

ACKNOWLEDGEMENT

This work was supported by the Scientific and Technological Research Council of Turkey (TÜBİTAK), Grant No: EEEAG-115E943. Recep Çakmak thanks to TÜBITAK for funding PhD scholarship supplied by Science Fellowships and Grant Programs Department (BİDEB).

REFERENCES

[1] B. Davito, H. Tai, and R. Uhlaner, "The smart grid and the promise of demand-side management," *McKinsey on Smart Grid*, vol. 3, pp. 8-44, 2010.

[2] P. Bertoldi and B. Atanasiu, "Electricity consumption and efficiency trends in the enlarged European Union," *IES–JRC. European Union*, 2007.

[3] P. Bertoldi, J. López-Lorente, and N. Labanca, "Energy Consumption and Energy Efficiency Trends in the EU-28 2000–2014," in *Scientific Reports of the Joint Research Centre, European Commission, JRC101177*, 2016.

[4] I. E. Agency, "Electricity information 2017," *International Energy Agency*, 2012.

[5] H. Farhangi, "The path of the smart grid," *IEEE Power and Energy Magazine*, vol. 8, no. 1, pp. 18-28, 2010.

[6] P. Palensky and D. Dietrich, "Demand Side Management: Demand Response, Intelligent Energy Systems, and Smart Loads," *IEEE Transactions on Industrial Informatics*, vol. 7, no. 3, pp. 381-388, 2011.

[7] J. S. Vardakas, N. Zorba, and C. V. Verikoukis, "A Survey on Demand Response Programs in Smart Grids: Pricing Methods and Optimization Algorithms," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 152-178, 2015.

[8] R. Cakmak and I. H. Altas, "Scheduling of domestic shiftable loads via Cuckoo search optimization algorithm," in *2016 4th International Istanbul Smart Grid Congress and Fair (ICSG)*, 2016, pp. 1-4.

[9] R. Çakmak and İ. H. Altaş, "Optimal scheduling of time shiftable loads in a task scheduling based demand response program by symbiotic organisms search algorithm," in *2017 Saudi Arabia Smart Grid (SASG)*, 2017, pp. 1-7.

[10] T. Logenthiran, D. Srinivasan, and T. Z. Shun, "Demand Side Management in Smart Grid Using Heuristic Optimization," *IEEE Transactions on Smart Grid*, vol. 3, no. 3, pp. 1244-1252, 2012.

[11] H. Mortaji, S. H. Ow, M. Moghavvemi, and H. A. F. Almurib, "Load Shedding and Smart-Direct Load Control Using Internet of Things in Smart Grid Demand Response Management," *IEEE Transactions on Industry Applications*, vol. 53, no. 6, pp. 5155-5163, 2017.

[12] S. M. S. Basnet, W. Jewell, and H. Aburub, "Effects of Direct Load Control and Residential PV System on Demand Response and Its Cost Benefit Analysis," in *2015 Seventh Annual IEEE Green Technologies Conference*, 2015, pp. 117-124.

[13] C. Chen, J. Wang, and S. Kishore, "A Distributed Direct Load Control Approach for Large-Scale Residential Demand Response," *IEEE Transactions on Power Systems*, vol. 29, no. 5, pp. 2219-2228, 2014.

[14] P. Sanchez-Martin, G. Sanchez, and G. Morales-Espana, "Direct Load Control Decision Model for Aggregated EV Charging Points," *IEEE Transactions on Power Systems*, vol. 27, no. 3, pp. 1577-1584, 2012.

[15] L. A. Zadeh, "Fuzzy logic= computing with words," *IEEE transactions on fuzzy systems*, vol. 4, no. 2, pp. 103-111, 1996.

[16] R. Cakmak and I. H. Altas, "The effect of integration types on FLC based MPPT systems," in *2013 IEEE INISTA*, 2013, pp. 1-4.

[17] I. H. Altas and A. M. Sharaf, "A novel maximum power fuzzy logic controller for photovoltaic solar energy systems," *Renewable Energy*, vol. 33, no. 3, pp. 388-399, 2008/03/01/ 2008.

[18] I. H. Altas and J. Neyens, "A fuzzy logic decision maker and controller for reducing load frequency oscillations in multi-area power systems," in *2006 IEEE Power Engineering Society General Meeting*, 2006, p. 9 pp.

[19] C. V. Chandran, M. Basu, and K. Sunderland, "Comparative study between direct load control and fuzzy logic control based demand response," in *2016 51st International Universities Power Engineering Conference (UPEC)*, 2016, pp. 1-6.

[20] A. Keshtkar, S. Arzanpour, and F. Keshtkar, "An autonomous system via fuzzy logic for residential peak load management in smart grids," in *2015 North American Power Symposium (NAPS)*, 2015, pp. 1-6.

[21] Z. Wu, S. Zhou, J. Li, and X. P. Zhang, "Real-Time Scheduling of Residential Appliances via Conditional Risk-at-Value," *IEEE Transactions on Smart Grid*, vol. 5, no. 3, pp. 1282-1291, 2014.

[22] M. M. Rahman, S. Hettiwatte, and S. Gyamfi, "An intelligent approach of achieving demand response by fuzzy logic based domestic load management," in *2014 Australasian Universities Power Engineering Conference (AUPEC)*, 2014, pp. 1-6.

[23] E. Hamid, P. Nallagownden, N. B. M. Nor, and M. A. L. Muthuvalu, "Intelligent demand side management technique for industrial consumer," in *2014 5th International Conference on Intelligent and Advanced Systems (ICIAS)*, 2014, pp. 1-6.

[24] I. H. Altas and A. M. Sharaf, "A generalized direct approach for designing fuzzy logic controllers in Matlab/Simulink GUI environment," *International Journal of Information Technology and Intelligent Computing, Int. J. IT&IC*, vol. 1, no. 4, pp. 1-27, 2007.

A System of Choice For Pre-Six Sigma Projects with the Use of the Paraconsistent Logic

Giovanna Albertini[1], Jair M. Abe[2], Caique Z. Kirilo[3], Kazumi Nakamatsu[4], Luiz A. de Lima[5]

Abstract –**Companies of products and services, implement the six sigma methodology in several scenarios, however, without taking into account factors (organizational climate, organizational culture) that are fundamental to the success of the implementation in the pre-project phase and even in the selection of its six sigma projects. The purpose of this study is the development of a system that uses the Paraconsistent Decision Method to study the feasibility of its implementation in Six Sigma in a given scenario, making the decision making more precise. The Paraconsistent Decision Method allows the support of paraconsistent logic in the (pre-project) phase of choice in consideration of six sigma projects; we have the quest to enable improvement in success accuracy in scenarios where there are factors (organizational climate, organizational culture) critical of success. This article aims to contribute to the constant search for quality (reduction of defects) and mitigation of costs by companies in low-quality scenarios (defects in products and services).**

Keywords –**Six Sigma; Quality; Paraconsistent Annotated Evidential Logic Eτ; Paraconsistent Decision-Making Method.**

I. INTRODUCTION

According to Mikel Harry, he recognizes as a six-sigma methodology process improvement that achieves defect levels of 3.4 ppm (parts per million) for critical quality characteristics of customers. Deming in 1990, in his vision of states, reinforces that in every process there is some variation, in greater or lesser quantity; the key to improving processes is to attack and reduce the cause of variation systematically. From the tools applied logically and structured in a scenario that has the essential for the operation and proper performance of the system, in this scenario, a scene with an excellent organizational climate and an ethical corporate culture, preventing human factors can affect system performance. [2]

The question is related to the fact that the organizational culture and organizational climate can be considered as unstable and ephemeral since both are mostly human and suffer constant changes which can affect the behavior of the system.

Noticeably, or not, the most significant difficulty in the deployment of Six Sigma is in exercising our knowledge and their tools, where the system depends on both the team collaboration and the environment, as well.

There are fundamentally human interactions, where these interactions may not suffer from the human inconsistencies or attitudes as vitiate the data obtained through the Six Sigma projects. [6].

Problems and inconsistencies occur naturally in the scenario with humans, not impeding the ability of reasoning or human thought, the system can perform its knowledge of the situation, together with the humans correctly when finding themselves in a scenario that meets their needs entirely. [6]

Given this assumption, we have sought to establish the feasibility of implementing the Six Sigma system, considering the critical success factors, the organizational climate, the organizational culture and the scenario. [6]

Considering that the decision-making has always been a painful process for both the machine and the human, the vast amount of data, possibility and possible results made this task a problem that needed something new to resolve; it needed a system capable of accurately calculate and show the possible scenarios, a method to support decision-making. [6].

However, in addition to a support system for the decision, a precise system, capable of calculating all the inconsistencies of the scenario, working with a calculation which includes all the variables and brings.

As a result, the feasibility of the System of choice for pre-Six Sigma projects with the use of Paraconsistent Logic becomes patent, since such logic has the ability to process uncertain, inconsistent and even incomplete data in a non-trivial way.

Hence we have chosen the said logical system as the logic underlying our studies.

[1]Giovanna Albertini is Student at Undergraduate Program in Analysis and Development of Systems at Paulista University, Rua Amazonas da Silva 737, São Paulo - SP, Brazil, E-mail: anne.albertini@live.com

[2]Caique ZanetiKirilo is Assistant at Software Engineering Research Group at Paulista University, Rua Antonio de Macedo505, São Paulo-SP, Brazil, E-mail: caiquez.kirilo@hotmail.com

[3]Jair Minoro Abe is Full Professor at Graduate Program in Production Engineering at Paulista University, Rua Dr. Bacelar 1212, São Paulo - SP, Brazil, E-mail: jairabe@uol.com.br

[4]Kazumi Nakamatsu is Full Professor at Human Science and Environment/H.S.E at University of Hyogo, Hyogo, Japan, E-mail: nakamatu@shse.u-hyogo.ac.jp

[5]Luiz A.de Lima is Student at Graduate Program in Production Engineering at Paulista University, Rua Dr. Bacelar1212, São Paulo - SP, Brazil, E-mail: aula.prof.luiz@gmail.com

II. THEORETICAL BACKGROUND

A. Six Sigma

In the mid 80's, Six Sigma was born in the company Motorola. Directly and indirectly, the company, at that time, was spending around 10% and 20% of revenues in low quality. After studying the scenario, the bond between the experience of apparent failure on clients and, also, the knowledge of internal defects in their factories, Motorola started to be aware of the fact that the low quality obtained a significant impact on its profitability of primary line. [2]

Soon after its deployment at Motorola, Six Sigma has different settings that in short, were linked to efficiency in processes and operations, the improvement of business processes, achieving excellence in our processes. [3]

However, the primary objective of Six Sigma continues to lead the continuous improvement of the process of troubleshooting methodology, being documented and verifiable repetition. [3]

Another definition that can be attributed to this system, which is the definition of a management philosophy, which seeks to achieve challenging objectives considered, reduction of defects in products, using processes and services, through a careful analysis of the results obtained and data collection. [1]

The level of the Six Sigma identification is taken into account as main inputs: total opportunities (number of units tested * possible quantities of opportunities) along with the number of defects found. In a given hypothesis (errors found in production) as shown in table 1, we considered the total of opportunities = 1; then we had the perception of how impotent means the search for the 6sigma level, which represents the almost total extinction of defects, and consequently to the almost 100% success.

TABLE1. PROJECT SIX SIGMA WITH TOTAL OPPORTUNITIES = 1 AND FORMULA.

Sigma Level	DPMO-Defects per Million Opportunities(DPO x 1.000.000)	% Error - Six Sigma	% No Error - Six Sigma
6	3,4	0,00034%	99,99966%
5	233	0,02330%	99,97670%
4	6210	0,62100%	99,37900%
3	66807	6,68070%	93,31930%
2	308538	30,85380%	69,14620%
1	691492	69,14620%	30,85380%

(Source: Author)

In the Six Sigma system is used the tool DMAIC (Define, Measure, Analyse, Improve and Control)

Defines: an accurate definition of the scope of the project;

Measure: Find the focus of one or more problems in the scenario;

Analyze Definition of the causes of each problem;

Improve: Evaluate, present and calculate possible solutions to questions;

Control: Ensure that the answer will keep for a long-term goal. [6].

The logical way to use the DMAIC tool, follow the steps as shown in figure:

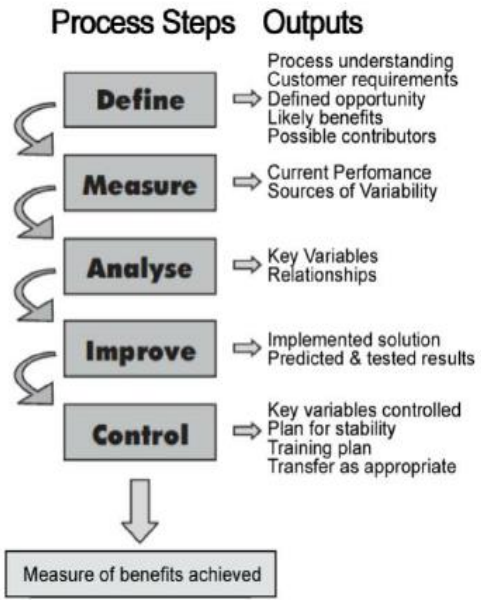


Fig. 1 – *DMAIC – Source:* [10]

B. Success Factors of Six Sigma

It can be identified as factors affecting the system: assigned projects and the environment in which it is being implemented, team preparation and top management, lack of structure and necessary knowledge to work with the system, lack of leadership and team monitoring. Add to that the internal processes of the company. All this leads to the prevention of the achievement of objectives and improvement in the operations and products of the company. [2]

The leadership can be singled out as essential and indispensable for achieving the success of Six Sigma. Monitoring progress and ensuring team commitment is monitored through meetings. Such commitment constitutes one of the fundamental tasks that an active leadership and senior management need to realize. [2]

In addition to the performance of the high administration, customer focus, the use of a structured method and the proper infrastructure are considered the factors of success of Six Sigma. [2]

C. Organizational Climate

The organizational climate can be roughly defined as the work environment, the corporate environment, and psychological atmosphere. Within this environment, it is easier to detect the effects of climate change on people, affecting mainly the performance and teamwork, both significantly essential pillars for the performance of the system Six Sigma, which detect for what reasons the environment is this way.[4]

Even when, understood that the organizational climate is fundamental of inconsistencies and unforeseen changes, makes it essential for the study and the importance of balance in the environment that the system works mainly with human interactions and develops its methodology in the team. [4]

It makes the current mood is the motivation of the members which, as a result, make the environment more productive and satisfying, generating positive effects and animation, collaboration and interest.

Changes happen all the time, preventing the balance still and stable. However, control the variation and seek that doesn't happen an extreme contrast, making the climate with foci of disinterest, depression, dissatisfaction, in more severe cases, which may lead to strikes, nonconformism, unrest among the members of the scenario that consequently also become dissatisfied with the company. [8]

The organizational climate must be studied and thoroughly analyzed by the administrator, then, toil to encourage their decisions, and then find it necessary, interfering in the environment to generating positive changes and gradual climate and organizational culture. [9]

D. Organizational culture

Speak of regulatory climate makes consequent need to speak of corporate culture since one refers to the other. [5]

Organizational culture is what influences and defines the regulatory environment. Would the reasons by which, the atmosphere is the climate in which is, he is a particular climate or not, the study of the culture, is the study of attitudes, habits, gestures, speech, among many others, that establish the environment and team collaboration among themselves. [5]

After setting a set of norms, values, and beliefs that guide and normalize the behavior of particular team, becomes noticeable that culture is broader than the organizational climate. The importance of organizational culture is the significant influence that it has on the environment and people. [9]

If the environment is detrimental to the team and the processes, changes must also come from the culture, essential points for a motivational change are communication, competence, commitment, continuity, and understanding. [9]

E. Paraconsistent Logic

More commonly known as non-classic logic, the paraconsistent logic goes beyond classical logic with calculations and results, which defy the concepts defined by classical logic. Had as precursors, the Polish logician Jan Łukasiewicz (1878-1956) and the Russian logician Nikolai A. Vasiliev (1880-1940) in 1910, published, regardless, jobs where they sought to deal with the possibility of a logic which does not exclude ab initio the contradictions. However, only in 1948, the Brazilian logician Newton C. A. da Costa and the Polish logician StanisławJaśkowski, achieve, independently, build the Paraconsistent Logic. [7]

The project will be used to Paraconsistent Logic Noted Evidential - Eτ, in search of proofs and technical procedures, paraconsistent logic works, in the form of degrees of evidence or belief, with propositions accompanied by annotations. [7]

Paraconsistent logic is a fundamental instrument in various areas and subjects due to its capacity of dealing with uncertainties, inconsistencies, and paracompleteness. [7]

The results obtained through formulas and calculations are positioned on the grid of the paraconsistent logic as shown in Fig. 2.

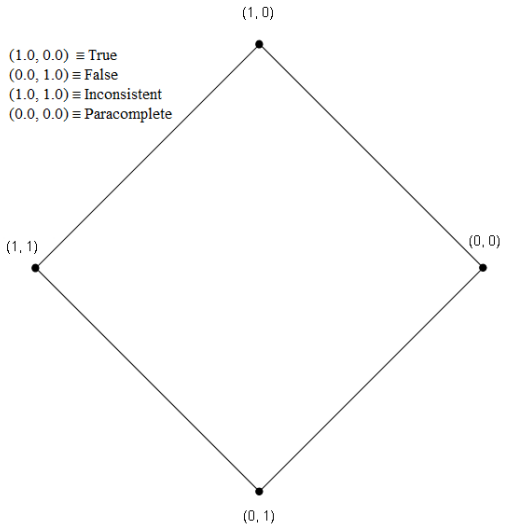


Fig. 2 - *Paraconsistent Logical Lattice- Source:* [11]

F. Paraconsistent Method of Decision

The Paraconsistent Method of Decision (MPD) was developed by Carvalho (2006) through their studies. To recognize the factors that influence in the enterprise, causing the success or failure, in other words, what can influence the decision of continuity of particular project or not. [7]

It was possible to recognize that specific factors may present different results, as favorable conditions, in other cases, unfavorable terms, or else, can still submit circumstances indifferent to the project. [7]

TABLE II
EXTREME AND NON-EXTREME STATES

Extreme States	**Symbol**
True	V
False	F
Inconsistent	T
Paracomplete	⊥
Non-extreme states	**Symbol**
Quasi-true tending to Inconsistent	QV→T
Quasi-true tending to Paracomplete	QV→⊥
Quasi-false tending to Inconsistent	QF→T
Quasi-false tending to Paracomplete	QF→⊥
Quasi-inconsistent tending to True	QT→V
Quasi-inconsistent tending to False	QT→F
Quasi-paracomplete tending to True	Q⊥→V
Quasi-paracomplete tending to False	Q⊥→F

The MPD receives data from the members of the decision-making process, as the experience, uses the so-called "experts" for evaluation, making them essential tools in the assessment of a specific issue. Moreover, through the information obtained, performs the calculation considering all the possibilities, not only of the members, as well as the scenario and the company. [7]

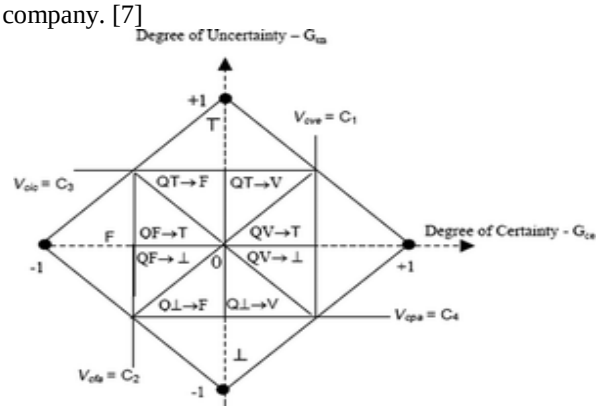


Fig.3.Extreme and non-extreme States. Source[13]

III. The Project

This study proposes the development of software that can calculate the feasibility of pre-projects of Six Sigma system through the use of the Paraconsistent Method of Decision, aiding in the decision-making process. By using the Paraconsistent Method of Decision, a questionnaire is considered to collect the necessary data on the project.

The user will define the experts who will provide the information on the project and the importance (weights) of each expert, making the report of a particular expert more relevant, in comparison with the other information from other experts. Once completed the questionnaire, it will be done the calculations with the evidence degrees, and it will be delivered the result of viability to the user.

Whereas it is necessary to calculate many variables, the software will be responsible for providing more accurate information essential for the decision-making process. Obviously precision and accuracy of the results are paramount in this process, and of utmost importance for the scenario.

To reflect the joint influence of all factors with weight in each decision, one must take into account the Global Analysis and are collected by the favorable and contrary evidence degree.

The calculation of the Global Analysis can be extracted by the weighted average of the evidence of conviction and uncertainties resulting from all the factors. When the weights in each decision are equal, the Global Analysis should be calculated by the arithmetic mean of the evidence of belief and uncertainty, becoming the geometric center.

At this point, the study advances and reinforces the importance of data collection by forms filled by experts to the implementation of algorithms represented in flowcharts in a way to implement in any computational technology and that support the decision support by the proposed system. The decision-making process consists of choosing one of several alternatives. The unified process of annotated paraconsistent logic is proposed as an aid in the decision-making of recounting, as follows:

TABLE III. UNIFIED MACRO PROCESS PARACONSISTENT ANNOTATED LOGIC

Item	Process	SubProcess
A	Definition	Define Proposition; Define Factors; Define Section; Define Database;
B	Transformation	Generate Normalization; Use Evidence (favorable and unfavorable);
C	Calculation	Calculate Maximization; Calculate Minimization; Calculate Evidence (Resultant Min, Resultant Max); Calculate Degree (Gce: Certainty, Gco: Contradiction); Calculate Globals Analysis (Gce: Certainty, Gco: Contradiction);
D	Parameterization	Parametrize Limitvalues;
E	Processing	Process Para-Analyzeralgorithm;
F	Decision-making Support	Assists decision-making;

The use of Paraconsistent Logic Annotated as support in decision-making in implementing six sigma projects should fill a significant gap in the demands for products and services that are based on the six sigma methodology. In this new proposed form, factors of climate and/or culture should be taken into account in the implementation of the six sigma by managers who decide success.

IV. Discussion Of Results

The study for the development of software capable of bringing the Paraconsistent Method of Decision to calculate the inconsistencies of the scenario and the people who are part, brought more reliability and accuracy to the decision-making process, giving due importance to the calculations and the results obtained.

The study necessary for the development was about the whole process from the pre-project the decision of deployment of the system Six Sigma. The approach by the proposed system must be based on the form that meets propositions able to foment data in the possibility to allow the use of paraconsistent logic and to obtain results that will aid in the whole of decision making by six sigma projects.

Other ways of representing the paraconsistent logic with possible implementation in a particular programming

language are to launch the use of the flowchart, where we have:

In this stage of the flowchart, there is an excellent possibility of being quasi-true tending to the inconsistent, or inconsistent tending to the True,

because the Gce and Gco conditions result in some response and when there is no possibility to answer, it follows in the "Y" flow to explore the possible answers offered by the paraconsistent logic

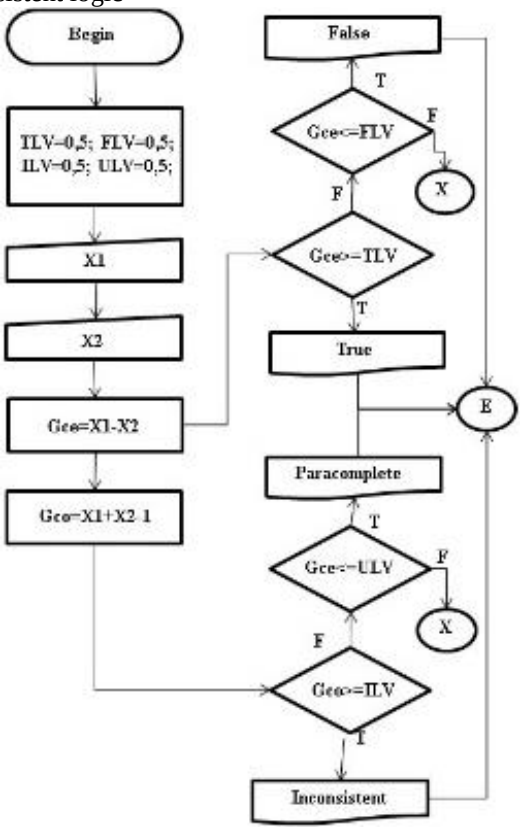


Fig. 4. Paraconsistent logical flowchart: True, False, Inconsistent, Paracomplete

The flowchart (Fig. 4) shows that there is a possibility of being quasi-true tending to the inconsistent, or quasi-inconsistent tending to the True because the Gce and Gco degrees conditions result in some response. When there is no possibility to answer, it follows in the "Y" flow to explore the possible answers offered by the structure of paraconsistent logic.

The flowchart (Fig. 5) shows that there is a possibility of being quasi-true tending to the inconsistent, or inconsistent tending to the True because the Gce and Gco conditions result in some response. When there is no possibility to answer, it follows in the "Y" flow to explore the possible answers offered by the paraconsistent logic.

The next flowchart (Fig. 6), there is a possibility of being quasi-true tending to the Paracompleteness or Paracompleteness tending to the True, since the Gce and Gco conditions result in some response. Moreover, when there is no possibility to answer, it follows in the stream "Z" to explore the possible answers offered by the paraconsistent logic.

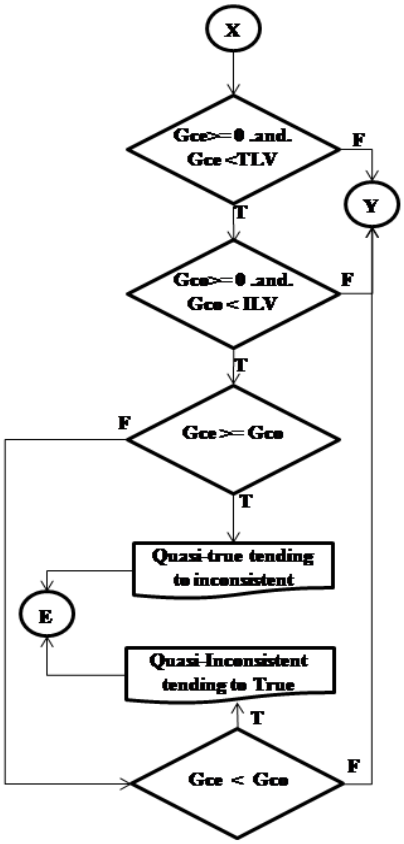


Fig. 5. Paraconsistent logical flowchart: Quasi True tending to the Inconsistent, Inconsistent tending to True

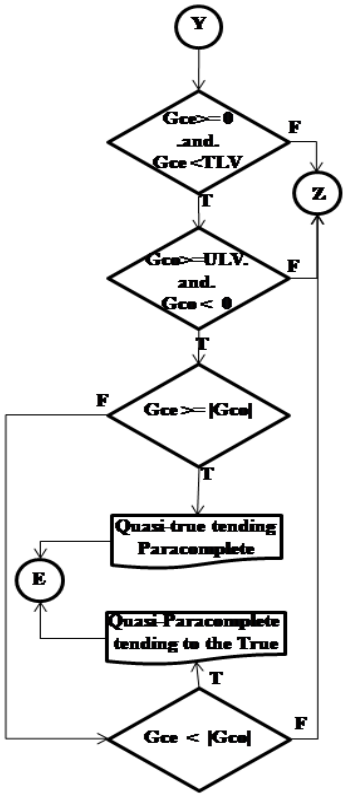


Fig. 6. Paraconsistent logical flowchart: Quasi True tending to Paracompletenessn, Paracompletenessn tending to the True

The flowchart (Fig. 7), there is an excellent possibility of being quasi-false tending to full Paracompletenessn, or full Paracompletenessn tending to false, because the Gce and Gco conditions result in some response. Moreover, when there is no possibility of an answer, it follows in the flow "W" to explore the possible answers offered by the paraconsistent logic.

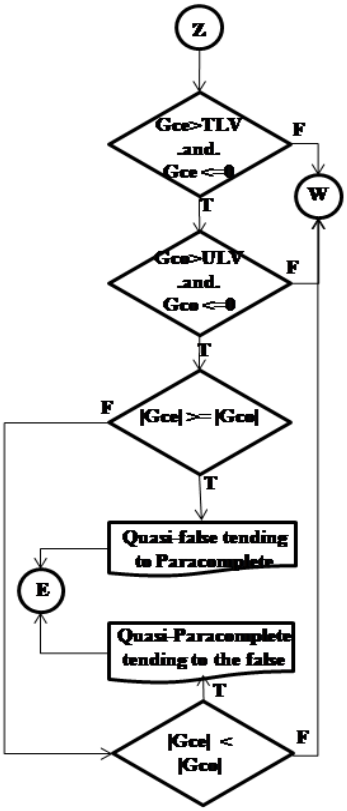


Fig. 7. Paraconsistent logical flowchart: Quasi-False tending to Paracompletenessn, Paracompletenessn tending to False

In this final step of the flowchart (Fig. 8), there is a possibility of being quasi-false tending to the inconsistent, or inconsistent tending to false, because the Gce and Gco conditions result in an answer.

The functionalities of registering data on organizational climate, registering data on organizational culture, consulting indicative (Artificial Intelligence) of support on climate and/or culture, should reflect in the System of choice for pre-Six Sigma projects.

With this, the flowchart helps in the construction of the algorithm that must reflect the use of the Paraconsistent Annotated Logic.

For this, it is necessary to 'translate' what we need into logic Eτ language, with a proposition that reflects the problem and their favorable and contrary evidence, as required by the underlying logic considered.

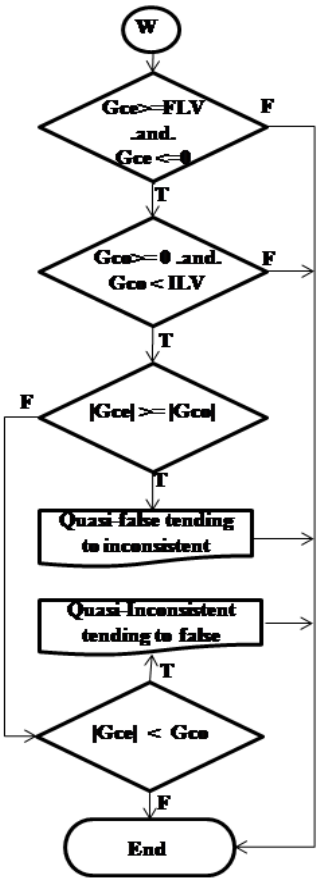


Fig. 8. Paraconsistent logical flowchart: Quasi False tending to Inconsistent, Inconsistent tending to False (Source: Luiz A. de Lima).

V. Final Considerations

Inconsistencies and human errors continue making the decision-making processes involved, as well as affect the production within an organization. Calculate is not enough, it is necessary to make these calculations automated, easy access to the user. Make the decision-making process more accurate, reliable and fast. The production and operations grew to become the most common errors within the activities, the search for the improvement of operations and the quality of the same, brought the study and development of Six Sigma, which proved to be a useful tool and produced results that demonstrate the improvement in processes and production. In addition to this study, in order to support managers for the implementation of the six sigma methodology, we seek artificial intelligence techniques and, in particular, parachutist logic, aid in decision making with more accuracy and even allowing the refuse in the implementation of six sigma projects, when considering factors such as climate and / or organizational culture.

ACKNOWLEDGMENT

Thanks to the Software Engineering Research Group and the University Paulista.

REFERENCES

[1] D. Nevalainen, L. Berte, C. Kraft, E. Leigh, L. Picaso, & T. Morgan, “Evaluating Laboratory Performance on Quality Indicators With the Six Sigma Scale”, Archives of Pathology & Laboratory Medicine: April 2000, Vol. 124, No. 4, pp.516-519, 2000.

[2] J. Antony, R. Banuelas, "Key ingredients for the effective implementation of Six Sigma program", Measuring Business Excellence, Vol. 6 Issue: 4, pp.20-27, 2002.

[3] M. Van Houtte, “Tracking effects on school achievement: A quantitative explanation regarding the academic culture of school staff”, American Journal of Education, 110(4), pp.354–388, 2004.

[4] D. Venanzi, B.P. Laporta, “Learn six sigma”, South American Development Society Journal, [s.l.], v. 1, n. 2, ISSN 2446-5763, pp. 66 - 84, 2017.

[5] J. Prosser, “The evolution of school culture research” In J. Prosser (Ed.), School culture, London: Paul Chapman, pp.1–14, 1999.

[6] C.Z. Kirilo, “Método Paraconsistente de Decisão aplicado ao Seis Sigma”, São Paulo, 2017, to appear.

[7] J.M. Abe, “Paraconsistent logics and applications”, In: 4th International Workshop on Soft Computing Applications.. - [s.l.] : IEEE, pp. 11–18, 2010.

[8] R. Payne & D. Pugh, Organizational Structure and Climate. In M. Dunnette (Ed.), “Handbook of Industrial and organizational psychology” Chicago: RandMcNally, pp.1125–1173, 1976..

[9] F.T.T. Phua, The role of organizational climate in socially embedding construction firms’ sustainability goals, Construction Management and Economics, (1), 2018.

[10] https://www.researchgate.net/figure/The-Six-Sigma-DMAIC-Process-and-Key-Outputs-17_fig4_273524820

[11] J.M. Abe, S. Akama, and K. Nakamatsu Introduction to Annotated Logic – Foundations for Paracomplete and Paraconsistent Reasoning, Series Intelligent Systems Reference Library, Vol. 88, Springer International Publishing, ISSN 1868-4394, Ed. 1, 190 pages, 2015.

[12] http://www.demneuropsy.com.br/detalhe_artigo.asp?id=95

[13] http://www.revistaespacios.com/a18v39n09/18390915.html

Organizational Culture Assessment Using Paraconsistent Logic Method of Evaluation

Caique Z. Kirilo[1], Jair M. Abe[2], Kazumi Nakamatsu[3], Marcelo Nogueira[4], Luiz Carlos Machi Lozano[5] and Luiz A. de Lima[6]

Abstract – **The present study aims to present a tool to analyze the organizational climate adherence through the integration of success codes and the Paraconsistent Decision Method. In this way, contributing to a previous scenario analysis that can return more precise feedback of the organizational culture conditions of the organization. The Paraconsistent Decision Method has the axillary function of the decision-making process, so when integrated into a corporate climate analysis survey the result is a more precise survey, where contradictions are treated in a different way, extracted to the maximum the information collected.**

Keywords – **Organizational culture; Paraconsistent Annotated Evidential Logic Eτ; Paraconsistent Decision-Making Method.**

I. INTRODUCTION

The consideration of a new quality management system will have, always, to a greater or lesser degree, rivals that will strive to hinder or even prevent the change, attitudes that can characterize ill will or even sabotage. Part of the literature relates to interpersonal trust many organizational variables such as citizenship behavior, cooperation, communication quality, performance, and troubleshooting. The trust involves expectations of benevolence and dependence on another party involved, where the actions of who trust are utterly related to their perceptions and attitudes of the target of trust [1].

Even the trust and organizational culture are distinct levels variables, micro, and macro; respectively, there are several indications of this relationship. Studies that analyze the relationship between trust and organizational culture, organizational culture as an antecedent of confidence, while considering the faith as a variable to facilitate processes of change in the culture of organizations [1].

Considering the organizational culture mainly as what is in the Organization, the confidence that employees have in the organization can contribute to anyone change. If it is appropriate to consider that the trust stimulates beneficial behaviors in the context of the work, it is possible to say that one's absence can encourage helpful little or nothing response [1].

Based on these notes it is possible to identify the relationship between reliability and organizational culture. Therefore, it is necessary to use a tool that can measure the success factors of the corporate culture to improve the security of the actors of the organization.

After this explanation, this study presents itself divided into five chapters. The theoretical framework is set out in Chapter 2; Emphasized the methodology in Chapter 3. Section 4 shows a discussion of the results, followed by their respective analyses. Moreover, finally, is given in Chapter 5 the final considerations of the present research.

II. THEORETICAL BACKGROUNDS

In this paragraph is presented all the contents that were used as the basis for this research, after vast bibliographical research on Paraconsistent logic and Organizational Culture was unable to collect the necessary material to develop a reliable path for the present work.

A. Organizational Culture

In the social sciences, culture can be understood as that which holds individuals of the same group in specific patterns of behavior, are their conscious elements or not of existence and culture. Manifestations, represented in some ways either by attitudes, gestures, habits, that is, the various actions expressed by members of a group, that make these individuals part of the Group [1].

Many authors narrow the relationship between organizational culture and organizational climate as is the case with Coda [2] [3], Souza and Schneider [4]. There are also authors that deal with environment and culture as synonymous, i.e., there's no way to talk about climate or culture separately.

According to Oliveira [5], talking about organizational

[1]Caique Zaneti Kirilo is Assistant at Software Engineering Research Group at Paulista University, Rua Antonio de Macedo, 505, Pq. São Jorge, São Paulo, Brazil, E-mail: caiquez.kirilo@hotmail.com

[2]Jair Minoro Abe is Full Professor at Graduate Program in Production Engineering at Paulista University, Rua Dr. Bacelar, 1212 - Vila Clementino, São Paulo, Brazil, E-mail: jairabe@uol.com.br

[3]Kazumi Nakamatsu is Full Professor at School of Human Science and Environment/H.S.E at University of Hyogo, Hyogo, Japan, E-mail: nakamatu@shse.u-hyogo.ac.jp

[4]Marcelo Nogueira is Researcher at ALGORITMI Centre at School of Engineering, University of Minho Campus Azurém, 4800-058 Guimarães, Portugal, E-mail: marcelo@noginfo.com.br

[5]Luiz Carlos Machi Lozano Assistant at Software Engineering Research Group at Paulista University, Rua Antonio de Macedo, 505, Pq. São Jorge, São Paulo, Brazil, E-mail: luizmlozano@gmail.com

[6]Luiz A.de Lima MSc student at Graduate Program in Production Engineering at Paulista University, Rua Dr. Bacelar, 1212 - Vila Clementino, São Paulo, Brazil, E-mail: luiz.lima@wcisp.com.br

climate refers almost necessarily speaking also to corporate culture as if they were conceptually inseparable.

Some of the authors who are organizational climate and organizational culture as synonyms are Katz and Khan [6] and [7] Milioni Toledo. Katz and Kahn conceptualize that environment, or corporate culture reflects the norms and values of the formal and informal system of the Organization, namely, the concept becomes vital to be outlined the formal and informal structure of the organization.

Organizational culture involves a complex range of assumptions and values. Such issues are discovered or invented by a particular group of people during attempts to adapt to external conditions or internal. Eventually, they become strong enough to be passed on to other members of the Organization, whether they are new members or not, the correct way to perceive, think, and feel in respects to such situations [8].

The intellectual development of people is a factor impacting on organizational culture because it is nothing more than a set of ideas, knowledge, behavior patterns and techniques learned in the group set [8].

The noticeable side of organizational culture is formal and can be compared to the tip of the iceberg, that is readily identifiable because it is equipped with technologies, artifacts, and creations. The other side, the imperceptible, is not easily recognizable because it consists of values, customs, and assumptions that are deeply rooted and implicit within the Organization [9].

To define culture as a set of values, norms and beliefs that standardize and guide the behavior of the particular group, note that the concept of culture is more vast and broad definition of organizational climate [9].

Motta and Caldas [10] denote the organizational culture is a set of beliefs, expectations, and values that are shared between its participants and left as a legacy for future generations, and these rules of conduct apply since high even the most simple operator manages.

Every organization, even if informally, create one's own culture that for his time reflects the standards and formal values. The concept of motivation leads us to the idea of organizational climate, people are always in search of satisfaction of their needs and keep their emotional balance.

The organization can be compared to a social system that has different types of participants, which are represented by all the elements that receive incentives and bring contributions to the Organization's existence.

There are four types of participants, employees, investors, suppliers, and customers. Each participant assumes his role at any given time in the Organization, not necessarily needing to be inside her to act, but all relate to the Organization of the reciprocal form [8].

TABLE I
BUSINESS PARTNERS (CHIAVENATO, 2002)

Participants (Partners)	Contributions(Personal Investments)	Incentives (Expected Return)
Employees	Contribute to work, dedication, personal effort, performance, loyalty, attendance.	Motivated by salary, benefits, awards, praise, and recognition opportunities.
Investors	Contribute to money in the form of shares, loans, financing.	Driven by profitability, liquidity, dividends, the return on investment.
Suppliers	Contribute materials, raw materials, technologies, specialized services.	Motivated by prices, terms of payment, billing, warranty of new business.
Customers	Contribute money to the purchase of products/services and one's consumption or use.	Motivated by price, quality, terms of payment and satisfaction of needs.

A process of change, to be successful, must be well planned, well publicised, well justified and well executed. It is of paramount to consider the climate and culture of the organization. Need to count on the commitment of top management, without which inexorably fail. The responsibility is the desire for real and evidenced by visible evidence. The team responsible for the change must have a delegation of authority to senior management. The difference must start with the points that have a higher probability of success so that one can produce the effect demonstration, which will facilitate the most problematic aspects [11].

Six determinations of change should be considered:

Understanding: understanding the importance of change on the part of senior management;

Commitment: decision to make the transition by the persuasion of its significance;

Competence: the ability to perform the technical and administrative change;

Correction: the absence of errors in the process of change;

Communication: information and justification to all involved about the need and importance of change;

Continuity: ensuring that there will be no setback after implementation.

The changes for the better comes from new concepts and not the adoption of new methodologies. These methodologies help only in organizations that understand their real problems and can change their paradigms in favor of progress [11].

The teams involved in the process of the six determinations of change can come across incomplete, ambiguous, vague and often contradictory to a greater or lesser degree, making data analysis for decision making. A non-Classical Logic can present more reliable results for this scenario. The following will be given to Paraconsistent logic as a helper tool for analysis of these scenarios.

B. Paraconsistent Logic Annotated Evidential Eτ

This topic introduces as purpose to introduce the Paraconsistent Logic Annotated and Evidential, that will be the basis for subsequent studies.

Approximately, paraconsistent logic allows contradictions in theories based on them without trivialization. It is possible to note that Nikolai Vasiliev and Jan Łukasiewicz discovered Paraconsistent logician 1907 at Aristotelian level while Stanisław Jaśkowski in 1948 and Newton da Costa in 1958 independently introduced propositional level and higher-order logic, respectively [12].

da Costa has developed a family of Paraconsistent Logic, propositional and predicate calculus of first-order set theory, that is, all the standard logic levels [13].

The process of decision making is a rational process, in which a plan of action is chosen based on various. Every decision-making process produces a choice. The decision refers to the process of selecting a coherent way in certain situations [14]

All decision-making, in a company, affects the general condition. The path to be chosen must be parsed; one should support the decision-making process by setting ways that must be traversed and think about what is affected by that decision [14]

By analyzing the real world, uncertain and inconsistent situations, we notice that in most of them we have partial knowledge of the facts. However, it does not prevent the development of human reasoning, that is beyond binary relation of truth and falsehood. The need to demonstrate and handle situations of contradiction raised an underlying logic to formal systems, called Paraconsistent Logic [15]

The decision-making process is an excellent response. Some people have an ease with the decision-making process. However, others attach to the problem a disproportionate value to their reality that wrong choices are made. [14]

The Annotated Evidential And Paraconsistent Logic Et is a class of Paraconsistent Logic whose language contains propositions of type p(μ, λ), where p is a proposition and (μ, λ) is an annotation constant. μ indicates the favorable degree of evidence and λ the contrary evidence expressed by the proposition p. The evidence degrees of μ and λ range between the real numbers 0 and 1 [16].

The processing of input data by application of minimizing and maximizing connectives between the Atomic Formulas A and B, that define the resulting state of the output.

Considering two groups of experts A = (E1, E2) and B = (E3, E4), it can be shown the connective OR application, represented by the disjunction A v B:

E1 (μ1, λ1) OR E2 (μ2, λ2) = (max {μ1, μ2}, min {λ1, λ2}) = air (μ1, λ1)

E3 (μ1, λ1) OR E4 (μ2, λ2) = (max {μ1, μ2}, min {λ1, λ2}) = (μ2, λ2)

Then, the application of connective AND, among the signs noted in air and BR, representing the conjunction ∧ AIR BR:

R = (μ1, λ1) AND BR (μ2, λ2) = (min {μ1, μ2}, Max {λ1, λ2}) = R (μ1, λ1)

After applying the maximization and minimization, the degrees of certainty and uncertainty are obtained by Degree of certainty: Gce(μ, λ) = μ-λ; Degree of Uncertainty: Gun (μ, λ) = μ + λ-1.With the values of Gce and Gun obtained, identifies the logical State arising through the analysis of τ in lattice Fig. 2.

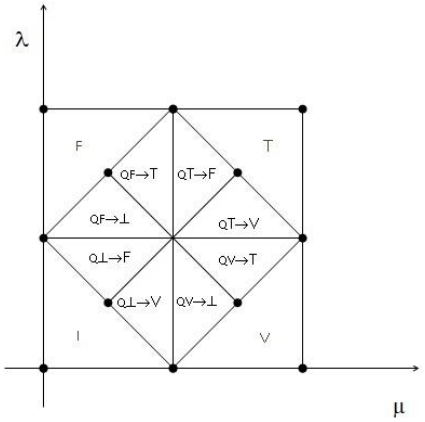


Fig. 2 – Extreme and non-extreme States Source: [15]

TABLE II - EXTREMES STATES SOURCE: [15]

Extreme State	Symbol
True	V
False	F
Inconsistent	T
Paracomplete	⊥

TABLE III NON-EXTREMES STATES - SOURCE: [15]

Non-extreme State	Symbol
Quasi-true tending to the Inconsistent	QV→T
Quasi-true tending to Paracomplete	QV→⊥
Quasi-false tending to the Inconsistent	QF→T
Quasi-false tending to Paracomplete	QF→⊥
Quasi-inconsistent tending to the True	QT→V
Quasi-inconsistent tending to False	QT→F
Quasi-paracomplete tending to True	Q⊥→V
Quasi-paracomplete tending to False	Q⊥→F

C. Paraconsistent Method of Decision

Based on studies of [20], you can synthesize Paraconsistent method definition of decision (MPD), which is a method that assists decision-making using Paraconsistent Logic.

Paraconsistent method of Decision was developed by [21], which sought to identify factors that influence the success or failure of a project, namely, that end up affecting the decision to carry out a plan or not. The analysis made it possible to identify what attributes can in some cases indicate favorable conditions in other unfavorable and other circumstances indifferent. These factors may be of different orders: economic, social, legal, environmental, political, technical, among others [21].

MPD uses as "input" (in) the experience of the participants in the decision-making process that are called experts as an

essential tool of assessment of the particular issue, enabling or precludes a situation any [17].

Starting a problem, question or note, which gets its name from the proposition, the method determines the need to finish the so-called factors, which as the name implies are the factors that impact on the viability or infeasibility of this proposition [18].

The factors can be severed to increase the accuracy of the analysis of a particular factor, sections created can extract more from the knowledge of the experts who are evaluating the [17].

Paraconsistent decision method consists of basically eight steps [20]:

1. Define the degree of demand that is parameterized on the decision-making process.

2. Define the factors that impact the proposition that will be parsed.

3. Set the sections that constitute the factors, to explain better the factor limits; there is no limit of sections to give the factor or a pattern to be followed.

4. Form the database, which can be formed by the weights also assigned factors and by evidence favorable factors and evidence to the contrary, that are deposited to each factor and its sections; such weights and opinions are taken from people who are considered experts in the field of knowledge that the proposition is inserted.

5. Carry out field research to establish, in which condition each of the factors.

6. Get the of the favorable evidence degree (ai, R) and the contrary evidence degree (bi, R) ($1 \leq i \leq n$) for each of the factors (Fi) and the sections (Spj) considered.

For that consider applications of the techniques of maximizing (MAX operator) and minimizing (MIN operator) of logic andτ.

7. Obtain the degree of favorable evidence (a_w) and the degree of evidence to the contrary (b_w) of the global analysis representing the chosen factors in the lattice t.

8. Finally, decide with the aid of the data obtained.

The theoretical basis for the MPD (Paraconsistent Method of the decision) is based on the rules of maximizing and minimizing of the Paraconsistent Annotated Evidential Logic Eτ.

Such rules are applied to favorable evidence degrees (μ) and contrary evidence degrees (λ).

The application of the rules of maximization and minimization can be performed as follows:

1. Making the maximization of degrees of evidence to a set of notes, to get:

The best evidence that is favorable (The highest value of favorable evidence μ)

The worst evidence that would be contrary (The highest value of contrary evidence λ)

2. Also, we consider the degree of certainty:

Gce = μ - λ, which in a way, reflects how much the information contained in this set allow to infer the truth or the falsity of the premise. (This form is more intuitive and leads to more predictable and consistent results).

Maximizing the degree of certainty (Gce) is seeking:

The best evidence that is favorable (The highest value of favorable evidence μ)

The best evidence would be contrary (The lowest value of favorable evidence λ)

Moreover, consequently, minimizing search:

The worst evidence that is favorable (The lowest value of favorable evidence μ)

The worst evidence that would be contrary (The highest value of favorable evidence λ)

The MPD, as a tool to aid decision making based on paraconsistent logic plays a vital role in the treatment of the views of its members, taking into account its contradictions and that in some instances it is significant for decision-making more accurate, therefore, has much to contribute to human relations within the Six Sigma. Six Sigma has tools contribute to the improvement of quality, but these tools are subject to human intervention, which can change the results and create a false impression of success, however, which does not hold in the long run.

III. Application

A. The Survey

The survey presented in this chapter is an adaptation of the work of Liza f. de Carvalho [23], which adapted your work, translating and improving the tool (OCAI ®). We consider now a well-known tool for the diagnosis of an organization's organizational culture Organizational Culture Assessment Instrument (OCAI) and consists of four sections composed of questions are assigned scores of 0 to 100. These scores are then calculated and will serve as a diagnosis of the current culture and preferred. The process used in the diagnosis is simple: first answered the questions knowing the current conditions of the Organization, then responds again, but now having in mind the situation, Quinn and Cameron [24] stress that was necessary during the distribution of some variables coming up so the two dimensions. The first dimension differentiates on the one hand effectiveness criteria focused on flexibility, discretion, and dynamism, of principles which emphasize the stability, order, and control. Some organizations are seen as active as remain in constant movement, adaptable and organic. Others are considered adequate if they remain stable and static [23]. The second dimension differentiates effectiveness criteria which emphasize internal guidance, integration, and unity, of external guidance criteria, differentiation, and rivalry. Some organizations work harmoniously focusing on interior features while others know the success and are recognized for work with other organizations external to itself [23]. So, if form four quadrants and born around these dimensions. These quadrants represent forces or opposing dynamics, typeset by Quinn. Each quadrant represents basic assumptions, guidelines, and values, the same elements that make up an organizational culture. The OCAI ®, as well, is an instrument that allows the diagnosis of the dominant orientation of the organization based on the types of nuclear culture [24]. The goal of the OCAI ® is to evaluate six dimensions of organization: Dominant Characteristics, organizational leadership, people management, Organizational, Strategic emphasis on Uniqueness and Success criteria, sorting into four types of dominant corporate culture: Team, Innovation,

processes and results and their respective foci, represented by the matrix below:

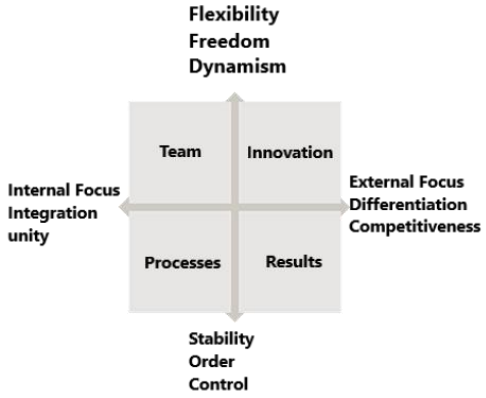


Fig. 3 – OCAI Array Source: (Cameron and Quinn, 2006)

Fill in the fields "I believe" with the percentage of how much you believe in the affirmation and the field "I do not believe" with the percentage of how much you do not believe in the statement		
Dominant Features		
	I believe	**I do not believe**
The organization is very familiar. People tend to share enough of themselves.		
The organization is very dynamic and enterprising. People are willing to commit to taking risks.		
The organization is very results-oriented. One of the most significant concerns is implementing the work. People are competitive and results-oriented.		
The organization is very controlled and structured. Formal and established standards guide people.		
Organizational Leadership		
In an organization, leadership is usually supporting example, counseling, and training		
In an organization, leadership is usually an example of entrepreneurship, innovation and willingness to take risks.		
In an organization, leadership is usually an example of objectivity, combativeness and results orientation.		
In an organization, leadership is usually an example of coordination and continuous management of efficiency.		
Management of employees		
The employee management style is characterized by promoting teamwork, consensus, and participation.		
The employee management style is characterized by valuing the autonomy, the risk, and the creative spirit		
The employee management style is characterized by developing the competitiveness, promote the results and achievement of goals.		
The management style is characterized by employees to ensure job security, compliance, and stability of relations		
Organizational Cohesion		
Loyalty and mutual trust are what keeps the organization cohesive. The commitment to the organization is high.		
What keeps the cohesive organization's commitment to innovation and growth. What's important is being on the cutting edge.		
What keeps the cohesive organization's commitment to achieving and exceeding goals. Win, be energetic and combative, are current themes.		
What keeps the cohesive organization are the formal rules and policies. Stability and balance are critical.		
Strategic Emphasis		
For the organization, it is essential to the development of people. Value trust, openness, and participation.		
For the organization, it is vital to the acquisition of new capabilities and the creation of new challenges. Value the new experiences and to prospect for new opportunities.		
For the organization, it is crucial to competitiveness and performance. The emphasis is on overcoming the ambitious goals and on market supremacy.		
For the organization, it is essential the permanence and stability. Focusing on the control, the efficiency and the fluidity of the operations.		
Success criterion		
The success of the organization is defined based on the development of human resources, the teamwork, the degree of involvement of its employees and concern for people.		
The success of the organization is defined based on the uniqueness and novelty of their products and services. It is a leading product and innovation.		
The success of the organization is defined based on conquest and supremacy of market. The posts are market leaders and competitive advantage.		
The success of the organization is defined based on efficiency. The crucial aspects are the satisfaction of commitments, planning and cost control.		

B. Collection of the data

After collecting the data obtained through the survey, there will be an array of knowledge compatible with a paraconsistent system, which works as follows:

• Step 1: receipt of the information.

The information is obtained using two independent variables, which are between 0 and 1, the first being the degree of favorable evidence and the second, the degree of evidence to the contrary.

• Step two: Data Processing.

The data are processed using the following equations:

a) GCT = (λ + μ)-1 to find the degree of contradiction.

b) GC = (μ-λ), to find the degree of certainty

• Step three: conclusion

To perform the completion, the following conditions:

a) and there is a high degree of contradiction, there is no certainty yet about the decision. Therefore, they must seek new evidence.

b) and there is a low degree of Contradiction, one can formulate the conclusion since it has a high degree of certainty.

IV. Discussion Of Results

This Work, with a broad view on the subjects discussed, invites the reader to reflect on the use of paraconsistent logic as a way to improve analysis and assessment of organizational culture. By studying the organizational culture, both in theory and in practice it was possible to analyze that even being a qualitative process based on reviews and human sensors, it is necessary that there are statistical techniques of high reliability to maintain organizational culture healthy.

A. Analysis of the results

The bibliographic survey was of importance to elucidate all the way by which the research would pass, many authors have dedicated much of their lives to consolidate the concepts employed in this research, denoting the importance of latent areas that have been addressed.

The resulting survey work analyses the organizational culture by assisting in paraconsistent, method such factors are of concern for the success of the survey presented because its perspective differs from common surveys.

Affirmations instead of questions were used to put the expert in the proposed scenario, and so he returns with the data of how that scenario is compatible with his reality.

Thus creating a context analysis more comfortable for the respondent, there are no wrong answers, including its contradictions will be used in the decision-making process.

V. Final Considerations

Times have changed, the way an organization behaves is not the same as years ago, where previously the Moldovan culture organization of employees, today the environment shapes the Organization's culture. Even when the focus of the organization is not directly related to improving the organizational culture, as in the implementation of Six Sigma, such factors nowadays are indispensable for the successful implementation of the quality improvement process, because the Organization is made by individuals who are entirely affected by organizational culture. However, in practice, there is no significant improvement. The Surveys can be reworked and customized according to each case study, keeping the main base and adding the desired score to each particular job.

Acknowledgment

FCT – Fundação para a Ciência e Tecnologia supported part of this work in the scope of the project: PEst-OE/EEI/UI0319/2013 by Portugal and University Paulista - Software Engineering Research Group by Brazil.

References

[1] E. G. S. Nunes, "Cultura Organizacional e confiança dos empregados na Organização: O Caso de uma Multinacional Alemã," in Instituto De Ciências Humanas E Sociais - Mestrado Em Gestão E Estratégia De Negócios, Seropédica, RJ, UFRRJ, 2009.

[2] R. G. Silva, "Empresa e Família: instituições que se entretecem na continuidade de uma cultura organizacional", MSc Dissertation (in Portuguese), Faculdade de Gestão de Negócios, Federal University of Uberlândia, 2010.

[3] R. Coda, “Pesquisa de clima organizacional: uma contribuição metodológica” (in Portuguese), São Paulo, 1992.

[4] E. L. P. D. Souza, “Diagnóstico de clima organizacional”, (in Portuguese) Rio de Janeiro: Revista de Administração Pública, 1977.

[5] B. Schneider and D. T. Hall, “Correlates of Organizational Identification as a Function of Cancer Pattern and Organizational Type”, Ithaca: Administrative science quarterly, 1972.

[6] M. A. Oliveira, *Pesquisas De Clima Interno Nas Empresas: O Caso dos Desconfiômetros Avariados* (in Portuguese), São Paulo: Nobel, 1995.

[7] D. Katz and R. L. Khan, *Psicologia social das organizações*, São Paulo: Atlas, 1976.

[8] F. M. B. Toledo, *Dicionário de Recursos Humanos*, São Paulo: Atlas, 1986.

[9] I. Chiavenato, *Iniciação à Teoria das Organizações*, Barueri: Manole, 2010.

[10] M. S. D. O. Mello, “A Qualidade do Clima Organizacional como Variável Interveniente no Desempenho do Trabalho Humano: Um estudo de caso na empresa Herbarium”., Santa Catarina, 2004.

[11] F. C. P. &. C. M. P. Motta, *Cultura Organizacional e Cultura*, São Paulo: Atlas, 1997.

[12] P. L. d. O. Costa Neto and S. A. Canuto, *Administração Com Qualidade: Conhecimentos Necessários para a Gestão Moderna*, São Paulo: Blücher, 2010.

[13] J. M. Abe, S. Akama, and K. Nakamatsu, *Paraconsistent Intelligent-Based Systems - New Trends in the Applications of Paraconsistency*, 1 ed., vol. 1, Switzerland: Springer International Publishing, 2015.

[14] F. R. &. A. J. M. Carvalho, Tomadas de Decisão com Ferramentas da Lógica Paraconstistente Anotada, São Paulo: Edgard Blucher Ltda., 2011.

[15] T. Shimizu, Decisão nas Organizações, vol. 2 ed., São Paulo, SP: Atlas, 2006.

[16] J. M. Abe, H. F. Lopes, and K. Nakamatsu, "Paraconsistent artificial neural networks and EEG.," International Journal of Knowledge-Based Intelligent Engineering Systems, vol. 17, no. 2, pp. 99-111, 2013.

[17] J. M. Abe, J. I. da Silva Filho, U. Celestino and H. C. d. Araújo, Lógica Paraconsistente Anotada Evidencial Et, J. M. Abe, Ed., Santos: Comunicar, 2011.

[18] J. I. Da Silva Filho, "Métodos de Aplicações da Lógica Paraconsistente Anotada com Dois Valores - LPA2v com Construção de Algoritmo e Implementação de Circuitos Eletrônicos," Tese de doutorado apresentada a EPUSP, 1999.

[19] F. R. Carvalho, "Aplicação de lógica paraconsistente anotada em tomadas de decisão na engenharia de produção", Doctor Thesis (in Portuguese), University of São Paulo, São Paulo, 2006.

[20] F. R. de Carvalho and J. M. Abe, Fuzzy Decision Method and its Comparison with the Paraconsistent Decision Method, to appear, 2018.

[21] J. M. Abe, "Paraconsistent Artificial Neural Networks: An Introduction," in Knowledge-Based Intelligent Information and Engineering Systems, vol. 3214, M. Negoita, R. Howlett and L. Jain, Eds., Springer Berlin Heidelberg, pp. 942-948, 2004.

[22] W. G. Zikmund, B. J. Babin, J. C. Carr and M. Griffin, *Business Research Methods*, Ninth Edition ed., C. Learning, Ed., South-Western College Pub, 2010.

[23] K. S. Cameron and R. E. Quinn, *Diagnosing and Changing the Organizational Culture: Based on the competing values framework*, San Francisco: Wiley (Jossey Bass), 2006.

An Expert System for Uncertain, Inconsistent, and Paracomplete Data Decision-Making

Jair M. Abe[1], Kazumi Nakamatsu[2], Seiki Akama[3], and Ari Ahrary[4]

Abstract – **Nowadays deal with big data or lack of data are an ordinary matter. We need more formal and generic tools to deal with. In this review paper, we discuss a skeleton of an expert system which can deal with such concepts, uncertain, inconsistent and paracomplete data without the danger of trivialization. Such skeleton leans on paraconsistent annotated evidential logic Eτ. More specifically, in the underlying of the lattice of truth-values and an algorithm called para-analyzer.**

Keywords – **Expert systems; Inconsistency; Paracompleteness; Uncertainty; Decision-Making.**

I. Introduction

The contradictions or inconsistencies, as well as uncertainties, are common when we make the description of parts of the real world. The control systems used in Automation and Robotics and the Expert Systems used in AI perform this description, in general, with base in the classical logic. The classical logic makes this description considering only two states. These common binary systems cannot manipulate the contradictory situations appropriately. Paraconsistent logic was born of the need to find means for giving a non-trivial treatment to the contradictory situations. The Paraconsistent logics have presented results that make possible to consider the inconsistencies in its structure in a non-trivial way. Abe and collaborators, has shown some applications of the concept of paraconsistency in areas mentioned above [2], [4], [5].

II. The Paraconsistent Annotated Logic Eτ

The atomic formulas of the paraconsistent annotated logic $E\tau$ is of the type $p_{(\mu, \lambda)}$, where $(\mu, \lambda) \in [0, 1]^2$ and [0, 1] is the real unitary interval (p denotes a propositional variable). There is an order relation defined on $[0, 1]^2$: $(\mu_1, \lambda_1) \leq (\mu_2, \lambda_2) \Leftrightarrow \mu_1 \leq \mu_2$ and $\lambda_1 \leq \lambda_2$. Such ordered system constitutes a lattice that will be symbolized by τ. A detailed account is to be found in [1], [3].

[1]Jair M. Abe - Graduate Program in Production Engineering, Paulista University, R. Dr. Bacelar, 1212, CEP 04026-002 São Paulo – SP, Brazil, E-mail: jairabe@uol.com.br.
[2] Kazumi Nakamatsu - School of Human Science and Environment/H.S.E. – University of Hyogo, Japan, nakamatu@shse.u-hyogo.ac.jp
[3] Seiki Akama, C-Republic, Tokyo, Japan, E-mail: akama@jcom.home.ne.jp
[4] Alireza Ahrary - Faculty of Computer and Information Sciences - Sojo University, Kumamoto, Japan, E-mail: ahrary@ieee.org

$p_{(\mu, \lambda)}$ can be intuitively read: "It is believed that p's belief degree (or favorable evidence) is μ and disbelief degree (or contrary evidence) is λ." [1]

So, we have some interesting examples:

- $p_{(1.0, 0.0)}$ can be read as a true proposition.
- $p_{(0.0, 1.0)}$ can be read as a false proposition.
- $p_{(1.0, 1.0)}$ can be read as an inconsistent proposition.
- $p_{(0.0, 0.0)}$ can be read as a paracomplete (unknown) proposition.
- $p_{(0.5, 0.5)}$ can be read as an indefinite proposition.

Note. The concept of paracompleteness is the "dual" of the concept of inconsistency.

The consideration of the values other than evidence, such as belief degree and disbelief degree is made, for example, by experts that use heuristics knowledge, probability [12] or statistics [13].

The output can be of two types: situations of *extreme states* that are, False, True, Inconsistent and Paracomplete, and the situations of *non-extreme states*, all these situations are represented in the lattice presented in the next figure:

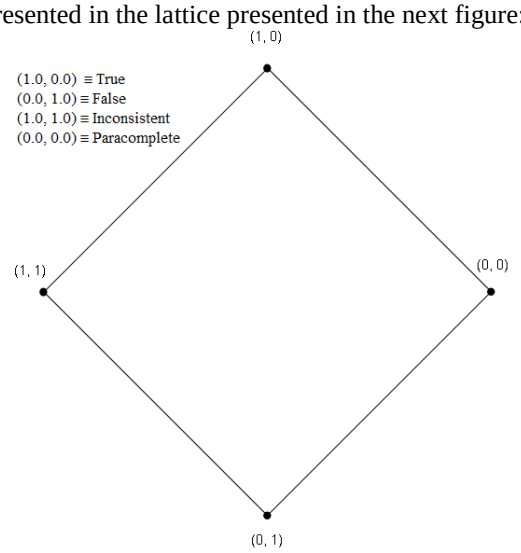


Fig. 1. Lattice τ

There is a natural operator in the lattice $\sim : |\tau| \rightarrow |\tau|$ is defined as $\sim [(\mu, \lambda)] = (\lambda, \mu)$. Such operator works as the "meaning" of the logical negation of the logic $E\tau$.

Also we have the operations OR and AND:

(μ_1, λ_1) OR $(\mu_2, \lambda_2) = (\text{Max}\{\mu_1, \mu_2\}, \text{Min}\{\lambda_1, \lambda_2\})$

(μ_1, λ_1) AND $(\mu_2, \lambda_2) = (\text{Min}\{\mu_1, \mu_2\}, \text{Max}\{\lambda_1, \lambda_2\})$

Where Max and Min are the usual maximization and minimization operations on real numbers with usual order.

The usual cartesian system can represent the lattice τ.

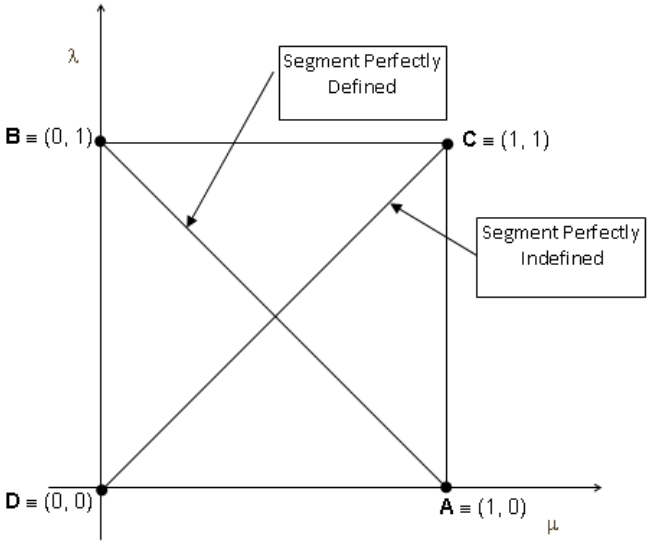


Fig. 2. Segments Perfectly Defined and Undefined

We can consider several important segments:
Segment *DB* - segment perfectly defined: $\mu + \lambda - 1 = 0$
Segment *AC* - segment perfectly undefined: $\mu - \lambda = 0$
Segment *DB* - segment perfectly defined: $\mu + \lambda - 1 = 0$
Segment *AC* - segment perfectly undefined: $\mu - \lambda = 0$
Uncertainty Degree: $G_{un}(\mu, \lambda) = \mu + \lambda - 1$; Certainty Degree: $G_{ce}(\mu, \lambda) = \mu - \lambda$;

To fix ideas, with the uncertainty and certainty degrees we can get the following 12 regions of output: *extreme states* that are, False, True, Inconsistent and Paracomplete, and *non-extreme states*. All the states are represented in the lattice of the next figure: such lattice τ can be represented by the usual Cartesian system (Figure 4).

These states can be described with the values of the certainty degree and uncertainty degree using suitable equations. In this work, we have chosen the resolution 12 (number of the regions considered according to the Figure 1), but the resolution is entirely dependent on the precision of the analysis required in the output, and it can be externally adapted according to the applications.

So, such limit values called Control Values are:

V_{cic} = maximum value of uncertainty control = C_3
V_{cve} = maximum value of certainty control = C_1
V_{cpa} = minimum value of uncertainty control = C_4
V_{cfa} = minimum value of certainty control = C_2
In this paper we have used: $C_1 = C_3 = ½$ and
$C_2 = C_4 = -½$.

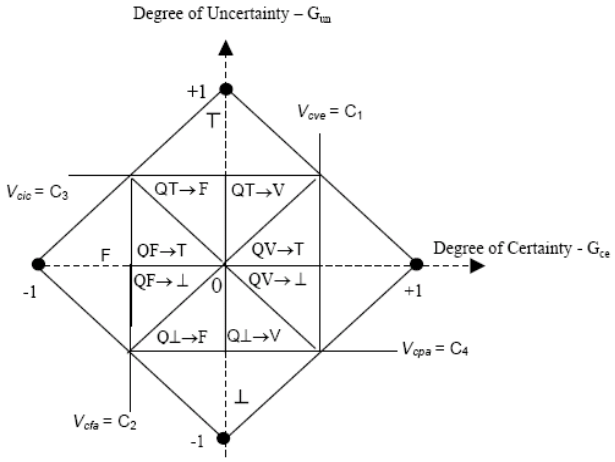


Fig. 3. Representation of the certainty degrees and uncertainty degrees.

With the values in the lattice, some regions can be considered in the unitary square of the Cartesian plan that will define the *outputs resulting states*.

These states can be described with the values of the certainty degree and contradiction degree using the equations.

In this work, we have chosen the resolution 12 (number of the regions (states) considered according to in Figure 4), but the resolution is entirely dependent on the precision of the analysis required in the output. Also, the resolution (states) can be easily modified according to each application and accuracy requested.

TABLE I
EXTREME AND NON-EXTREME STATES

Extreme States	**Symbol**
True	V
False	F
Inconsistent	T
Paracomplete	⊥
Non-extreme states	**Symbol**
Quasi-true tending to Inconsistent	QV→T
Quasi-true tending to Paracomplete	QV→⊥
Quasi-false tending to Inconsistent	QF→T
Quasi-false tending to Paracomplete	QF→⊥
Quasi-inconsistent tending to True	QT→V
Quasi-inconsistent tending to False	QT→F
Quasi-paracomplete tending to True	Q⊥→V
Quasi-paracomplete tending to False	Q⊥→F

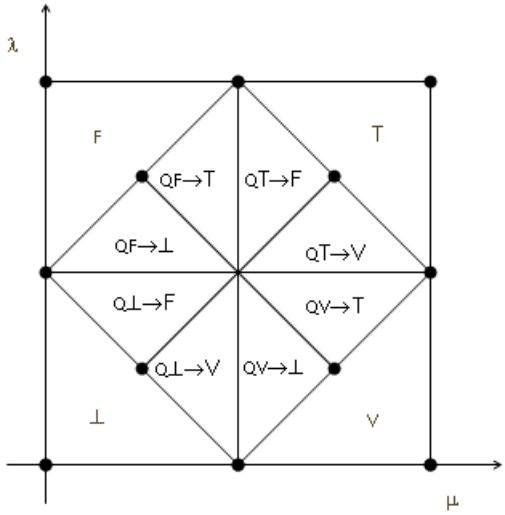


Fig. 4. Representation of the *extreme* and *non-extreme* state's regions

To make easier the recognition of each region, each one received a denomination in agreement with its proximity to the extreme states points of the lattice.

The algorithm that expresses the calculations of the inputs μ and λ is:

III. Algorithm Para-analyzer

In what follows, we present the algorithm para-analyzer [10]. The primary concern in any analysis is to know how to measure or to determine the certainty degree regarding a proposition if it is False or True. Therefore, for this, we take into account only the certainty degree G_{ce}. The uncertainty degree G_{un} indicates the measure of the inconsistency or paracompleteness. If the certainty degree is low or the uncertainty degree is high, it generates an indefinite.

The resulting certainty degree G_{ce} is obtained as follows:

If: $V_{cfa} \leq G_{un} \leq V_{cve}$ or $V_{cic} \leq G_{un} \leq V_{cpa}$ $\Rightarrow$ G_{ce} = Indefinite

For: $V_{cpa} \leq G_{un} \leq V_{cic}$

If: $G_{un} \leq V_{cfa} \Rightarrow$ G_{ce} = False with degree G_{un}

$V_{cic} \leq G_{un} \Rightarrow$ G_{ce} = True with degree G_{un}

The algorithm Para-analyzer is as follows:

```
   */ Definitions of the values */
   Max_vcc = C1 */ maximum value of
certainty Control*/
   Max_vctc = C3 */ maximum value of
uncertainty control*/
   Min_vcc = C2 */ minimum value of
certainty Control */
   Min_vctc = C4 */ minimum value of
uncertainty control*/
   */ Input Variables */
   μ
   λ
   */ Output Variables */
   digital output = S1
   Analogical output = S2a
   Analogical output = S2b
   * / Mathematical expressions * /
   being:
   0≤ μ ≤ 1 and 0≤ λ≤ 1
   Gun(μ; λ) = μ + λ - 1;
   Gce(μ; λ) = μ - λ
   * / determination of the extreme
states * /
   if Gce(μ; λ) ≥ C1 then S1 = V
   if Gce(μ; λ) ≥ C2 then S1 = T
   if Gun(μ; λ) ≥ C3 then S1 = F
   if Gun(μ; λ) ≤ C4 then S1 = ⊥
   */ determination of the non-extreme
states * /
   for 0 ≤ Gce < C1 and 0 ≤ Gun < C2
   if Gce ≥ Gun then S1 = QV→T
   else S1 = QT→V
   for 0 ≤ Gce < C1 and C4 < Gun ≤ 0
   if Gce ≥ | Gun | then S1 = QV→⊥
   else S1 = Q⊥→V
   for C3 < Gce ≤ 0 and C4 < Gun ≤ 0
   if |Gce | ≥ | Gun | then S1 = QF→⊥
   else S1 = Q⊥→F
   for C3 < Gce ≤ 0 and 0 ≤ Gun < C2
   If |Gce | ≥ Gun then S1 = QF→T
   else S1 = QT→F
   Gct = S2a
   Gce = S2
   */ END */
```

In this way, contradictory, paracomplete, and uncertainty information can be treated in a close approach to the reality, through combinations of evidence.

The external adjust permitted in the defined regions of the unitary square of degrees, make the applications of the "Para-analyzer" more natural and more faithful when in the elaboration of control systems for Automation areas, Artificial Intelligence, and Robotics. The Para-analyzer also allows optimization and offers good controllability of operations, including crucial situations of the real world. The visualization through the Hasse's diagram of the lattice, with the axis of values of certainty degrees and the contradiction degrees, gives a more realistic vision of the situations through sensor information of the environment at any moment, portraying several conditions more entirely and faithfully. Therefore, the fundamental importance of the algorithm presented is to show that the Paraconsistent logic is applicable in real systems.

Some examples of applications are in Expert Systems, Neural Networks, Robotics and Artificial Intelligence [2], [8].

IV. Application to Increase Industrial Equipment Availability Through Maintenance

Industrial equipment needs a permanent availability through conservation. Information about the material is captured by many devices such as sensors, visual cameras, etc. Such information almost always contains uncertainties and conflicts. We show how the Para-analyzer can be useful to increase equipment availability through maintenance.

Suppose that three experts are making the maintenance of some equipment. The chief engineer receives an amount of information, each of them is a proposition with a certainty degree and uncertainty degree (or if it is the case, favorable evidence and contrary evidence degrees): for instance, last maintenance, past recordings, etc.

The information, whatever they come, can have agreements, disagreements or even uncertainties. The Para-analyser proposed can perform paraconsistent reasoning that will analyze each evidence for the favorable and contrary evidence. A suggested form for this implementation is the use of a maximization analysis with the connective OR and minimization with the connective AND among the three experts' information. The figure 5 display this implementation.

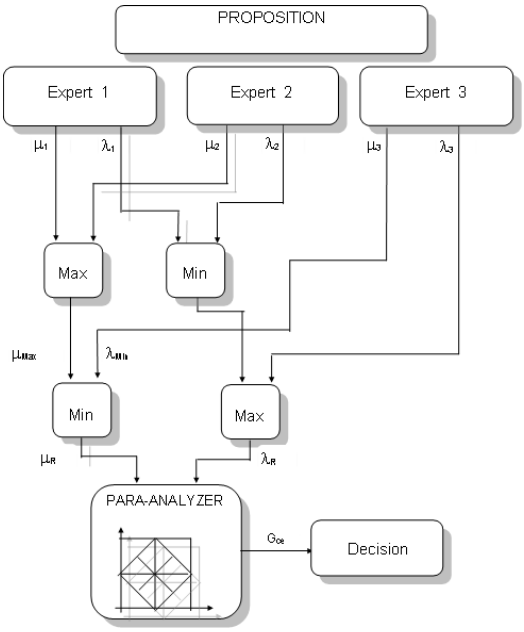


Fig. 5. Symbolic net of the information given by three experts.

The next table II displays the net with more details, where it stands out the action Para-analyzer in the information brought by the three experts.

TABLE II
RESULTING ANALYSIS BY THE THREE EXPERTS

E1	E2	OR	E3	AND	RESULTING STATE
⊤	⊤	⊤	⊤	⊤	A
⊤	V	⊤	V	V	B
⊤	F	⊤	F	F	C
⊤	⊥	⊤	⊥	⊥	D
V	⊤	⊤	⊤	⊤	E
V	V	V	V	V	F
V	F	⊤	F	F	G
V	⊥	V	⊥	⊥	H
F	⊤	⊤	⊤	⊤	I
F	V	⊥	V	V	J
F	F	F	F	F	K
F	⊥	F	⊥	⊥	L
⊥	⊤	⊤	⊤	⊤	M
⊥	V	V	V	V	N
⊥	F	F	F	F	O
⊥	⊥	⊥	⊥	⊥	P

-

E1 = Expert 1
E2 = Expert 2
E3 = Expert 3

OR = maximization
AND = minimization

Description of each resulting State of the table:

State A - All the three experts are presenting inconsistent values. Therefore the system should go back requesting new information.

State B - Two of the three experts are presenting true values. Therefore the system concludes that the proposition is true.

State C - Two of the three experts are presenting false values. Therefore the system concludes that the proposition is false.

State D - Two of the three experts are presenting indefinite values. Therefore the system should go back requesting more information for the two experts that present amount of insufficient data.

State E - Two of the three experts are presenting inconsistent values. Therefore the system should come back requesting new information for the two experts that present contradictions.

State F - All the three experts are presenting true values. Therefore the system concludes that the proposition is true.

State G - Two of the three experts are presenting false values. Therefore the system concludes that the proposition is false.

State H - Two of the three experts are presenting indefinite values. Therefore the system should come back requesting more information for the two experts that present amount of insufficient data.

State I - Two of the three experts are presenting inconsistent values. Therefore the system should go back requesting new information for the two experts that present contradictions.

State J - All the three experts are presenting true values. Therefore the system concludes that the proposition is true.

State K - Two of the three experts are presenting false values. Therefore the system concludes that the proposition is false.

State L - Two of the three experts are presenting indefinite values. Therefore the system should come back requesting more information for the two experts that present amount insufficient of information.

State M - A specialist is presenting indefinite values and the other ones two they are presenting inconsistent values. Therefore the system should come back requesting more

information for the specialist than he/she has the little amount and new information for the two that present contradictions.

State N - A specialist is presenting indefinite values and the other ones two they are presenting true values. Therefore the system concludes that the proposition is true.

State O - A specialist is presenting indefinite values and the other ones two they are showing false values. Therefore the system concludes that the proposition is false.

State P - All the three experts are presenting indefinite values. Therefore the system should come back requesting more information because all are coming in insufficient amounts for analyses.

The return conditions described in the analysis of table II they do not necessarily force that the system specialist paraconsistent should make new searches of evidence, but, to the opposite, until they can be conclusive depending on the resulting values of the certainty degrees and uncertainty.

The values of certainty degrees and of uncertainty they can decide the conclusion in the cases in that, after the analysis, are obtained resulting states of inconsistencies or inconclusive. The certainty degree Gce and the one of uncertainty degree Gun are values that can be compared with the resultants of other analyses and the system can then, to decide for that that presents a smaller uncertainty degree or greater certainty degree, depending on the application of the project.

In this example, the paraconsistent expert system that analyses the proposition "the equipment need to make maintenance," examines various information. The result brings as answer a been resulting suitable for two values of certainty degree Gce and of uncertainty Gun that gives full conditions to the control System to make a decision and to choose when is the instant to stop.

V. Conclusion

In this paper, we have sketched an expert system based on Paraconsistent Annotated Evidential Logic Eτ. Due to the properties of logic Eτ, such expert systems can deal directly with imprecise, inconsistent, and paracomplete data, analyzing more reliable and realistic data, without the need to use extra-logical devices.

A suitable combination of the para-analyzer algorithm can be used as a tool in a flexible ways, can be applied in problems of greater complexity.

Acknowledgment

Thanks to J.I. da Silva Filho for useful comments.

References

[1] J.M. Abe, “Foundations of Annotated Logics”, Doctor thesis (in Portuguese) University of São Paulo, Brazil, 1992.

[2] J.M. Abe, *Paraconsistent Intelligent Based-Systems: New Trends in the Applications of Paraconsistency*, editor, Book Series: “Intelligent Systems Reference Library,” Springer-Verlag, Vol. 94, ISBN:978-3-319-19721-0, 306 pages, Germany, 2015.

[3] J.M. Abe, S. Akama, K. Nakamatsu, *Introduction to Annotated Logics - Foundations for Paracomplete and Paraconsistent Reasoning*, Series Title Intelligent Systems Reference Library, Volume 88, Springer International Publishing, Copyright Holder Springer International Publishing Switzerland, ISBN 978-3-319-17911-7, Edition Number 1, 190 pages, 2015.

[4] J.M. Abe, “Some Aspects of Paraconsistent Systems and Applications”, Logique et Analyse, 157, 83-96, 1997.

[5] S. Akama, *Towards Paraconsistent Engineering*, Intelligent Systems Reference Library, Volume 110, 234 pages, 2016, ISBN: 978-3-319-40417-2 (Print) 978-3-319-40418-9 (Online), Series ISSN 1868-4394, Publisher Springer International Publishing, DOI: 10.1007/978-3-319-40418-9

[6] J.I. da Silva Filho, G.L Torres & J.M. Abe, *Uncertainty Treatment Using Paraconsistent Logic - Introducing Paraconsistent Artificial Neural Networks*, IOS Press, Netherlands, Vol. 211, ISBN 978-1-60750-557-0, doi: 10.3233/978-1-60750-558-7-I, 328pp., 2010.

[7] J.I. da Silva Filho, “Métodos de interpretação da Lógica Paraconsistent Anotada com anotação com dois valores LPA2v com construção de Algoritmo e implementação de Circuitos Eletrônicos”, Ph.D. Thesis (in Portuguese), University of São Paulo, São Paulo, 1999.

[8] N.F. Reis, “Método Paraconsistente de Cenários Prospectivos”, PhD thesis, in Portuguese, Paulista University, 2014.

[9] J.P.A. Prado, “Uma Arquitetura em IA Baseada em Lógica Paraconsistente”, PhD Thesis (in Portuguese), University of São Paulo, 2006.

[10] J.I. da Silva Filho, J. M. Abe, “Paraconsistent analyzer module”, International Journal of Computing Anticipatory Systems, vol. 9, ISSN 1373-5411, ISBN 2-9600262-1-7, pp.346-352, 2001.

[11] P. Jackson, *Introduction To Expert Systems* (3 ed.), Addison Wesley, p. 2, ISBN 978-0-201-87686-4, 1998.

[12] A.P. Dempster, “Generalization of Bayesian inference”, Journal of the Royal Statistical Society, Séries B-30, pp.205-247, 1968.

[13] R.O. Duda, P.E. Hart, K. Konolid & R.Reboh, *A computer-based Consultant for Mineral Exploration*, TR, SRI International, 1979.

Checking of Real-Time Control Logic Specifications using Logic Programming and SMT solvers

Andrey Tyugashev[1] and Dmitrii Zheleznov[2]

Abstract –**In this paper the information about application of SMT solvers in area of Mission Critical Software verification are given. Rules of verification are based on Real-Time Control Algorithm's Logic. Required specification can be feasible or non-feasible on defined basis of functional control processes. In proposed approach, feasibility of the specification is being checked by SMT solver Z3. SMT Solver is called from special Java application through API.**

Keywords –**control logic; functional process; logical vector; real-time control algorithm; logic programming; SMT solver.**

I. Introduction

The modern technical object such as airplane, submarine, spacecraft or nuclear power station can be reviewed as 'system of the systems' including a lot of subsystems, actuators, sensors, other devices. Like an orchestra playing symphony, all of these devices should co-function in harmonic manner to produce a useful outcome. Each instrument must to start play at a right time. In orchestra, the conductor performs control functions. In modern complex technical complexes, the control system should provide the same functionality. The human could be involved in the process in case of automated control, or not be involved in case of automatic system. Discussing complexity level of control system we can note that in according to Ashby's Law of Requisite Variety [1], "Variety absorbs variety", so the complexity of control system should be adequate to complexity of controlled object. Control system realizes corresponding control algorithms. The 'input data' for control algorithms is so-named 'control logic'. In fact, this logic is representation of coordinated functioning of all units needed to achieve the goal of our system. The 'coordinated' word means here both semantic coordination related to physical restrictions and logic of actions, and coordination in time. The time characteristics of control logic should be adequate to speed of ongoing physical processes associated with the controlled technical complex [2-5].

The very important problem for control logic of complex technical object is evaluation of its parameters and checking if these values are correspond to existing physical and technological constraints. This issue is actual both at design stage when the key question is feasibility of requirements, and during operation of existing technical object when we need, for example, to analyze performance. This paper is focused on timing (synchronization) parameters, and degree of use of accessible resources (level of workload/overload). The problem has an additional importance due to its straight connection to dependability/safety issues.

Today, as a rule, the control logic's evaluation is being performed by human. Unfortunately, the number of parameters which must be analyzed, for example, for modern spacecraft, can be very big and exceeds the human opportunities. The purpose of the work is to provide automation to this process. We utilize two approaches for evaluation of the control logic – use of SMT solvers, and logical programming.

Herewith, we can review potentially useful approach connected with apply of existing SMT automation tools to provide assistance to specialists responsible for control logic's evaluation [6-7]. The very popular and promising technology today is Satisfiability Modulo Theories (SMT) approach. SMT supported by a lot of commercial and free solvers such as ABSolver, Alt-Ergo, Barcelogic, MathSAT, CVC, OpenSMT, Simplify, STeP, Yices, Z3, etc. We can specify the existing constraints using smt-lib formal language, and then get the answer if the system satisfies (sat) the constraints, or not (unsat). The system even can calculate the values of the variables which provide satisfiability.

In this paper we want also to remind about power and opportunities provided by logic programming. In fact, internal logical inference machine provide us with opportunities comparable with features of modern SMT solvers. Moreover, the logic programming systems are very close by their nature to specificity of Real-Time Control Algorithm's Logic applied to checking of properties of control algorithms. So, we present the corresponding example of application of logic programming in our domain.

II. Method

A. Real-Time Control Logic

In previous papers [3,6,7] we had proposed the semantic model for real-time control algorithm. The model represents control actions by the set of following tuples:

$$RTCL = \{< f_i, t_i, \tau_i, l_i >\}, i=1..N \qquad (1)$$

f_i represents an identifier of functional process to be executed, and: t_i – time of f_i begin (non-negative integer), τ_i – its duration (non-negative integer). l_i is a 'logical vector' defining whether process should be executed. The logical vector consists of logical variables within checked values: ($\alpha_1=0$, $\alpha_2=1$, $\alpha_3=0$, $\alpha_4=H$, $\alpha_5=H$). Herewith, 1 and 0

[1]Andrey Tyugashev is with the Dept. of Applied Mathematics and Computer Science, Samara State Transport University, 2V Svobody Street, Samara 443066, Russia, E-mail: a.tyugashev@samgups.ru.

[2]Dmitrii Zheleznov is Rector of Samara State Transport University, 2V Svobody Street, Samara 443066, Russia, E-mail: rektorat@samgups.ru Blvd, Sofia 1000, Bulgaria.

corresponds to True and False, and 'H' value means that execution of the process is not depends on value of this logical variable. The presence of logical variables in the model allows specifying a set of options of implementation of the algorithm (including normal and abnormal situations).

Some parameters can be specified by a known constants, some be initially unknown and stated as variables.

The constraints and requirements for real-time control logic can be specified using language of CA formal theory (calculus of real-time control algorithms) proposed by A.A. Kalentyev [3-4]. The extended version of this theory developed by author [6] – Real-Time Control Algorithm's Logic allows its usage for real control logic specification and verification.

We focus on synchronization of functional processes to be executed. The synchronization of two processes can be expressed by following operators: coincidence by begin (named CH from Russian abbreviation), coincidence by end (named CK), direct following (→), time uncrossing (<>), precedence (<), strict precedence (<<), the overlap with the specified shift (H), parameterized following with the specification of the delay (3A). Table 1 unites short reference descriptions of them.

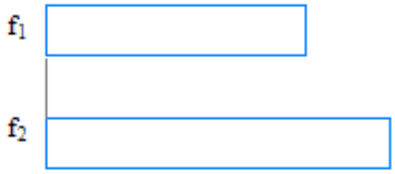


Fig. 1. Coincidence 'begin-begin' CH

The sense of operators becomes quite clear after looking at Fig1-6.

Fig. 2. Coincidence 'end-end' CK

The operators: <. << and <> expressed 'soft' bindings where times of processes' begins and ends may vary in some intervals.

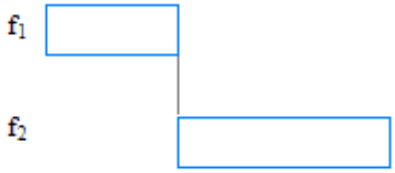


Fig. 3. Direct following

Fig. 4. Strict precedence

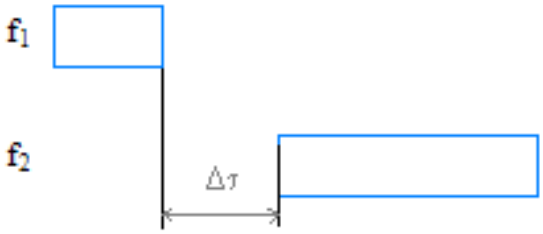


Fig. 5. Parameterized following

Special operator <*l*> means logical incompatibility of actions, i.e. the processes cannot be found in the same case of execution. This is means that the same logical variable has value 1 in one vector, and 0 in another.

Fig. 6. Parameterized overlap

These formal calculi are strong associated with algebraic models or real-time control algorithms [6].

TABLE I OPERATORS OF RTCL

Name	Mean	Signature
CH	'begin-begin'	$(UA_1, UA_2) \rightarrow UA$
CK	'end-end'	$(UA_1, UA_2) \rightarrow UA$
$\rightarrow$	direct following	$(UA_1, UA_2) \rightarrow UA$
H	parameterized overlay	$(UA_1, UA_2, int) \rightarrow UA$
3A	parameterized following	$(UA_1, UA_2, int) \rightarrow UA$
@	absolute time binding	$(UA, integer) \rightarrow UA$
$\Rightarrow$	qualification by logical condition	$(condition, UA) \rightarrow UA$

In some cases (due to values of involved variables, and specification to be checked) specification can be feasible with certain parameters of functional processes, but unfeasible with other parameters. The reader can found more detailed description in [7].

Example 1. For the following synchronization requirements: $f_1\ CH\ f_2$; $f_1 \rightarrow f_3$; $f_4\ CK\ f_5$; $f_3 \rightarrow f_4$; $f_2 \rightarrow f_5$, and parameters' values $\tau_1 = 20$, $\tau_2 = 100$, $\tau_3 = 200$, $\tau_4 = 10$, $\tau_5 = 50$, the specification is not feasible due to violation of $f_2 \rightarrow f_5$ requirement (this fact is obvious when we look at Fig. 7).

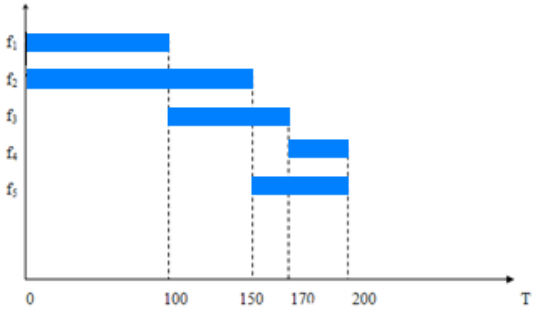


Fig. 5. Example of feasibility checking

But if we have the another parameters, for example, $\tau_1 = 100$, $\tau_2 = 150$, $\tau_3 = 70$, $\tau_4 = 10$, $\tau_5 = 50$, specification becomes feasible (see Figure 8).

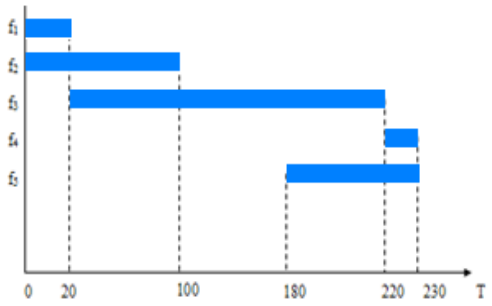


Fig. 6. Example of feasibility checking

The very important point is that this model can be applied not only for real-time spacecraft's flight control software (domain where it was initially developed), but for representation of any sort of activity/processes performed by human, robots, various mechanisms, etc. In other words, the presented model is invariant to nature of performer. But at the same time, the model has enough expressive power for adequate representing of Real-Time control logic's complex features in 'time space' and 'logical space'.

B. *Ways of utilization of SMT solvers functionality*

It is not a wonder that the fundamental mathematical objects such as integers, rational and real numbers, vectors, and matrix are supported by existing SMT solvers by default. Consequently, if we will know how we can transform requirements applicable to control logic into requirements applicable to mentioned objects, then we have possibility to utilize functionality of available SMT solvers.

To do this, we use the following transition from relations between functional processes described as formulas of RTCL, to equations and inequalities on numbers.

TABLE II
ASSOCIATIONS BETWEEN THE CONTROL LOGIC REQUIREMENTS AND INEQUALITIES AND EQUATIONS WITH NUMBERS

RTCL formulae	requires	comment
$f_i\ CH\ f_j$	$t_i = t_j$	equation of numbers
$f_i\ CK\ f_j$	$t_i + \tau_i = t_j + \tau_j$	equation of numbers
$fi \rightarrow f_j$	$t_i + \tau_i = t_j$	equation of numbers
$3A(fi, fj, \delta)$	$t_i + \tau_i + \delta = t_j$	equation of numbers
$H(fi, fj, \delta)$	$t_i + \delta = t_j$	equation of numbers
$fi < fj$	$t_i < t_j$	inequality of numbers
$fi << f$	$t_i + \tau_i < t_j$	inequality of numbers
$f_i <> f_j$	$t_i + \tau_i < t_j \vee t_j + \tau_j < t_i$	disjunction of inequalities
$f_i <l> f_j$	set of boolean equations	logical incompatibilities of FPs (see above)

C. *SMT based Software Tool Prototype*

Some of free SMT solvers provide API for calling them from user software. Some of them, for instance, Z3, accessible through Internet, the user can online specify required or unwanted properties using smt-lib language. Using this opportunity, we tried to apply functionality provided by SMT solver, for control logic checking. For this purpose, the software tool prototype was developed. Using the prototype, we have successfully validated prospectiveness of this approach.

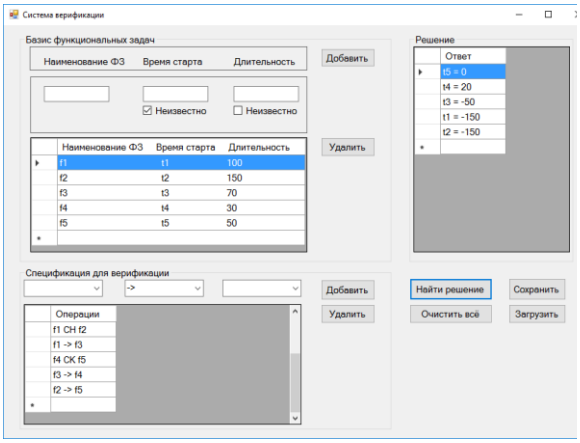


Fig..7. Screenshot of developed software tool prototype

Example 2. The software prototype coded in Java 8, it has intuitively understandable ease user interface. The screenshot is presented in Fig. 3 (interface uses Russian).

First, user sets values of model variables in corresponding input fields. Then he needs step-by-step input specification to be verified. For this purpose, graphical user interface elements allow choosing operations of RTCAL logic. Transformation of specification represented in this from, into SMT solver smt-lib language, is being performed automatically. There are also buttons for trying to evaluate of feasibility, saving and loading of given specification, etc. The result from Z3 in form 'sat' or 'unsat' is decoded, and if the specification is feasible, parameters' values are shown in special window.

D. Utilization of Logic Programming

Follow the monograph [4], let us try to analyze possibility of application of logic programming system's power for our purposes. It is well known that, for example, Prolog logic programming system is equipped with internal logical inference machine. Moreover during looking for answers to user specified questions, Prolog automatically finds values making answer positive (if inference machine system cannot found appropriate values, it returns answer 'No').

This feature provides us with a chance to use Real-Time Control Logic formalism in couple with logic programming system similarly we did it with SMT solvers. The operators of RTCL formulas have to be transformed into Prolog language predicates. The semantics of RTCL can be simply introduced in Prolog terms due to ease using of lists which are main data structure in Prolog language. It is convenient for us due the semantics of RTCAL is formed by tuples which can be reviewed as some equivalent to lists. The logical vector, due to its nature, can be simply represented by list as well. So, we can integrate the specially developed Prolog pre-built program module with specification to be verified, also presented in Prolog language. Then we specify the parameters of basic functional processes of control algorithm, using Prolog language. Doing this, we use variables for unknown parameters. After that, we formulate question (goal) for Prolog system, and get the answer, including values of unknown parameters allowing specification to be feasible.

Example 2.
with Prolog input
CK(f5,f3).
CK(f4,f3).
begin_time(f5,10).
duration(f3,50).
duration(f5,90).
and goal ?CK(f3,X), user gets the answer 'Yes' and values X=f3, X=f5, X=f4, with the goal ?begin_time(f3,X). user gets 'Yes, X=50'.

III. CONCLUSION

We have shown how the algebraic and logical based models of real-time control logic can be applied for feasibility checking using Satisfiability Modulo Theories solvers. The Real Time Control Logic presented in the paper, is a product of evolution of ideas formulated by A.A. Kalentyev in his early formal calculus of control algorithms, and algebra of real-time control algorithms. The proposed approach uses transformation of formal specification represented in terms of RTCL formulas, into equations and inequalities with integers. Then we convert them in SMT solver compatible smt-lib language. The paper presents prototype of software tool supporting the approach and based on calling Z3 SMT solver through its application programming interface.

ACKNOWLEDGEMENT

Author wants to gratefully acknowledge Professor Anatoly Kalentyev who introduced him into world of science and particular area of space technologies and who found basis of Real Time Control Algorithm's Logic.

REFERENCES

[1] W. Ashby, *Introduction to Cybernetics*, – New York, Chapman & Hall, 1956.
[2] R.N. Akhmetov, V.P. Makarov, A.V. Sollogub, "Principles of the Earth Observation Satellites Control in Contingencies", Information and Control Systems, vol. 1, pp 16-22, 2012.
[3] A.A. Tyugashev, "Integrated environment for designing real-time control algorithms", Journal of Computer and Systems Sciences international, vol. 2(45), pp. 287-300, 2006.
[4] A.A. Kalentyev, A.A. Tiugashev, *Application of CALS technologies in Lifecycle of Complex Control Software*, Samara, Samara Centre of RAS Publishing, 2006 (in Russian).
[5] A. Tyugashev, I. Ilyin, I. Ermakov, "Ways to improve quality and reliability of software in aerospace industry", Large-Scale Systems Control, pp. 288–299, vol. 39, 2012. (in Russian).
[6] A. Tyugashev, "Use of graph-based and algebraic models in lifecycle of real-time flight control software", Proceedings of Mathematical Modeling Session at the International Conference Information Technology and Nanotechnology (MM-ITNT 2017) pp 306-311, Samara, Russia, 19.11.2017.
[7] A. Tiugashev, "Build and evaluation of real-time control algorithms in case of incomplete information about functional processes' parameters", Proceedings of 2017 XX IEEE International Conference on Soft Computing and Measurements (SCM'2017), pp. 179-185, Saint Petersburg, Russia, 2017.

CEEPUS WORKSHOP “MODELLING, SIMULATION AND COMPUTER-AIDED DESIGN IN ENGINEERING AND MANAGEMENT”

Power consumption evaluation of FPGA devices based on I/O and clock frequency

Marsida Ibro[1] and Luan Karçanaj[2]

Abstract – **In this paper, models to estimate and evaluate power consumption on FPGA devices will be introduced. FPGA-based electronic systems are becoming more attractive due to reconfigurability and high performance. Due to better fabrication techniques, nowadays FPGA elements contain many programmable blocks CLB (Configurable Logic Blocks) and other functionalities such as memory chip (on-chip memory), DSP blocks, clock synthesizer etc. But before the design of complex electronic systems is more valuable to make evaluations and estimations on power consumption. In order to make the right evaluation we have taken into consideration the internal resources like I/O and clock frequency, due to the complexity and design of such architectures. To achieve these goals, we will take into account other factors and results will show that the power consumption is reduced.**

Keywords – ***FPGA, Power consumption, Power consumption estimation, Power consumption models, I/O and Operation frequency***

I. Introduction

Field programmable gate arrays (FPGAs) provide an alternative solution for intensive calculations in digital signal processing applications (DSPs). The FPGA structure consists of two main components: logic blocks that implement the combinatorial part of on-chip memory design and memories. Logic blocks include Look-Up Tables (LUT) and information storage elements. These two elements found in configurable logic blocks (CLBs) make the FPGA architecture ineffective.

FPGA devices can be used to implement many applications with high throughput, but high performance and power consumption remain today's challenge for electronic systems, mainly on mobile devices. In this paper, methods that help estimate power consumption based on I/O (input/output) and operation frequency will be presented. In FPGA elements we need to take in consideration dynamic power which is the power consumed only when the transistors are working or not working (idle). Dynamic power loss is proportional to the number of transistors.

Due to the requirements for designing DSP systems, programmable logic devices are becoming very necessary. Programmable logic devices, such as FPGAs, offer high performance and therefore are faster than the traditional microprocessors. The reconfigurability of FPGA devices requires the usage of many internal hardware components.

[1]Marsida Ibro is with the Faculty of Information Technology at "Aleksandër Moisiu" University of Durrës, L1, Rruga e Currilave, Durrës, Albania, E-mail: marsidaibro@uamd.edu.al

[2]Luan Karçanaj is with Professional College of Tirana, "Xhanfize Keko", Kompleksi Xhura, 1021 Tirana, Albania

To implement a logic circuit on FPGA devices we need to use more transistors compared to ASICs (Integrated Application Specific Specifications). This leads to higher power consumption in FPGA and this is one of the areas where ASIC are superior to FPGA. As FPGA devices have higher power consumption, today is one of the main areas of scientific research.

In order to optimize power consumption of FPGA devices we can use CAD (Computer-Aided Design) architectures and circuits. The main focus for manufacturers and scientific researches related to FPGA and CAD, is to improve efficiency and performance [1, 2, 3].

Dynamic power is consumed as a result of the toggle rate in logic circuit signals during operation. These transitions increase not only in proportion to logic transitions activity of the signals in the circuit but also the charging / discharging of capacitors. Power leakage occurs even when circuits are in the idle state. Power consumption optimization and reduction of FPGAs offers many advantages because most of the today's applications are portable or battery-powered and mobile applications have limited standby power in some μA.

II. FPGA Architecture

FPGAs are built-in configurable circuits that can be used to design digital circuits. FPGAs configuration is done with VHDL (VHSIC Hardware Description Language) or Verilog programming languages. FPGAs contain configurable logic blocks that provide reconfigurability and programmable interconnects hierarchy is shown in Fig.1.

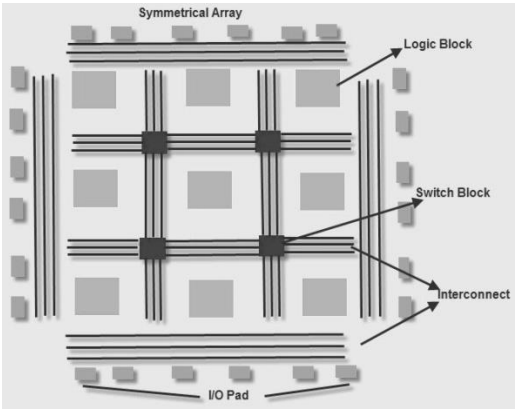


Fig. 1. FPGA architecture

The trends in terms of FPGA architectures technology comprise components such as DSP blocks, processors, and high-speed transceivers to realize the System On a Programmable Chip (SoPC).

In FPGA, connections between elements are realized by elements such as SRAM (Static Random Access Memory), flash cells or EEPROMs (Electrically Erasable Programmable Read-Only Memory) that directly connect components via interconnections.

III. Related Works

Nowadays, reducing power consumption and dynamic power leakage is a challenge on FPGA-based circuit design. One of the most effective techniques is to use partial reconfigurability. FPGA power consumption can be optimized even at the architectural level. The pipeline stages can reduce power consumption by about 25% to 40% [4, 12].

Reducing power consumption due to reconfigurability in FPGA-based DSPs is done by reducing unnecessary calculations from internal resources [5, 6]. Since FPGA architecture uses CMOS (Complementary Metal-Oxide-Semiconductor), a three level architecture is proposed for transistors. Also, different logic blocks structure using 4-LUTs with 3-inputs with multiplexers and DFFs (D flip-flop) can be used [7, 8, 9, 10].

IV. Methodology

To achieve accurate predictions, probability techniques are used, which are based on the conditions of transitions activity estimation for logic circuits. To reduce power consumption, following internal components have been taken into consideration like logic cells (LUT, flip-flops, and registers), interconnections, I/O cells, clock frequency and memory. The method used to measure the power consumed by any internal component in FPGA is:

1. Initially, we take the simplest case with the minimum number of elements, fixed toggle rate (12.5%) and constant room temperature.
2. While considered internal elements are variable, other components are constant to view the direct impact of each FPGA internal component. This procedure is accomplished using simulators software`s of each manufacturer.
3. The first element to be measured is interconnection because is used to make logic or I/O connections.
4. Finally, the procedure is repeated for each element. The measurement results from the proposed models are compared with the models of the manufacturers and will help to understand which components should be considered for reducing power consumption.

After using our methods, we will compare the results with those of the manufacturer's models as shown in Fig.2. To make the measurements we have used Spartan 6 kit board and Xilinx ISE Design tool. Xilinx Spartan-6 FPGA (XC6SLX9-TQG144) has 9152 logic elements, Onboard I2C serial EEPROM 24C04, MAX232 serial communication transceiver and 50 MHz Oscillator.

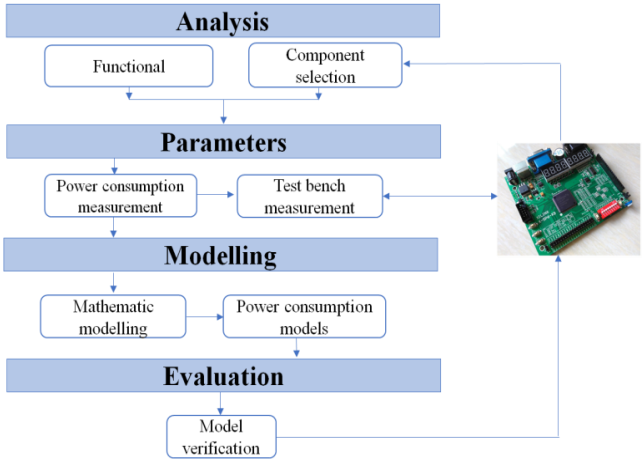


Fig. 2. Evaluation and measurement methods

V. Proposed Models for Power Consumption

Now, we will introduce commonly used methods to calculate, and estimate power consumption. These methods usually use probability estimates for glitch detection and then apply the glitch reduction method using flip-flop with negative triggering impulses. Before we define the models for power consumption will use the vendor measurements shown in Fig. 3.

As mentioned before, delays caused by logic gates can be risky when this function is used to control clock impulses. Most of the models used to explain power consumption for integrated IC circuits are based on equations derived from the CMOS inverter analysis [11].

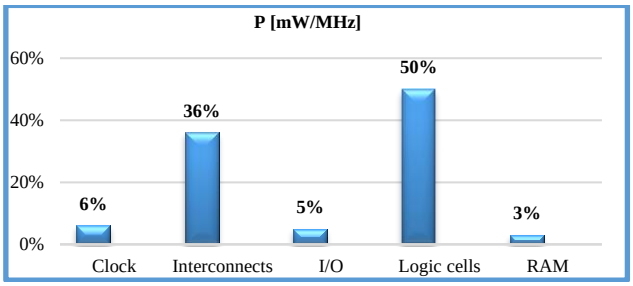


Fig. 3. FPGA components power consumption measured by vendors

In Eq. 1 is calculated the saturation current in PMOS (p-channel MOSFET) and NMOS (n-channel MOSFET) transistors:

$$I_{DSp} = -\beta_p \left[(V_{IN} - V_{DD} - V_{Tn})(V_o - V_{DD}) - \frac{1}{2}(V_o - V_{DD})^2 \right]$$
$$I_{DSn} = \beta_n \frac{(V_{IN} - V_{Tn})^2}{2} \quad (1)$$
$$\beta_p = k_p \frac{W_{eff}}{L_{eff}},\ V_{GS} = V_{IN} - V_{DD},\ V_{DSp} = V_o - V_{DD}$$

where W_{eff} is channel effective width, L_{eff} is channel effective length, k_p is a parameter that depends on process. This factor is calculated $k_p = \mu C_{ox}$, where C_{ox} is oxide capacitance for unit length and $C_{ox} = \varepsilon_0 / l_{ox}$, where l_{ox} is oxide width of CMOS gate. V_{Tn} is threshold voltage for MOSFET transistor, V_{GS} is gate-source voltage, V_{DD} is supply voltage in

drain terminal and V_{IN} is input voltage. In Eq. 2 is calculated the static power:

$$P_{static} = P_{leak} + P_{dp} \quad (2)$$

Dynamic power is calculated by the equation Eq. 3:

$$P_d = \frac{C_L \cdot V_{DD}^2}{T} = C_L \cdot V_{DD}^2 \cdot f \quad (3)$$

where f is clock frequency.

Dynamic power consumption and I/O power dominates overall total power consumption on FPGA. Often FPGAs designs require usage of higher clock frequency f. I/O commutation for high speed data and logic switching for larger frequencies became the dominant factors in power consumption for FPGA elements [12]. To effectively reduce total FPGA power, both static and dynamic power must be addressed while ensuring the FPGA's performance still meets design requirements. Xilinx mathematical model for power consumption is:

$$P = (M \cdot C \cdot tog_{LC} \cdot f_{max} \cdot P_{CLB}) + (G \cdot f_{max} \cdot P_{CLK}) + (L \cdot tog_{LC} \cdot f_{max} \cdot P_{LL} \quad (4)$$

where M is the number of flip-flops per CLB, C is the number of CLBs, tog_{LC} is toggle rate, f_{max} is maximum clock frequency, P_{CLB} is power to frequency ratio of the CLB, G is the number of global clock, P_{CLK} is power to frequency ratio of the global clocks, L is the number of long lines and P_{LL} is the power to frequency ratio of the long lines.

Dynamic power consumption source is the current needed for load capacity LC charging/discharging. The average dynamic power P_d required to load capacitance charging /discharging LC during clock period T is given in Eq.5:

$$P_d = \frac{1}{T}\int_0^T i_o(t) \cdot v_o(t)\, dt \quad (5)$$

An extra power loss is caused due to the short circuit current. This current depends on the rising edge and input voltage drop. If we assume that the rising and falling edge are equal, the power consumed by the short-circuit current is given in the Eq.6:

$$P_{SC} = I_{mes} V_{DD} \quad (6)$$

Another equation relates short circuit power for short circuit is also proportional to the frequency of operation. Short-circuit power consumption is reduced by about 20% of the total power consumption and will be given in Eq.7:

$$P_{sc} = \frac{\beta}{2}(V_{DD} - 2V_T)^3 \tau f \quad (7)$$

Input/output (I/O) elements allow logic circuits in FPGA devices, to communicate with external devices. I/O circuits are important because can limit the speed of the circuit and power consumed. There are several types of I/O circuits, such as input and output buffers, clock distribution, a clock buffer, and low-swing I/O [13, 14]. For circuits that contain several I/O, dynamic power consumption is calculated in Eq.8:

$$P_{input} = \alpha \cdot N_i \cdot E_{ii} \cdot f \quad (8)$$

where α is transition states, N_i the input numbers, E_{ii} is internal energy in W/Hz and f is operation frequency. Dynamic power consumption by the output circuit is expressed in Eq.9:

$$P_{dinamike} = \alpha\left(N_o E_{io} + N_o C_o \cdot {V_{DD}}^2\right) \cdot f \quad (9)$$

VI. Results and Discussions

A critical element in electronic circuits reliability is that a part of power is transformed in heating during the operation. Thermal characteristics of a circuit depend on equipment and used packaging, operating temperature, and operating current. Although these two factors, the power and time of the delay, are related to each other, differ in function of the power supply voltage. It is therefore very important to predict the performance with the increase of the delay time when the food source is reduced. This is important since the use of food resources at the level. low impact on the reduction of power consumption. Using the ring oscillators, we measure the current, voltage over frequency to obtain the maximum delay when using the lowest possible supply voltage.

Fig. 4 shows the block diagram of the ring oschillator which will be used for Spartan-6 device.

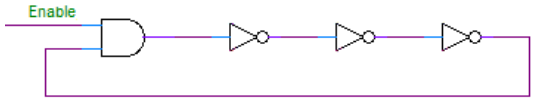


Fig. 4. Block diagram of ring oscillator

To measure the power consumed we follow the steps below: 1) we estimated power consumption of the designed circuit, 2) we calculated the maximum power for the device and 3) packaging and compare it with the estimated power values and with power measured by manufacture simulator. Based on the mathematical models we have used Xilinx ISE Design to make a comparison between our models and the manufacture tools to evaluate power consumption.

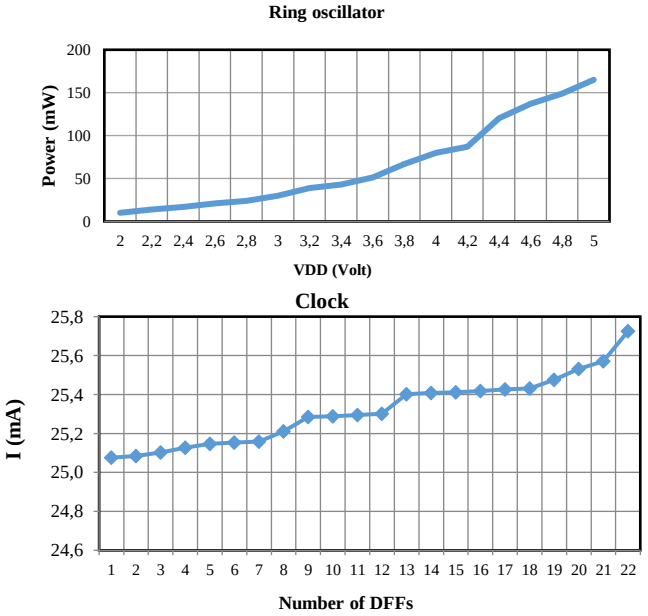


Fig. 5. Power measurements for different supply voltages and Current measurements of the clock signal for different used number of DFFs

After the measurement on the FPGA board, we have compared the result between the two methods. In Fig. 6., we can see the results taken from the methodology used to analyze and measure the power consumption of the main components in the architecture of the FPGA devices.

Fig. 7. shows power consumption of I/O components over the supply voltage. These measurements are done to get a whole picture about the effects of the inputs/outputs and to understand potential consideration in using models for estimation of power consumption.

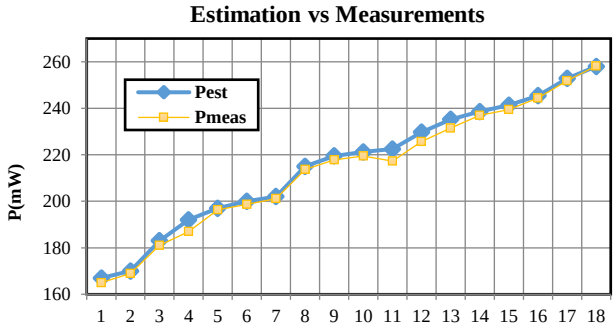


Fig. 6. Comparison of estimated and measured power consumption

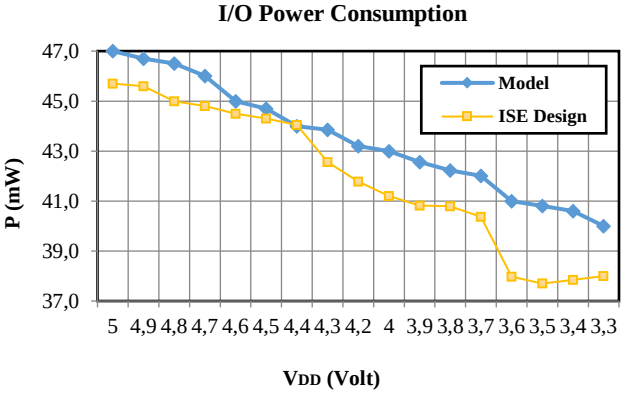


Fig. 7. I/O Power consumption

VII. Conclusion

FPGA architectures are composed of different electronic structures, such as CMOS, Pass-Transistor and SRAM. The power consumption behavior of such devices is much more complex than the power behavior of only CMOS circuits. In this paper, we have take in account, static power, dynamic power and short-circuit currents can be negligible. In FPGAs, power behavior should have a non-negligible DC component due to direct path currents that increase because of the pass-transistor structure, and due to leakage currents, that become important when increasing the number of equivalent ASIC gates. Results show that our measurements on power consumption models are good because they estimate power dissipation and show that power consumption depends on clock frequency and the number of logic cells used to design applications.

References

[1] Zh, Xia, M. Hariyama and M. Kameyama, "A Low-Power FPGA Based on Self Adaptive Multi-Voltage Control", ISOCC, 2013, Conference Proceeding pp. 166-169

[2] Sh. Ishihara, Zh. Xia, M. Hariyama and M. Kameyama, "Evaluation of a Self-Adaptive Voltage Control Scheme for Low-Power FPGAs", *Journal of Semiconductor Technology and Science*, Vol. 10, No. 3, September 2010, pp. 165-175

[3] V. George, H. Zhang and J. Rabaey, "The Design of a Low-Energy FPGA", ISLPED 1999 International Symposium on Low Power Electronics and Design, pp. 188-193, San Diago, USA.

[4] F.G. Wolff ; M.J. Knieser ; D.J. Weyer ; C.A. Papachristou, "High-Level Low Power FPGA Design Methodology", IEEE Conference October 2000, pp. 554 - 559 Dayton, OH, USA

[5] Li Shang, A.S Kaviani, K.Bathala, "Dynamic Power Consumption in VIRTEX-II FPGA Family", Conference Proceedings 24-26 February 2002, 10th International symposium on Field-programmable gate arrays pp.157-164, Monterey, California, USA

[6] J.H. Anderson, F.N. Najm, "Active Leakage Power Optimization for FPGAs", IEEE Transaction on Computer Aided Design of Integtrated Circuits and Systems, Vol. 25, No. 3, March 2006, pp 423 - 437.

[7] Yu Zhou, Sh. Thekkel, S. Bhunia, "Low Power FPGA Design using Hybrid CMOS-NEMS Approach" Low Power Electronics and Design (ISLPED), pp. 14-19, 27-29 August 2007, Portland, USA.

[8] J. Lamoureux. W. Luk, "An Overview of Low-power Techniques for Field-Programmable Gate Arrays", IEEE NASA/ESA Conference on Adaptive Hardware and Systems, pp. 338 – 345, 22-25 June 2008, Noordwijk, Netherlands

[9] Zh. Xia, Sh. Ishihara, M. Hariyama," Dual-rail/single-rail hybrid logic design for high-performance asynchronous circuit", IEEE International Symposium on Circuits and Systems, pp. 3017 - 3020 20-23 May 2012, Seoul, South Korea

[10] Sh. Ishihara, Zh. Xia, M. Hariyama, M. Kameyama, "Architecture of a Low Power FPGA based on a Self-Adaptive Voltage Control", IEEE International SoC Design Conference (ISOCC), pp. 274–277, 22-24 November 2009, Busan, Korea (South).

[11] K.S. Han, D.I. Jeon, K.S. Chung, "Ultra Low Power and High Speed FPGA Design with CNFET", International Symposium on Communication and Information Technologies (ISCIT), pp. 828–833, 2-5 Oct. 2012, Gold Coast, QLD, Australia

[12] S.A. Raja Ram, D. Jerish Solomon, "Reducing Power in Reconfigurable Processors using Dual V_{DD}", International Journal for Research and Development in Engineering (IJRDE), ISSN: 2279-0500, pp- 211-216, 2014.

[13] P.A.Kamble, M.B. Mali, "Low Power Techniques for High Speed FPGA", International Journal of Scientific and Research Publications (IJSRP), Volume 5, Issue 3, March 2015.

[14] B. Sankar, Dr. C.N. Marimuthu, "Power Optimization using Body Biasing Method for Dual Voltage FPGA", International Research Journal of Engineering and Technology (IRJET), e-ISSN: 2395 -0056 Volume: 02, Issue: 01, April 2015.

Lab-Based Learning on Wireless Network QoS Case Study: WSN for real time environmental monitoring

[1]Zhilbert Tafa, [1]Artan Salihu, [1]Blerona Cakolli, [1]Edmond Hajrizi

Abstract –**With a growing number of wireless applications for real time continuous event monitoring, the concept of Internet of Things (IoT) is becoming reality. However, in contrast to the fixed wired networks, the performances of wireless networks are much more unstable. In the scope of the advanced computer networks and wireless communication courses, the graduate computer science students are taught on network performance analytics and wireless sensor networks, respectively. This paper presents a project based approach of researching on both topics. Specifically, it brings up the synergic effects on multidisciplinary learning by evaluating the performance of a wireless network for environmental (greenhouse) monitoring. For this purpose, a typical wireless sensor network for medium-size greenhouse is simulated by using OPNET network modeller.**

While the educational effect was fully achieved, the simulations show satisfying network performance for the given purpose. The results also show the influence of the network settings (such as network topology, the wireless transmission distances, and the number of nodes) on Quality of Service (QoS) parameters.

Keywords –**Greenhouse, WSN, ZigBee, QoS, environmental monitoring, learning.**

I. INTRODUCTION

The IoT involves various information and communication technologies towards enabling the efficient human interaction with the environment. An IoT module/network collects the information from the surrounding objects with the aim to further process it and/or visualize it to the users.

Wireless Sensor Network (WSN) is the technology of choice regarding the IoT data acquisition infrastructure. It is comprised of a number of small battery supplied nodes that sense the information, locally process it, and wirelessly transmit it to the user and/or data centre. The critical part of a WSN, i.e., its greatest power consumer, is the Radio Frequency (RF) transmission module. Hence, choosing the right wireless technology for long term power-autonomous monitoring system is of crucial importance. ZigBee, a personal-area communication technology, is becoming a standard for low power wireless transmission at the 2,4 GHz unlicensed ISM band. The ZigBee transceivers are designed to have low cost, low power consumption, high security, and low data rates.

[1] The authors are with the Faculty of Computer Science and Engineering at UBT, Lagjia Kalabria, Prishtine 10000, Kosovo, E-mails: zhilbert.tafa@ubt-uni.net; artan.salihu@ubt-uni.net; bleronacakolli@gmail.com; ehajrizi@ubt-uni.net.

The technology supports for three topologies. Star and tree topologies are used in local, small to medium coverage networks, while mesh topology is used to cover greater areas in the multi hop manner. A ZigBee node can be a sensor, a router or a coordinator. Each of them contains microcontrollers and the RF transceivers, with some of them containing the sensing circuits as well. A sensor node transmits data from the sensors towards the router or the coordinator. A router simply routes the traffic, and is included only in tree and mesh topologies. The coordinator synchronizes the whole network and presents the central part of the topology.

An emerging WSN application area is the greenhouse monitoring systems. Previous systems of this purpose were based on the cable infrastructure, usually with a single sensing node at the centre of the greenhouse. The actual greenhouses are bigger in size and hence more measurement points are needed. With the development of the embedded systems and low power wireless technologies, the areas are now covered with the greater number of nodes and the bulky cabling has been replaced by wireless links. Depending on the various factors (such as sampling rate and transmitting/sleeping duty cycle, the used technologies and protocols, node-to-node distances etc.) today's wireless sensor nodes can last on batteries for weeks, or even years. As such, modern greenhouses use WSNs as a technology of choice for real time and continuous monitoring of the parameters such as air humidity and temperature, soil humidity and temperature, light intensity and concentration of the CO2.

To meet the application requirements, the wireless networking infrastructure should enable for the timely, accurate and reliable transmission of the sensor data. An appropriate traffic analytics can lead to the improvement on the network settings. A number of efforts has been made towards the QoS parameter evaluation for the greenhouse networking scenarios [1]-[5].

The paper presents a lab-based QoS analysis of a WSN for greenhouse monitoring by using OPNET modeller [7]. The motivation for the study is two-folded:

a) The academic/engineering argumentation - traffic analytics of the greenhouse monitoring network can lead to a better understanding of the networking problems and, consequently, to the network performance improvement.
b) Synergic multidisciplinary education effect - the OPNET is a powerful tool for network performance analytics in education [6]. Even though its support for the WSN simulation might be considered as a bit complex and incomplete; learning it from a realistic scenario will help students applying the QoS analytics on the other networking technologies and topologies.

Accordingly, the primary objective of the study is the performance analysis of a typical mid-size real time greenhouse monitoring system. The study is conducted in scope of the project that envisions students to learn by research. As such, another objective of the study is in the domain of education. Namely, in a research-oriented environment, students learn on how to deploy and set up the WSN, and how to analyse the network performances. On top of that, they are also able to understand the importance of each parameter as well as the factors that influence them in a WSN setup.

In order for the goals to be achieved, the theoretical background on WSNs, ZigBee, OPNET simulation framework and QoS are priory presented to the students in scope of the advanced computer networks and wireless communications courses.

The paper is organized as follows. After introducing the topic and the problem, Section II presents the simulation setup and details. The results are presented in Section III, while the Section IV concludes the paper.

II. THE SIMULATION ENVIRONMENT AND SETUP

A. The rationale behind OPNET for WSN in education

OPNET is a commercial general-purpose network simulator which is recently re-branded as a Riverbed Modeller by Riverbed Technologies [7]. In this paper, OPNET Modeller Academic Edition 17.5 was used. It supports design and modelling of communication systems as well as end-to-end system performance assessment by discrete event simulations. In addition to a wide range of ready to use models and functions, it provides a user with a Graphical User Interface (GUI) for node and network configuration, as well as for defining the underlying processes. Its graphical presentation capabilities make it easier to practice the network quality parameters, hence it is considered to be a more attractive and user friendly tool for students. It allows them to focus on results and understanding of performance evaluation metrics for WSN systems.

Other state of the art tools for simulation and QoS analysis of WSNs used by researchers and educators include ns-2 [8], OMNET++ [9] and TOSSIM [10]. However, ns-2 and OMNET++ require additional add-ons and modifications to provide reliable results for different WSN applications. Also, code in ns-2 is relatively difficult to manipulate with a poor graphical presentation capability. TOSSIM offers scalability and is specifically designed for WSNs prototyping, but it is limited to simulations of TinyOS applications. Simulations of protocols require writing TinyOS application, making it hard to troubleshoot and abstract for laboratory exercises. A comprehensive comparison of limitations together with advantages of WSN simulators is given in [11], revisited by a recent work in [12].

With the rapid increase of WSN applications and demand for hands-one skills in job market on the one hand, and laboratory equipment costs and space limitations on the other hand, OPNET provides an ideal tool to bridge the gap between theory and practice for teachers, practitioners and higher education institutions in general. The authors in [13] describe the effect of hands-on teaching using OPNET in a wireless communications undergrad class. By designing five simulation laboratories, they show how easy and beneficial to students is designing, deploying and assessing performance of complex and ever-changing wireless infrastructures. Based on surveys and student feedback, the authors in [14] claim the effectiveness of using OPNET simulation tool for teaching networking concepts. Participated students in this study show a significant increase in gaining both theoretical as well as practical skills in creating and analysing different networking scenarios.

We complement previous work by providing researchers and teachers a baseline of ZigBee implementation and performance analysis in OPNET. Thus, it can be easily replicated and extended to other case studies. Consequently, enabling students to practically understand other Low-Power Wide Area Network (LPWAN) technologies, its applications and last yet important appropriate QoS evaluation techniques one should choose.

B. The simulation scenario and setup

In order to observe the network QoS parameters, students consider a realistic environmental monitoring scenario, precisely a medium size 150mx200m rectangular greenhouse area. Although there are lots of ZigBee products in the market that are declared to cover much greater ranges (such as, for example, XBP24-API-001), we have adopted the theoretical limitations of the low power technology regarding the indoor and outdoor ranges. Therefore, the number of nodes is chosen accordingly to the field dimensions and general ZigBee range specifications (that claim the outdoor range of few tens of meters to 100m). Students set up the network by configuring one coordinator, two routers and 4 sensor nodes. The network topology is shown in Figure 1.

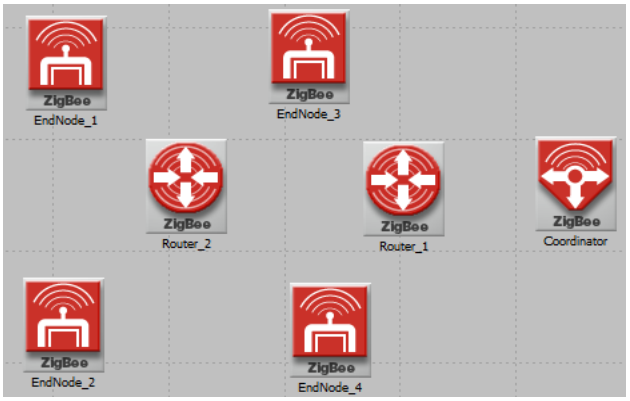


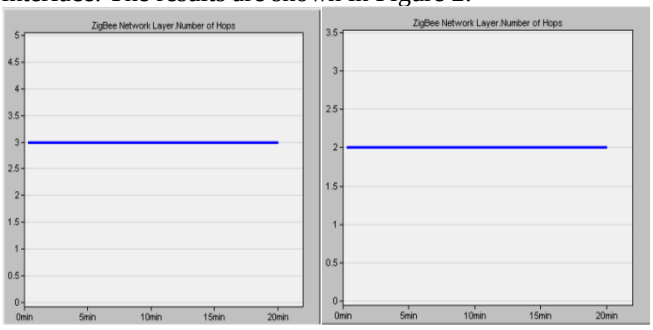
Fig. 1: Simulation setup

The coordinator is set to be data centre or to send data further through the Internet towards the end user. It sets the network in mesh-routing mode. The end-nodes are configured

to send data to the coordinator, i.e., with the coordinator set as the destination node. The routers are configured for the ad-hoc random connection manner. As such, the end nodes 1 and 2 are connected to the Router 2, while end nodes 3 and 4 are connected to the Router 1. Router 2 transmits data through Router 1 towards the coordinator.

III. THE SIMULATION RESULTS

The network parameter analytics were conducted at the end nodes 1 and 4, since they transfer through the different routers and number of hops to the destination. The simulation shows results for 20 minutes of the data transmission. To show the ad-hoc connection nature of the ZigBee network from the given topology, based on the nodes' distances, the numbers of hops towards the destination were firstly evaluated. The results were obtained from the network layer simulation interface. The results are shown in Figure 2.



(a) (b)

Fig.2:Number of hops for: (a) EndNode_1 and (b) EndNode_2

The average end-to-end delay were measured from the application point of view, i.e., from the application simulation interface. The results for end nodes 1 and 4, are shown in Figures 3 and 4, respectively.

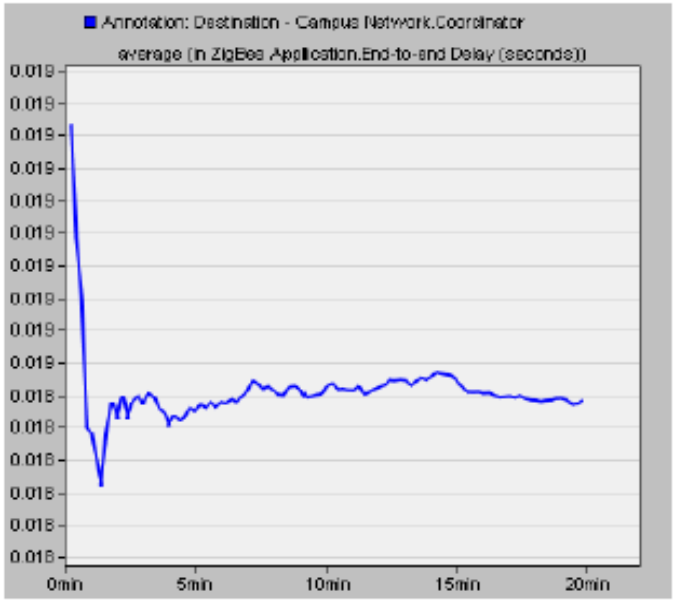


Fig. 3: End-to-end delay at EndNode_1

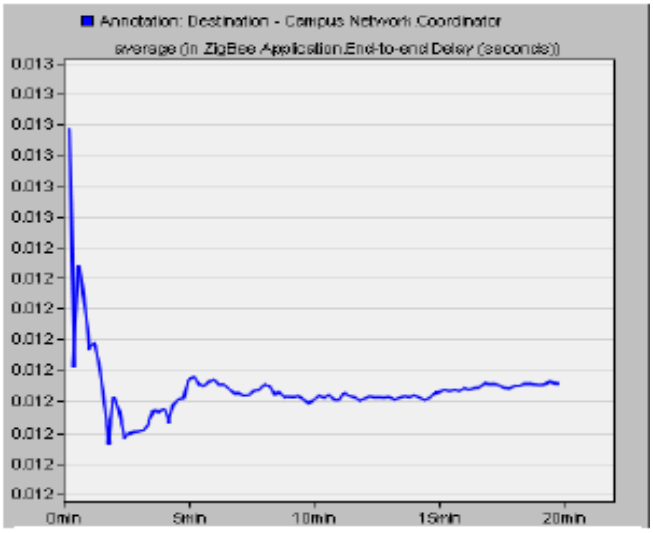


Fig. 4: End-to-end delay at EndNode_4

The results show that the average delay is higher at the beginning of the nodes' transmission. This can be explained by the on-demand nature of the routing protocol, and the initial route learning that happens at the beginning of the transmission. Also, as expected, when the delays get stabilized, the average delay of the EndNode_4 is lower than the average delay of the EndNode_1. But, this difference is not crucial.

Same as the delay, throughput depends on many parameters such as number of nodes, the network topology, the collision occurrence, etc. Throughput at the end nodes 1 and 4 were evaluated through the link layer simulation interface. The results are shown in figures 5 and 6.

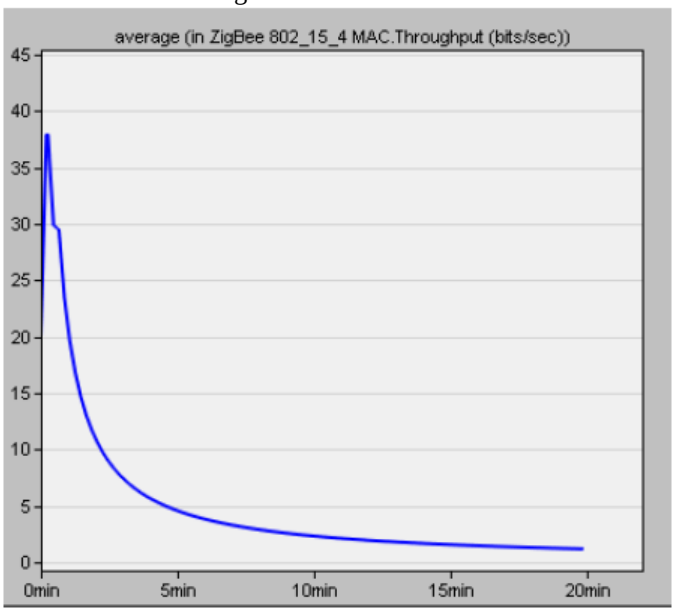


Fig. 5: Throughput at the EndNode_1

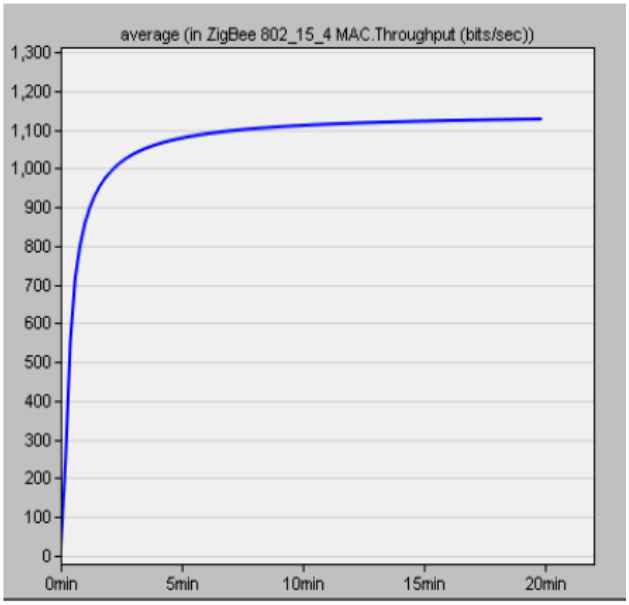


Fig. 6: Throughput at the EndNode_4

It can be noted that the number of hops, the position of the node in the topology, and the physical distance of the node from the destination, greatly influence throughput. The EndNode_4, which is much closer to the destination, reaching it with 2 hops, shows much greater throughput as compared to the EndNode_1. For the purpose of experimenting, the EndNode_3 and EndNode_4 were temporarily removed from the network. In this case, the throughput at EndNode_1 has improved significantly, but still hasn't reached the values of the EndNode_4.

IV. Conclusions and Future Work

The paper presents a QoS analysis of the WSN-based greenhouse monitoring system. Its workflow and methodology follows the concept of learning through the research projects. The work has few main outcomes. From the educational point of view, based on the tests and the research goals' fulfillment measure, the educational effect can be considered as fully achieved. On the other hand, from the simulation results, it can be observed that the number of nodes and the node-to-node distance greatly impact throughput while they don't influence the average delay that much. However, it can be concluded that the given QoS parameters would meet the network requirements for the purpose of a greenhouse monitoring system.

Future work will be focused on QoS analytics for intermediate nodes (routers) as well as on the real implementation of the WSN for greenhouse monitoring by using off-the-shelf components.

References

[1] M. Keshtgari, and A. Deljoo, "A Wireless Sensor Network Solution for Precision Agriculture Based on Zigbee Technology", Wireless Sensor Network, vol. 4, no. 1, pp.25-30, 2012.

[2] I. Lamprinos, and M. Charalambides, "Experimental assessment of Zigbee as the communication technology of a wireless sensor network for greenhouse monitoring" International Journal of Advanced Smart Sensor Network Systems (IJASSN), Vol 5, No.3/4, October 2015

[3] S. Vançin, and E. Erdem, Design and Simulation of Wireless Sensor Network Topologies Using the ZigBee Standard, International Journal of Computer Networks and Applications (IJCNA) Volume 2, Issue 3, May – June 2015

[4] J. Jiang, C. Wang, M. Liao, X. Zheng, J. Liu, C. Chuang, C. Hung, and C. Chen, "A wireless sensor network-based monitoring system with dynamic convergecast tree algorithm for precision cultivation management in orchid greenhouses", Precis. Agric., pages 1–20, 2016.

[5] I. Hoque, "Modelling and performance study of large scale Zigbee based Greenhouse monitoring and control network", Master. Massey University, (2013).

[6] D. Akbas, H. Gumuskaya, "Real and OPNET modeling and analysis of an enterprise network and its security structure", Procedia Computer Science, 3, pp. 1038 – 1042, 2011.

[7] OPNET official website, https://www.riverbed.com/gb/products/steelcentral/opnet.html.

[8] NS-2 official website, https://www.isi.edu/nsnam/ns/

[9] OMNET++ official website, https://www.omnetpp.org/

[10] Levis, P., Lee, N., Welsh, M., and Culler, D. Tossim: Accurate and scalable simulation of entire tinyos applications. In First ACM Conference on Embedded Networked SensorSystems, 2003

[11] M. Korkalainen, M. Sallinen, N. Karkkainen, and P. Tukeva. "Survey of Wireless Sensor Networks Simulation Tools for Demanding Applications". The Fifth International Conference on Networking and Services, 2009.

[12] I. Minakov, R. Passerone, A. Rizzardi, and S. Sicari, A comparative study of recent wireless sensor network simulators, ACM Trans Sens Netw 12, 2016.

[13] T. S. Chou, P. Lunsford, and K. Thomson, "Design of Simulation-Based Laboratories for Teaching Wireless Network Technologies," The Conference for Industry Education Collaboration, Palm Springs, CA, February 2015.

[14] Sh. Hublikar, and A. L. Kakhandki, and A. M. Kabbur. "OPNET: a Solution for Software and Hardware Networking." Journal of Engineering Education Transformations,2016

Combined Approach to Modelling and Simulation in Engineering Education

Maja Atanasijević-Kunc[1], Vito Logar[1], Matevž Pogačnik[1], and Vassil Guliashki[2]

Abstract – **Modelling, simulation, optimization, and control design of dynamic systems are very well established scientific fields which are of crucial importance for engineering education. The paper presents some activities realized through the CEEPUS project cooperation where several countries are involved, among others also Slovenia and Bulgaria. Common project activities include also international summer schools where students and professors from participating countries cooperate through education process. Some experiences and also e-learning development support initiated by such cooperation are described in this paper.**

Keywords – **Control Design, Dynamic Systems, Education, E-learning, Modelling, Optimization, Simulation.**

I. INTRODUCTION

Engineering education programs in most cases include also subjects describing different aspects of mathematical modelling, computer simulation, control design of dynamic systems, and optimization. All mentioned scientific fields are of interdisciplinary nature which means that similar or even completely equal approaches and methods can be used for solving problems from different areas. In addition the level can be adapted to suit all tree education levels: bachelor, master, as well as also doctoral (research) study.

Two years ago, the teaching staff from eight central and east European countries gained the so-called CEEPUS project entitled *Modelling, Simulation and Computer-Aided Design in Engineering and Management* [1]. Through the first year 12 universities took active role in starting the cooperation. Project activities enable also the organization of summer (and/or winter) schools. There are several important goals of such events:

- Students from all participating countries can obtain additional knowledge, representing an extension to their normal study program, which is also indicated in the final diploma certificate in most European countries, representing an important additional motivation for their study.
- Better and more active students have better options when searching and applying for employment.
- On the basis of new contacts established through such meetings, some interesting CEEPUS (but also Erasmus, or similar) exchanges can be organized for interested students, but also, of course, for teaching staff.
- Lectures and prepared materials offered to students during summer school can represent also the motivation to prepare the improvement of education process for all teaching staff at their home institutions and/or to enable some sharing environment suitable for different institutions and different study programs. This aspect is becoming more and more important because internationalization of studies is becoming increasingly important for all European countries.
- It is also important to mention social contacts which can be improved through such cooperation and which can motivate also additional scientific cooperation between involved institutions.

All mentioned goals contribute also to one very important aspect of education, namely e-learning, which extends existing teaching methodology to blended learning in such a way that education process is improved with some new and convenient possibilities. In spite of the fact that great improvement was made during the last decade, the optimal solution is still under investigation in different countries as well as in different scientific areas. In the paper some possibilities are analyzed which have the potential to represent an improvement of indicated international cooperation.

The paper is organized in the following way. In the next section education process organization at Faculty of Electrical Engineering, University of Ljubljana (FE-UL), Slovenia is presented first. Then some of the most important properties of e-learning program, which was developed at the FE-UL, are explained. In the section III some important information is presented regarding the first CEEPUS summer school entitled: *International Summer School in Modelling, Simulation and Computer-Aided Design*, which was organized at the FE-UL in 2017 [2]. In section IV, the DSS WebOptim is described as a tool for engineering education developed at Bulgarian Academy of Sciences (BAS). The paper ends with some concluding analytical remarks and plans for future work and cooperation.

II. ORGANIZATION OF EDUCATION PROCESS AT FE-UL INCLUDING E-LEARNING PLATFORM E-CHO

The majority of subjects at both bachelor (university and professional) study programs and also at master study program at FE-UL are organized in the form of lectures and laboratory exercises. At the end of the semester, first written and then also oral exams are organized. In some courses also colloquia are prepared during the semester to intensify

[1]Maja Atanasijević-Kunc, Vito Logar, and Matevž Pogačnik are with the Faculty of Electrical Engineering, University of Ljubljana, Tržaška 25, 1000 Ljubljana, E-mail: maja.atanasijevic@fe.uni-lj.si.

[2]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.

students' work. Very high colloquia's exam grade sometimes means that student has also passed written examinations.

In 2014/15, we have, in addition to the previously mentioned forms, introduced also the e-learning platform E-CHO for all the courses at all education levels. The development of this platform has long been conducted by the Laboratory for Telecommunications at FE-UL. Its efficiency was tested and complemented for the purposes of studies at FE-UL already in previous years [3-5]. Because of many positive experiences further development of the so-called combined or blended learning taking into account new possibilities, is now under investigation. The goals are numerous, but some of the most important are the following:

- gradual subject presentation,
- flexibility and multi-functionality,
- enhancement of motivation,
- step by step improvement of the platform E-CHO regarding specific experiences and arising needs,
- explicit subjects' connections, where possible,
- internationalization of some lectures, courses, or even study programs.

E-CHO is an e-learning Internet-based platform which associates and combines numerous e-learning functionalities. It can also be integrated into other web-based applications. This platform provides e-learning management (LMS – Learning Management System), e-learning content managing (LCMS – Learning Content Management System), tracking of the teacher's activity and progress, knowledge assessment/testing, e-learning standards' support (SCORM and QTI) and multilingualism. During the last few years improved versions of E-CHO were developed where special attention was devoted to the simplification of content development procedures, improvement of progress tracking tools and learning statistics reports, adaptive access – rights management, video and multi – media support, simplified third party integration through open interfaces, new mobile interfaces (tablets, smart-phones), and also improved knowledge assessment (new question types, simplified tests and questionnaires development and delivery).

Each professor and his/hers assistants, as well as all the students of FE-UL, can enter E-CHO through corresponding user name and password information. Starting window is illustrated in Fig. 1. Each professor can develop e-learning contents for the courses he/she is responsible for by clicking the button Classroom (see upper part of Fig. 1). This opens the window of the form as is presented in Fig. 2. In the left part of the window the titles of the subjects are displayed. Below also the access to the short manual is available and some answers to frequently asked questions. Titles of already developed contents are displayed in the next column. It is possible to observe all developed contents or only those which are connected to the certain subject. The user can develop contents in the following forms: file, figure, questions, folder with questions, web-page, url-address, video, different tests, group of contents, course, and video course.

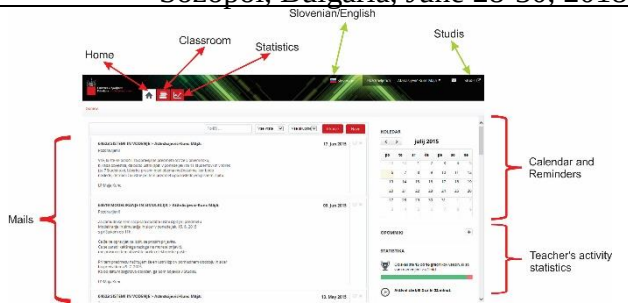


Fig. 1: E-CHO starting window

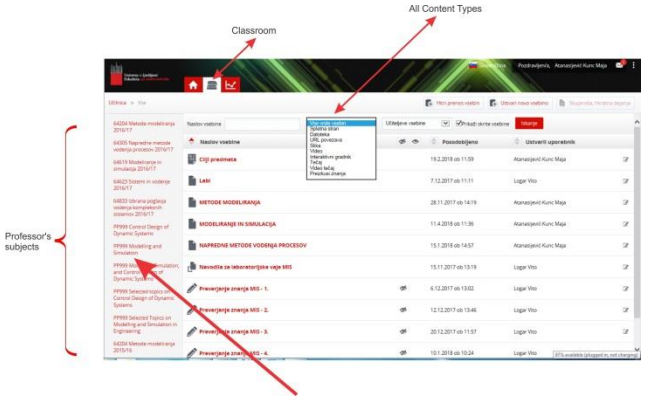


Fig. 2: E-classroom with professor's subjects and developed content elements

All prepared contents can be connected to one or more subjects in corresponding way. During development or during education process the chosen contents can be made visible only for developers. When desired this status can simply be changed and the chosen content is than delivered to the chosen students' group or groups.

It is important to mention that teaching staff and students can develop contents in the same way. Of course, if students are uploading their homework, projects, or exams these materials can be accessed by teaching staff only, but if their project results can represent interesting information to the whole group, it can also be made visible for everyone from the corresponding group. These properties are important because in this way developed contents can also be combined into different course forms and flexibility of education can be increased.

III. INTERNATIONAL CEEPUS SUMMER SCHOOL

As already mentioned, in 2017, the first *International Summer School in Modelling, Simulation and Computer-Aided Design* was organized at the Faculty of Electrical Engineering, University of Ljubljana. All professors who contributed to the mentioned summer school prepared corresponding materials, which were organized into lectures and laboratory exercises. Prepared materials were distributed among all summer school participants (students and teaching staff) in electronic form and were later also published in a paper form [2].

As all materials were prepared in electronic form we have decided to include them also into a new course-unit entitled *Modeling and simulation* as indicated in Fig. 2 because of the following reasons:

- The form of prepared materials includes lectures and laboratory exercises, which are frequently used education forms at FE-UL, but also at many other universities.
- Prepared materials can be accessed by guest professors from all the locations where Internet is available. They can be used inside the subject or separately, regarding certain situation or usage goal.
- Also the students to whom access is permitted can use the prepared materials.
- It is expected that summer school can be organized again and as material is well organized, also further extensions and/or suitable adaptations can be performed in a simpler and more transparent way regarding all the participants.
- Mathematical modelling is frequently connected also with the experiments involving real and/or virtual systems. Such experiments were already developed for some pilot plants and included into the education process at FE-UL through E-CHO platform [5, 6]. Additional similar possibilities are under investigation [7].
- As E-CHO supports also video elements, further summer/winter schools can be active also in this direction.

It is important to mention that one very frequently used activity (in addition to analysis) in modelling as well as in control design of dynamic systems is optimization. This important chapter can be included in many existing courses at FE-UL. But, an interesting possibility for engineering education represents also the so-called WebOptim tool which was developed at Bulgarian Academy of Sciences which has also participated at CEEPUS summer school and is presented in the following section.

IV. The DSS WebOptim as a tool for engineering education

The decision support systems (DSS) can be considered as a tool, contributing to the modern engineering education. A web-based DSS WebOptim (http://weboptim.iinf.bas.bg/) for solving multiple objective optimization problems has been proposed in [8] and developed (see [9,10]). The system WebOptim is designed in a modular principle, extensively using XML as communication standard and web services. The core system module is an original generalized interactive scalarizing method. It includes thirteen interactive methods. Most of the widely known scalarizing approaches (reference point approach, reference direction approach, classification approach etc.) are realized in the scalarizing method. The Decision Maker (DM) can choose the most suitable for him/her form to input his/her preferences: objective weights, aspiration levels, aspiration directions, aspiration intervals. This information could be changed interactively by the DM during the solution process. Depending on the DM's preferences form the suitable interactive optimization method is chosen automatically.

There are possibilities for distance/electronic education by means of WebOptim for students and post-graduate students in different areas: engineers, economists, persons making decisions (decision makers) in the management of enterprises and production processes, as well as all kinds of experts, who solve optimization problems in their activities (see [9]).

WebOptim may be used as a training system for distance learning (e-learning), providing flexible high qualitative online courses in single- and multiple-objective optimization methods. The solvers engaged in WebOptim allow performance of test experiments on pure continuous, pure integer, as well as on mixed-integer optimization problems.

The system WebOptim may be used both as an educational and as a research tool. For this reason it is designed as an open and extensible system. New methods, solvers, problem types and basically every kind of modules can be added at any time. With an Internet connection the students or postgraduate students can participate in online courses from any location in the world. The lecturer in the corresponding course has prepared series of common accessible examples with attached detailed description of the models and the methods, available for solving each example. It is suitable these series of examples to be used as training aid in the different courses. Starting by simple problems and continuing with always more and more complex problems, in this public profile the student can make steps from the fundamentals to advanced subjects. Each student has online access to the course content.

The system quality (see [11]) of WebOptim using 4 groups of criteria can be evaluated as follows: Group I (Learner Interface): Ease of use; User-friendliness; Ease of understanding; Operational stability; Group II (Learning Community): Ease of accessing shared data; Ease of discussion with teachers; Group III (System Content): Sufficient content; Useful content; Group IV (Personalization): Capability of controlling learning progress; Capability of recording learning performance.

V. Conclusion

From the presented the following can be concluded:

- E-learning has become an important aspect of engineering education at many different institutions.
- In spite of the fact that this learning approach has been world-wide adopted many institutions try to develop their own solutions.
- One possible explanation for such situation is the fact that it is also important how different contents are organized into education flow.
- Sometimes also hardware and software available at certain institution are of crucial importance.

- Different teaching approaches using different tools can represent an obstacle in international cooperation where larger number of institutions, teachers, and students should cooperate through such a learning environment.
- It is expected that good experiences in organization of summer schools like the described one inside CEEPUS project can indicate the potential problems but also uncover some good solutions which can also stimulate the development of new and more efficient programs, at least at frequently used activities at engineering field.
- The research of such solutions will also be one of our future goals as it is expected that in this way also our cooperation can be improved.

ACKNOWLEDGEMENT

The summer school and corresponding activities described in this paper were conducted by the project CEEPUS CIII-BG-1103-01-1617 Modelling, Simulation and Computer-aided Design in Engineering and Management.

REFERENCES

[1] http://www.ceepus.info/

[2] F. Breitenecker, N. Muškinja, V. Guliashki, G. Marinova, B. Rodič, G. Karer, E. Hajrizi, I. Perjési-Hámori, R. Svečko, M. Pogačnik, R. Žurbi, V. Logar, G. Maróti, P. Planinšič, G. Mušič, M. Atanasijević-Kunc, *International Summer School study material Modelling, simulation, and Computer-Aided design*, Ljubljana, Faculty of Electrical Engineering, 2017.

[3] S. Čorović, J. Bešter, D. Miklavčič, “An e-learning application on electrochemotherapy”, BioMedical Engineering Online, Vol. 8, No. 26, 2009.

[4] M. Papić, L. Zebec, M. Pogačnik, J. Bešter, M. Atanasijević-Kunc, V. Logar, “Personalized learning environment E-CHO”, EUROSIM Simulation News Europe, Vol. 22, No. 2, 17–24, 2011.

[5] M. Atanasijević-Kunc, V. Logar, R. Karba, M. Papić, A. Kos, “Remote multivariable control design using a competition game”, IEEE Transactions on Education, Vol. 54, No. 1, pp. 97–103, 2011.

[6] V. Logar, R. Karba, M. Papić, M. Atanasijević-Kunc, “Artificial and real laboratory environment in an e-learning competition”, Mathematics and computers in simulation, Vol. 82, No. 3, pp. 517 - 524, 2011.

[7] M. Atanasijević-Kunc, M. Papić, B. Zupančič, J. Bešter, "Combination of learning approaches in modelling and simulation", The 12th International Multidisciplinary Modelling & Simulation Multiconference, Bergeggi, Genova, Conference Proceedings, pp. 314-323, 2015.

[8] K. Genova, L. Kirilov, V. Guliashki, B. Staykov, D. Vatov, "A Prototype of a Web-based Decision Support System for Building Models and Solving Optimization and Decision Making Problems", Proceedings of Papers of XII International Conference on Computer Systems and Technologies CompSysTech’11, ISBN: 978-954-9641-52-3, Wien, Austria, 16-17 June, 2011, pp. 167-172.

[9] V. Guliashki, K.Genova, L. Kirilov, "The Decision Support System WebOptim in an E-Learning Context", Proceedings of Papers of the International Conference "Automatics and Informatics 2013", vol. 1, pp. 117-120.

[10] L. Kirilov, V. Guliashki, K. Genova, M. Vassileva, B. Staykov, **“**Generalized scalarizing model GENS in DSS WebOptim”, International Journal of Decision Support System Technology, vol. 5, issue 3, pp. 1-11, Special Issue from the Decision Support Systems Stream on the EUROXXV Conference in Vilnius, Guest Editors: F. Dargam, S. Liu, I. Linden, 2013.

[11] D. Y. Shee, & Y. S. Wang, **“**Multi-criteria evaluation of the Web-based e-learning system: A methodology based on learner satisfaction and its applications”, Computers & education, 50, 894-905, 2008.

Comparison of Real-valued and Complex-valued convolutional networks for TerraSAR-X Patch classification

Dušan Gleich[1], Danijel Šipoš[1], and Peter Planinšič[1]

Abstract – **This paper presents SAR patch categorization using real and complex valued (CV) deep Convolutional Networks (CNN) for categorization of Synthetic Aperture Radar (SAR) data. The categorization of Synthetic Aperture Radar (SAR) patches consists of feature extraction and classification. Over the past few years image categorization using deep learning became very popular, because it can handle large databases and has shown good recognition results. This paper presents deep convolutional networks for Synthetic Aperture Radar patch categorization. We have tested convolutional networks with20 layers. The CV-CNN consist in general of a real or complex valued input layer, output layer and one or more hidden layers. Hidden layers represent any combination of convolutional layers, pooling layers, activation functions, and are fully defined within complex valued domain. The custom database of patches was designed using 3 classes and parameters of CV-CNN were observed in order to achieve the best accuracy results.**

Keywords – **Synthetic Aperture Radar, Convolutional networks, patch categorization, deep learning.**

I. INTRODUCTION

Synthetic aperture radar (SAR) is an all-weather, night and day imaging system. SAR data sets are nowadays easily accessible from different airborne or spaceborne sources as it is Sentinel-1 and TerraSAR-X, which are mostly devoted to the wider scientific communities. Nowadays a high-resolution SAR images acquired from a spaceborne platforms can achieve resolution of 10 cm. SAR is particularly suitable for land cover classification, target detection, surveillance, land sliding, soil moisture etc. Because of scattering mechanism and speckle noise in SAR imagery, the interpretation and understanding of SAR images is different from visual photo analysis. The image understanding, and data interpretation of SAR data has been studied over the last few decades. The classical classification consists of statistical feature extraction and classification methods. Recently, classification using stacked neural networks (SNN) or shortly deep learning, which include deep belief network (DBN), convolutional neural network (CNN) and recurrent neural network (RNN) have shown very good recognition results using large databases with many categories.

By introducing a deep learning theory [1] for automatically learning features from a data sets using a multistage approach, a deep learning applications have become very attractive. A conventional neural networks and support vector machine (SVM) need feature extraction process to separate features within feature space, otherwise a deep learning does not need task-specific feature extractors and its because of this capable of learning features automatically from data sets. A deep neural network was applied to remote sensing image processing [2] and classification [3]. CNN as one of the typical deep learning models have achieved impressive performances in various fields [4]. It is difficult to separate classes within SAR images, because SAR images have complex scattering mechanisms and random speckle noise. In [5] a deep neutral network for SAR automatic target recognition (ATR) was applied and achieved the highest accuracy (more than 99%) on classification of ten-class targets compared with other cited methods at the time of this writing.

Some complex valued Convolutional Neural networks are not new and there has been many different approaches to complex-valued classification of real or complex valued problems [6]-[7]. Authors in [6] presented a variation of the CNN model with complex valued input and weights. The complex model as a generalization of the real model was proposed. The first investigation of Complex-Valued Convolutional Neural Networks (CV-CNN) for object recognition on Pol-SAR data was proposed in [8]. An architecture with only one single convolutional layer was used and showed promising results. A CV-CNN was presented in [9], by giving the full deduction of the gradient descent algorithm for training this type of networks. The comparison of CNN in the real valued domain - Real Valued Convolutional Neural Network (RV-CNN) for image classification was extended to the complex-valued domain. A deep complex valued CNN used for classification was applied to the Polarimetric SAR data. A comparison between a RV-CNN and CV-CNN was presented in [7].

This paper presents the deep convolutional network, which uses a real valued and complex valued approach. An influence of convolutional filters, number of layers of class recognition for complex and real valued SAR data was demonstrated.

II. REAL-VALUED CONVOLUTIONAL NEURAL NETWORKS

Convolutional neural network models are represented by utilizing various layers of the neural networks. A convolutional neural network consists of input layer followed by optional sub-sampling and regularization layers and ending in fully connected layers. The input into convolutional neural network

[1]Authors are with University of Maribor, Faculty of electrical Engineering and Computer Science, Koroška cesta 46, 2000 Maribor, Slovenia, E-mail: dusan.gleich@um.si.

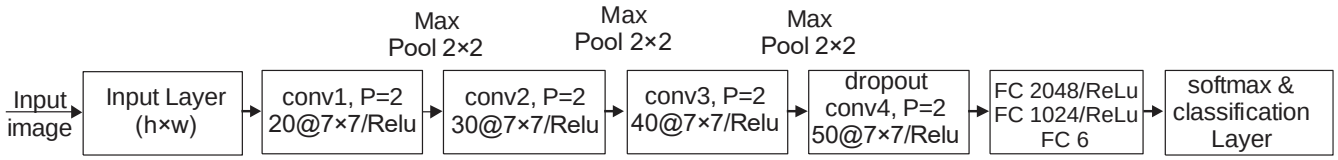


Fig. 1. Convolutional neural network with 20 layers.

is an image which is processed by several filters. The response of the filter is obtained by the convolution, therefore the convolutional response encodes the input and determine features by reducing dimensionality.

The responses of the filters represent inputs to non saturating activation function, which can drastically accelerate learning process. This is done by special rectified linear units (ReLU), which are involved in saturating nonlinearities process. Those functions were applied after every convolutional and fully connected layer. The final layer uses softmax activation to maximize the multinomial logistic regression objective. The result of filtering is usually sub-sampled in order to further reduce dimension of the features makes invariant features to translation. When the max-pool subsampling layers are applied a 2 × 2 max-pool layer divides the output into a set 2×2 cells, which are not overlapped. The maximum activated filter response is recorder into each cell. In this way the input dimensions is reduced by 2 and produced features are invariant to object translations.

The architecture of convolutional neural network is depicted in Fig. 1, which is composed of four convolution layers and three max pooling layers. Each of the first three convolution layers is followed by a max pooling layer, with a pooling size of 2 × 2 and a stride of 2 pixels. The ReLU nonlinearity is applied to every hidden convolution layer. The input image was filtered by 20 convolution filters of size 7×7 in the first convolution layer, resulting in 20 feature maps. The first pooling layer's outputs are sent into the second convolution layer, which has a convolution filter size of 7 × 7, leading to 30 feature maps. The filter size of the third convolution layer is 7 × 7, producing 40 feature maps of size. The fourth convolution layer includes 50 feature maps with a convolution filter size of 7 × 7, which brings out 50 feature maps. The dropout regularization technique is used before fourth convolutional layer. The ReLU nonlinearity was applied after each convolutional layer and after fourth convolutional layer, fully connected layer with dimension of 2048 was inserted before fully connected layers with 1024 and 6 units, respectively. After fully connected layers with 2048 and 1024 units, ReLU layer was applied. At the end of this convolutional network, softmax and classification layers were used. Within this convolutional network a dropout layer was used, which changes architecture and reduces over-fitting. The idea is to connect convolutional and fully connected layers so that hidden neuron outputs are deactivated with probability p during training. This probability was set to p = 0.5. The drop out layer reduces the co-adaptation of neuron. The dropout forces neurons to provide more robust contributions with combination of arbitrary active neuron. The set of neurons is changed randomly in the every epoch and the over-fitting is reduced by 1/(1−p), if it is compared with the network structure without dropout layer.

III. Structure of Complex-Valued Convolutional Neural Networks

In this section the CV-CNN architecture is introduced. Same as with the normal RV-CNN, the CV-CNN is also based on the 2D multichannel input or co-called channel maps. However, the main difference between the two is that each channel array value in CV-CNN is represented in the complex domain as well as hidden layers, which are convolutional filters, pooling filter and activation functions with complex valued inputs and outputs.

The convolutional layer indicates convolution between the sliding window and complex input patch, where the former acts as a filter bank. The result is a matrix where each output value is calculated with a complex dot product sum of the corresponding window and an input patch. Multiple different filter banks are used to search for different features of the specific region in the input patch. This output data can also be interpreted as output maps, which are then further connected to a nonlinear sigmoid or tanh activation function. In this particular approach the sigmoid function was used with the purpose of generating complex feature maps. The convolution result, with included previous layers outputs, can now be described as $O_i^{(l+1)} \in \Im^{W_2 \times H_2 \times I}$ and is calculated by

$$O_i^{(l+1)} = \frac{1}{1+exp(-\Re V_i^{(l+1)})} + j\frac{1}{1+exp(-\Im(V_i^{(l+1)}))}$$
$$V_i^{(l+1)} = w_{ik}^{(l+1)} \star O_k^{(l)} + b_i^{(l+1)} \quad (1)$$

where the filter banks are described with $\omega_{ik}^{(l)} \in \Im^{F \times F \times K \times I}$, input feature maps with $O_i^{(l)} \in \Im^{W_1 \times H_1 \times K}$ and the bias $b_i^{(l+1)} \in \Im^I$. Variable *l* represents the number of current layers, whereas * represents a convolution operator. $V_i^{(l+1)}$ represents weighted sum of inputs to the *i*th output feature maps in the layer *l* +1. The convolutional layer is determined by a number of feature maps *I*, filter size *F* ×*F* ×*K*, stride *S* and zero-padding *P*. The purpose of the pooling layer is to down-sample the patch resolution. Therefore, it is also known as a subsampling layer. With this task a spatial invariance is achieved, making the network insensitive to small shifts or distortions [11]. This is

mostly realized with the subsampling or max pooling function. The subsampling function averages values of the window, where the max pooling function takes the maximum value of the window. An average function was used and restructured for the complex domain as followed

$$O_i^{(l+1)}(x,y) = \underset{u,v=0,\ldots,g-1}{\text{ave}} O_i^{(l)}(x \cdot s + u, y \cdot s + v) \quad (2)$$

where *g* is the pooling size and *s* is the stride. (*x*,*y*) indicates the feature map location of *i*th units$O_i^{(l+1)}(x,y)$. The fully connected layer handles the classification after we calculate the final output of several convolutional and pooling layers. For the CV-CNN multiple fully connected layers have been used to connect each neuron with all the neurons in the previous layer [11]. The output can be described as

$$O_i^{(l+1)} = f(\Re(V_i^{(l+1)})) + jf(\Im(V_i^{(l+1)})) \quad (3)$$

$$V_i^{(l+1)} = \sum_{k=1}^{K} \omega_{ik}^{(l+1)} \cdot O_k^{(l)} + b_i^{(l+1)} \quad (4)$$

where *K* represents the number of neurons in the *l*th fully connected layer. In the final stage the result of multiple neurons is inserted to the output layer which then encodes the specific patch classes. Output of this layer is represented as a vector, described with $1 * C$, where *C* defines the number of classes and also represents the length of a vector. Vector also has to be a complex value, therefore scalar 1 from the previous equation has to be replaced with $(1+1 * j)$. The patch belongs to the class whose distance to vector value location is the shortest.

IV. Experimental results

4.1. Dataset

The custom database was designed, which consisted of 10 Single Look Complex data acquired in the Spotlight mode using TerraSAR-X satellite. Data were acquired over different urban, agriculture and forest areas with different incidence angles. A SAR patch data base was designed by an expert, which manually selected several patches per class, where a ground truth was known. 3 classes were selected: C01 Urban areas, C02 Forest, C03 River. 2000 patches per category were selected for training and 1000 patches for testing. The sample of SAR patches of all 3 categories are depicted in Fig. 2. In Figure samples of classes are depicted horizontally, starting with C1, C2 and C3.

4.2. CNN configuration

We designed 3 different configurations of CNN and changed patch size from 12×12, 24×24 and 48×48. The input was a complex valued image, single polarized in complex-valued format. For real valued approach we used only detected or amplitude data.

The layers of a deep convolutional network were learned in the training phase. Three different neural networks were learned, one, by one for three different sizes of input patches. The parameters of a convolutional are depicted in Fig. 1. Within the training process 1000 samples for each class were used and convolutional network was learned using 100 epochs. The goal of the paper was to investigate a structure of a deep neural network and verify its performances.

Tables 1 and 2 report the over all accuracy of the presented methods for 3 classes using a single complex-valued HH polarized datasets. 1000 images were used for testing stage and average accuracy is reported. Tables 1 and 2 show that the best accuracy was achieved using a complex valued approach.

TABLE I
MEAN ACCURACY OF CLASS RECOGNITION IN % FOR PATCH SIZES OF 12×12, 24×24 AND 48×48 USING A COMPLEX VALUED APPROACH

Class	12 × 12	24 × 24	48 × 48
C1	82	84	73
C2	86	89	64
C3	84	87	65
Total(%)	84	86	67

TABLE II
MEAN ACCURACY OF CLASS RECOGNITION IN % FOR PRESENTED METHODS USING REAL VALUED APPROACH.

Class	12 × 12	24 × 24	48 × 48
C1	82	81	74
C2	63	87	57
C3	68	86	63
Total(%)	71	84	64

From Tables I and II we can conclude that the highest accuracy was obtained with the complex-valued convolutional network using patch size of 24×24 pixels, then 86% of average accuracy was achieved. The real valued convolutional network performed the best the patch size of 24×24, providing accuracy of 84%. The patch size of 12×12 pixels provided accuracy of 84% and 71 % for complex valued and real valued case. The patch size of 48×48 pixels did not provide good recognition results. Experimental results showed that there is still a lot of improvements possible in the structure of deep convolutional networks. The best possible results could be achieved by combining real valued and complex valued parts of the convolutional networks.

V. Conclusion

In this paper a real valued and a complex valued convolutional network for categorization of SAR data were compared using different patch sized. Complex valued are real valued convolutional networks are not the same, but they should be very carefully designed. The experimental results showed that complex valued convolutional networks can archive better results in recognition in comparison to the real valued convolutional networks.

References

[1] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," Science, vol. 313, no. 5786, pp. 504–507, 2006.

[2] W. Diao, X. Sun, X. Zheng, F. Dou, H. Wang, and K. Fu, "Efficient saliency-based object detection in remote sensing images using deep belief networks," IEEE Geoscience and Remote Sensing Letters, vol. 13, no. 2, pp. 137–141, 2016.

[3] F. P. S. Luus, B. P. Salmon, F. van den Bergh, and B. T. J. Maharaj, "Multiview deep learning for land-use classification," IEEE Geoscience and Remote Sensing Letters, vol. 12, no. 12, pp. 2448–2452, 2015.

[4] X. Chen, S. Xiang, C. L. Liu, and C. H. Pan, "Vehicle detection in satellite images by hybrid deep convolutional neural networks," IEEE Geoscience and Remote Sensing Letters, vol. 11, no. 10, pp. 1797–1801, 2014.

[5] S. Chen, H. Wang, F. Xu, and Y. Q. Jin, "Target classification using the deep convolutional networks for sar images," IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 8, pp. 4806–4817, 2016.

[6] N. Guberman, "On complex valued convolutional neural networks," CoRR, vol. abs/1602.09046, 2016. [Online]. Available: http://arxiv.org/abs/1602.09046

[7] H. G. Zimmermann, A. Minin, and V. Kusherbaeva, "Comparison of the complex valued and real valued neural networks trained with gradient descent and random search algorithms," in European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 2011.

[8] R. Haensch and O. Hellwich, "Complex-valued convolutional neural networks for object detection in Polsar data," in 8th European Conference on Synthetic Aperture Radar, June 2010, pp. 1–4.

[9] C. A. Popa, "Complex-valued convolutional neural networks for real-valued image classification," in 2017 International Joint Conference on Neural Networks (IJCNN), May 2017, pp. 816–822.

[10] M. Wilmanski, C. Kreucher, and A. Hero, "Complex input convolutional neural networks for wide angle sar atr," in 2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Dec 2016, pp. 1037– 1041.

[11] Z. Zhang, H. Wang, F. Xu, and Y. Q. Jin, "Complex valued convolutional neural network and its application in polarimetric sar image classification," IEEE Transactions on Geoscience and Remote Sensing, pp. 1–12, 2017.

Design, Simulation and Fabrication of a Microstrip Band-Pass Filter

Ivaylo Vladimirov[1] and Zornitza Krasteva[2]

Abstract – **This document gives the design procedure for preparing a Microstrip band-pass filter using the software package ADS (shortened from Advanced Design System). In order for the filter to be fabricated one-sided FR-4 substrate, see-through scotch tape for the mask and sodium persulfate for the etching are used. In this paper it is made a comparison between the fabricated filter and the virtually designed one.**

Keywords – **ADS (Advanced Design System), design, fabrication, Mictrostrip band-pass filter.**

I. Introduction

New technologies for designing filter are being research all over the world to meet the growing demand for advance filter design with greater frequencies response and characteristic. Filter design often grapple with the trade-off among electrical and physical parameters such as physical size, insertion loss, loss variation, isolation, group delay, and production cost. Different techniques in designing filters have some advantage and disadvantages that need to be considered [1].

Microstrip is a type of electrical transmission line which can be fabricated using printed circuit board technology, and is used to convey microwave-frequency signals. It consists of a conducting strips separated from a ground plane by a dielectric layer known as the substrate. In such a technology, reciprocal and non-reciprocal, passive components are obtained by varying the configuration and dimensions of the printed metallic strips [2].

Microstrip filter design techniques are very compatible, quite durability, but the devices are not size efficient and with a high cost of production. They have been vastly researched all around the world[1].

A band-pass filter is a circuit which is designed to pass signals only in a certain band of frequencies while attenuating all signals outside this band. The LC rather than having a single element in each leg of the filter as in the case of the low pass and high pass filters, the band pass filter has a alternating parallel and series resonant circuit in each leg [3]. The parameters of importance in a bandpass filter are the low pass-band (f_{P1}) and the stop-band (f_{S1}) frequencies, the high pass-band (f_{P2}) and the stop-band (f_{S2}) cut-off frequencies, the bandwidth (B or BW), the center frequency (f_C), and the a maximum and minimum attenuations in respectively the pass-band and the stop-band [4].

[1]Ivaylo Vladimirov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: i_vladimirov@mail.bg.

[2]Zornitza Krasteva is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria. E-mail: zornica95@abv.bg

II. Parameters of the Filter

The goal is to design a microstrip Chebyshev band-pass filter that has to cover the following requirements: a maximum attenuation in pass-band (α_{max}) of 1 dB; a minimum attenuation in stop-band (α_{min}) of 30 dB; a low cut-off stop-band frequency (f_{S1}) of 1 GHz; a low cut-off pass-band frequency (f_{P1}) of 1.5Ghz; a high cut-off pass-band frequency (f_{P2}) of 2Ghz and a high cut-off stop-band frequency (f_{S2}) of 2.5GHz. All of the given parameters are shown in Fig.2.

For lowest cost, microstrip devices may be built on an FR-4 (standard PCB) substrate. However it is often found that the dielectric losses in FR4 are too high at microwave frequencies, and that the dielectric constant is not sufficiently tightly controlled.[2].

In this document a double sided FR4 substrate with a photo resistive layer on each side will be used. The photo restive coating is used to protect the microstrip stubs while etching the whole PCB substrate in the iron trichloride solution.

The FR-4 substrate has a relative dielectric constant of $\varepsilon_r = 4.2$, thickness of $H = 1.5\ mm$, loss tangent of $\tan\delta = 0.02$ and a metallization thickness of $t = 35\,\mu m$.

III. Design of the Filter

In the virtual designing of the filter, the software package ADS (Advanced Design System) is going to be used. ADS is an electronic design automation software system produced by Keysight Technologies [5]. It provides an integrated design environment to designers of RF electronic products such as mobile phones, pagers, wireless networks, satellite communications, radar systems, high-speed data links and even filters [6]. It has lots of functions that make design and production of all kinds of devices much easier.

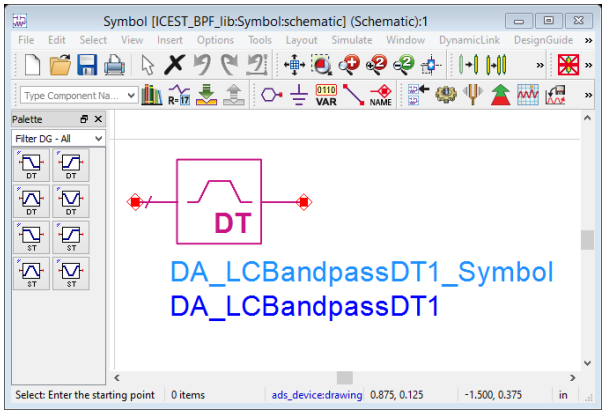


Fig. 1. The symbol of the bandpass filter and the "Filters DE-ALL" library

The design procedure in ADS begins with making a project which consist of choosing: a name; directory where the project will be stored; the libraries needed for the schematic; the technology that will be used. After that a schematic needs to me made. The ADS system has a really helpful library – "Filters DE-ALL" (Fig.1) which have interactive filter devices. By interactive it is meant that they use a function called "Filter Design Guide" to synthesize the filter structure consisting of lumped elements either singly terminated or not. The window of the function is shown in Fig.2. After the desired parameters (attenuations, cut-off frequencies, source and load impedances) are inserted and the resonance type is chosen (in this case Chebyshev) the order of the filter (n) (in this case its 3) automatically calculates. An order of three means that the filter will consists of 3 resonant circuits.

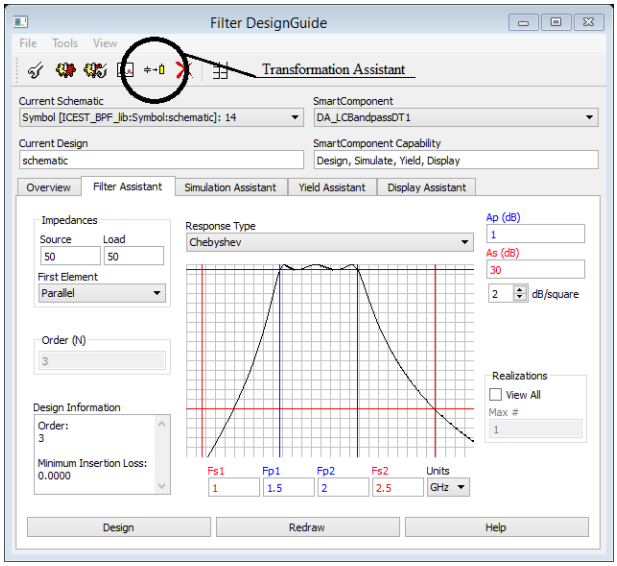


Fig. 2. The "Filter Design Guide" window consisting all the parameters of the filter

There are two topologies that can be used for the filters - the Pi and the T configurations. The "**Π**" (the Greek letter "Pi") configuration starts with a parallel resonant circuit while the "**T**"- with a series one [3]. "**Π**" filters present very-low impedances at high frequencies at both ends due to the capacitive shunting. "**T**" filters conversely have very-high impedances at high frequencies because of the inductive coupling [7]. For this filter a "**Π**" configuration is used.

After all the needed information about the device is filled or chosen in the "Filter Design Guide" window the "Design" button in the bottom right corner needed to be pressed. This triggers the automated design function which creates the filter prototype that consists only of lumped elements. The schematic of the prototype is shown on figure 3.

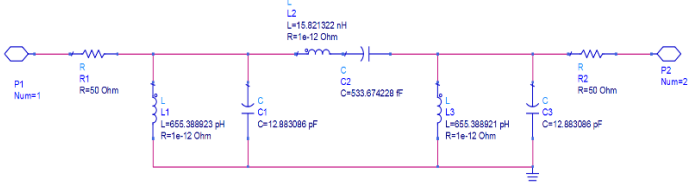


Fig. 3. Schematic of the filter's prototype.

As you can see it has three resonant circuits – a parallel circuit (consisting of L1 and C1) fowolled by a series one (consisting of L2 and C2) and again a parallel one (consisting of L3 and C3) at the end (this is the "**Π**" shape). The resistors R1 and R2 are respectively the source's and the load's resistances. They have a value of 50**Ω** that has been chosen in the "Filter Design Guide" window (Fig.2). The P1 and P2 are the input and output pins.

The next step is to use the "Transformation Assistant" window and transform the lumped elements into microstip lines. The button for opening this utility is in the "Filter Design Guide" window, located in the AJ_FilterToolbar (the exact position is circled on figure 2).

The "Transformation Assistant" utility is used to convert reactive lumped elements such as L – inductors, C – capacitors, and resonant circuits into either a transmition line (TLine) or microstrip line (Microstrip). It also can use the "Kuroda Identities" transformation that remodels a certain transmition line in another without changing any parameters or characteristic of the initial line. Figure 4 shows the window of the "Transformation Assistant" when transfiguring into a microstrip line.

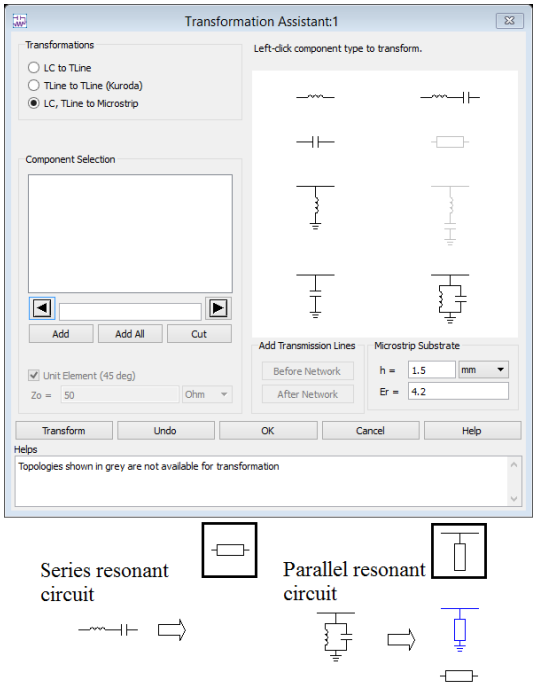


Fig. 4. The "Transformation Assistant" window and the used transformations

For our band-pass filter we are going to use the straight LC to microstrip line transformation. In order for it to be applied the parameters of the used substrate must be loaded onto the system. In the second paragraph it is said that the filter will be fabricated on a FR-4 substrate, so its characteristics (relative dielectric constant of $\varepsilon_r = 4.2$ and thickness of H = 1.5 *mm*) must be inserted. After that by left-clicking on the type of components that you want to transform a list on conversions is given (look at the bottom of figure 4). A series resonant circuit is going to become a series microstrip line, while a parallel resonant circuit – an open-circuited microstrip stub.

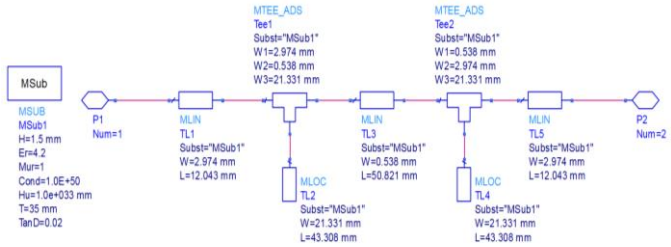


Fig. 5. Schematic of the microstrip filter

After all transformations are done and the lumped prototype is fully transmogrified into the microstrip filter the new schematic can be seen (Fig.5). The scheme has a new "MSub" block consisting of all the characteristics of the used substrate. The source resistance is transformed into the TL1 mirostrip line, L1 and C1 into TL2, L2 and C2 into TL3, L3 and C3 into TL4 and the load resistance into TL5. In order to optimize the design we have added two microstrip T-junctions (Tee1 and Tee2). They are fundamental passive components. in microwave and millimeter-wave circuits [8]. We have added them so that the lengths of the lines match better, they are not lossless and will change the end characteristics of the filter, but because of their small dimensions they won't worsen it much. TL1, TL3 and TL5 are MLIN, which stand for microstrip line and TL2 and TL4 are MLOC - open-circuited microstrip stub. All TLx components have the same parameters – their size – the length and the width. The Tee1 and Tee2 have the widths of all their adjacent components.

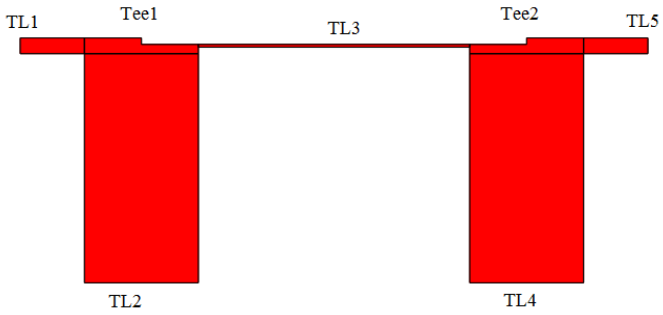


Fig. 6. "Gerber/Drill" file

In order to fabricate the filter a layout is needed to be generated. This is done by clicking the "Generate/Update Layout…" button in the "Layout" toolbar. This creates the actual two-dimensional filter. The exporting of the layout is done in from "Export…" in the "File" toolbar. There are a moderate amount of file types that the layout can be exported to, like: DXF/DWG, GDSII Stream Format, Gerber/Drill, IGES, MaskFile (.msk), Cadance/PCB etc. Figure 6 is an exported "Gerber/Drill" file. This file is usually used by some specialized software in order to drill the filter onto a PCB.

The different dimensions of the microstrip lines can be easily seen in figure 6.

Because any specialized software or hardware for the fabrication won't be used we are going to open the file with an ordinary picture processing software to print the layout onto some kind of a hard cardboard.

IV. Fabrication of the Filter

For the fabrication of the band-pass filter all items shown on figure 7 have been used.

The procedure begins by cutting a suitable portion of a 150x300mm PCB. The used piece is 61x130mm (Fig.7.1). After that a mask is need to be printed out, in this case the mask (Fig.7.2) is printed onto a carboard using a laser printer. In order to use this mask its middle part must be cutted out, which is done by a macket knife (Fig.7.3). For the mask to be used the chopped-out PCB's surface must first be prepared by scrubbing it down with the scouring pad of an ordinady kitchen sponge (Fig.7.4) and by putting a coat of scotch tape over the whole PCB. After the preparation has been done the mask (Fig.7.4) needs to be taped to the circuit board and using it and the knife (Fig7.3), the shape of the filter must be carved onto the scotch tape layer. The excess scotch tapemust be removed in order for the filter to be etched-out.

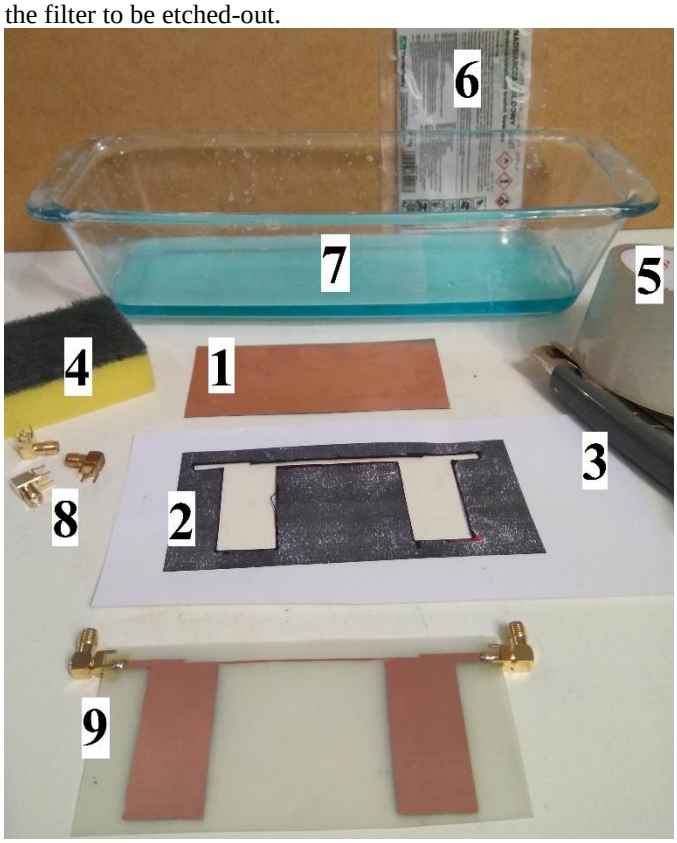


Fig. 7. Items needed for the fabrication of the filter; 1-Initial PCB; 2-Cardpoard mask; 3-macket knife; 4-kitchen sponge; 5-scotch tape; 6-sodium persulfate powder; 7- sodium persulfate solution; 8-SMA coaxial connectors; 9-the fabricated filter.

For the etching solution (Fig.7.7) to be made warm water , about 40˚ to 50˚C, must be combined with the sodium persulfate powder (Fig.7.6). Once the blend is homogeneous the circuit board can be inserted in and the process can begin.

Sodium peroxosulphate, short form NaPS (sodium persulfate is its trade name) is the salt of peroxodisulphuric acid. It is a strong oxidizer and it is almost non-hygroscopic and has particularly good ability to be stored for long time. It is easy and safe to handle. Although it is not combustible, it releases

oxygen easily and assists combustion of other materials. It is clear, so you can see how well it is working and is compatible with resist pens [9].

When all excess copper is removed the filter can be removed from the solution, washed with water to stop the etching process and wiped-out with a towel. After the device is all dried-up the scotch tape layer should be removed and the SMA coaxial connectors can be soldered on using a soldering iron and some lead solder.

The fabricated filter can be seen on figure 8.

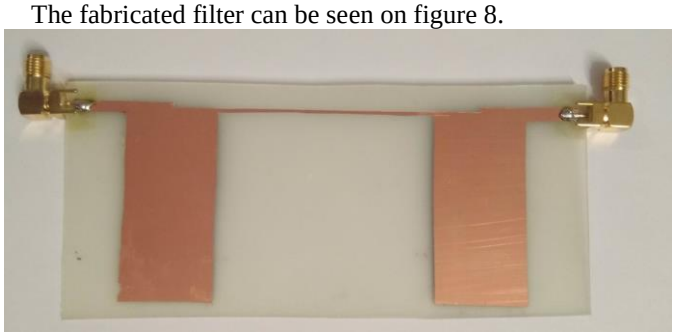


Fig. 8. The fabricated filter

V. SIMULATIONS AND COMPARISON

Comparison of measured from the fabricated filter and simulated results for the transmission and reflection parameters of the microstrip structure is shown in Fig.9. The filter responses were measured by a setting consiting of a Hewlett Packard 8753C Network Analyzer and a Hewlett Packard 85047A S Parameter Set at the frequencies from 900 kHz to 2.8 GHz . The measurement was performed using the SMA connectors. The simulated results are obtained by a Momentum simulation in ADS.

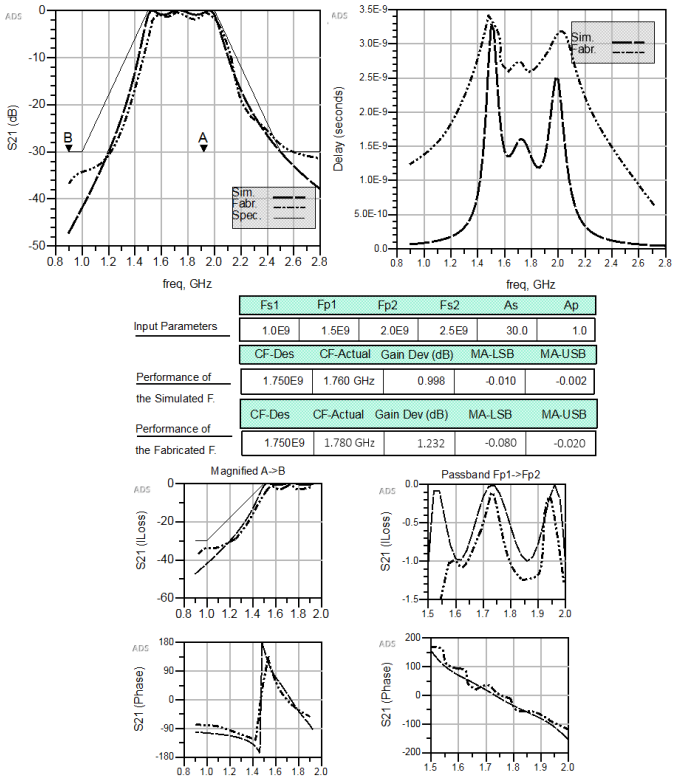


Fig. 9. Comparison results

The measured result shows a small difference in comparison with the simulated results.

The first graph on the left is the transmition characteristic (S21(dB)). The second graph on the right shows the delay of the filter. There are some big differences but considering the scale they are actually quite negligible. The two simulations are followed by a table consisting of all input parameters and the performance of both the simulated and fabricated filters in their central frequency. After the table there are two magnified parts of the S21 graphic (from point A to point B and the pass-band of the filter). On the bottom there are the phases of the signals for the same parts of the S21 graphic as above.

ACKNOWLEDGEMENT

The research described in this paper is partly realized within Erasmus+ mobility and CEEPUS program.

REFERENCES

[1] B. H. Ahmad, M. H. Mazlan, M. N. Husain, Z. Zakaria and N. A. Shairi, “Microstrip Filter Design Techniques: An overview”, VOL. 10, NO. 2, February 2015

[2] S. Halder, S. Sarkar, “A Low-Pass Filter Design using Microstrip”, India, Hildia: Haldia Institute of Technology 2012

[3] Ian Poole, “LC band pass filter circuit”

[4] Lawrence J. Christiano and Terry J. Fitzgerald, “The Band Pass Filter”, July 1999.

[5] “Design and Simulation of High Speed Digital”, 2015

[6] CommsDesign, "Evaluating Performance Tradeoffs in a Dual-Mode, W-CDMA/EDGE Digital IF Receiver", Oct 07, 2002

[7] Jon Gabay "Pi, T Filters Match RF Impedances", 2015-08-20

[8] G. R. Branner, B. Preetham Kumar, and Donald G. Thomas, Jr., “Design of microstrip T-junction power divider circuits for enhanced performance”, September 1995

[9] http://pcbhowto.com/, “PCB Etchant Solutions”

New Business Model for EDA Industry in Internet of Everything Future

Galia Marinova[1] and Aida Bitri[2]

***Abstract* - Internet of everything (IoE) is the newest concept transforming industries. Through this paper we aim to bring a clear understanding on how this driving force is transforming and reshaping electronic design automation (EDA) industry and how companies within industry are trying to find new strategies to adapt and profit in this technological disruption. New approaches and the features of a new business model for EDA industry in IoE are discussed.**

Keywords – **EDA industry, Internet of Everything, Business model.**

I. INTRODUCTION

EDA industry consists on companies designing and producing advanced design electronic tools and systems which can be used in different industries such as healthcare, automotive, defense, aerospace, etc.

For years, since its beginning EDA was a closed industry with traditional approaches at designing. With the latest developments and advancements in technology, the industry is experiencing changes, unknown fields and new challenges. The pressure is so high that businesses inside industry need to adjust, transform and create new business models and philosophies if they want to survive in this disruptive market. Cadence, one of the leaders in EDA, in 2010 released a paper, called EDA360, presenting its own vision on how the industry needs to cooperate and collaborate in order to deal with these innovations [1]. Other companies in EDA, like Synopsys and Mentor Graphics, are trying to find new products to extend their markets [2].

One driving force behind innovations and new products in EDA is the so called, Internet of Everything. IoE is used to describe the connection of all the things, processes, data and people through internet and automatic data exchange [3]. IoE is the extended version of Internet of Things (IoT) including Internet of People and wearable technologies. The term, first introduced by CISCO, 2013, covers not just the technological changes but at the same time explores the economic and strategic implications of IoE innovations for companies [3]. Complex technologies behind IoE bring huge advantages and changes to different industries and economies [4] [5]. IoE transforms existing electronic products and gives rise to new systems and products. At the same time, there are lots of challenges, especially in privacy, security, management of big data, cloud computing and new business models for users, developers, organizations and governments [6][7]. In the continuing sections we try to bring a clear view on how IoE is driving companies at EDA to explore new business models and opportunities.

[1] Galia Marinova is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria.

[2] Aida Bitri is with the Faculty of Information Technology at Aleksander Moisiu University of Durres, L.1, Rr. Currilave, Durres, Albania, Email: aida_bitri@yahoo.com

II. TRENDS AND CHALLENGES

Cisco estimates that 99.4% of physical objects that may one day be part of the IoE are still unconnected. With only about 10 billion out of 1.5 trillion things currently connected globally, there is vast potential to “connect the unconnected” [7][3]. These devices need the right technology, with different specifications on what are used until now. They need to be specialized for specific applications, to be secured, power efficient and to be connected to networks with different types of components and features. The IoT-devices market will require not only a single type of chip but a diverse of them.

Semiconductor devices will be needed for different aspects, especially in cloud, integration and connectivity including computing, sensors, communications and interactivity [8]. Their size must not be a problem and at the same time they need to be as expensive as mass-production devices. The market requires these devices to be produced in small batches with a low price and Moore’s Law may not be anymore the main driving force [9].

The range of applications within IoE go beyond customer electronics and include: wearable devices such as fitness or health accessories, smart home applications like heating, lighting, connected cars, smart cities and especially industrial automation. All these advanced will bring improvement to human life, safety and quality. The demand will grow for EDA and semiconductor industry and this indicates that companies will be more aggressive to find new ways to take advantage from these markets and products [3]. At the same time there is a need for privacy and security, weaknesses, that companies are still working on to fix.

These devices require complex integration. As Cadence, predicted, the market is hunger for integrated systems and for right sophisticated tools to design these systems [1]. The traditional methodologies, where each component is designed, analyzed and produced independently do not give to these companies’ competitive advantages. These methodologies can result in products with higher cost, poor quality and delay [10]. Industries need a more integrated system approach and tools where engineers can carry out an early system prototyping and analyze the interaction of each component while achieving the optimal design. The electronic industry depends on the efficiency of these EDA tools that could help build smaller,

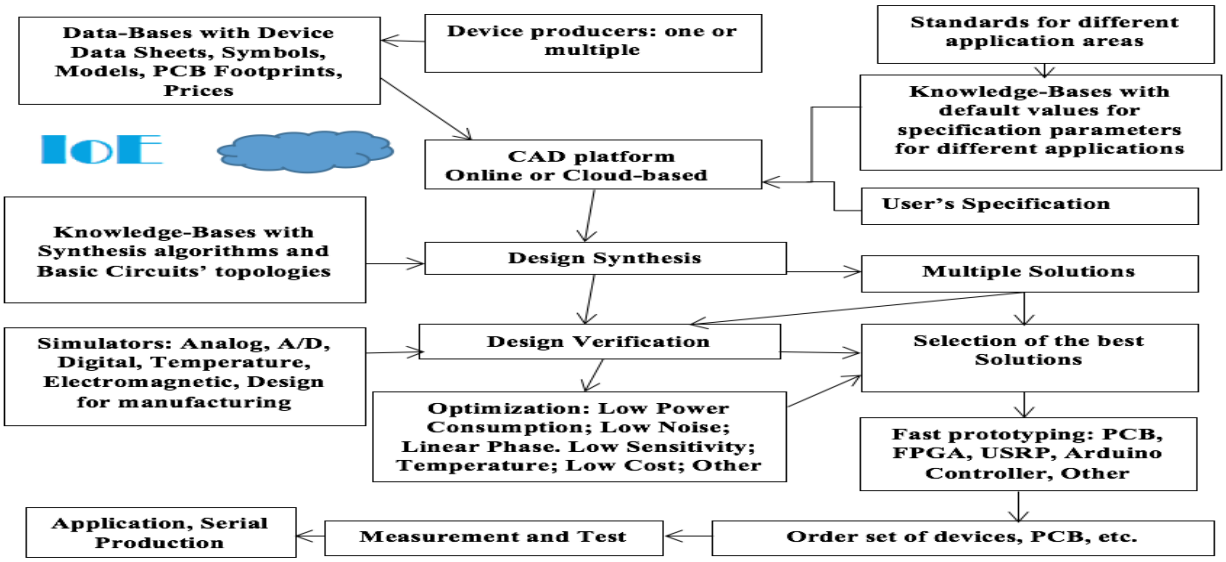


Fig. 1 Integration in EDA industry, Product-oriented design, Cloud and online exposition, IoE environment

faster, cheaper and more intelligent next generation products [10]. Semiconductor and EDA industries are the ones who will provide the necessary baseline for all the successful connectivity through electronic devices. They need to merge and cooperate together because the challenge of building the right infrastructure is a difficult one and industries responsible for this cannot realize it by their own work [10].

Providing connectivity for these small size devices, requires new standards of communications with higher speed and with more intelligence in reacting. That will lead to intelligent and heterogeneous networks [11].

III. A NEW BUSINESS MODEL FOR EDA INDUSTRY

After reading and analyzing the general situation of the industry, and especially the philosophies of the three main EDA companies, Cadence, Synopsys and Mentor graphics, we came up with some approaches for companies to follow and with a business model proposal for the industry. We will use business model canvas, created by Alexander Osterwalder [12] [13].

System integration. IoE includes drastic changes in the core technology and calls to think out of the box for new models and new products. According to Tom Beckley senior vice president of Cadence's Custom IC & PCB Group, the industry must focus in new tools and technologies because it is not anymore about "the chip", but about "the product" [14].

Vic Kulkarni, Director at Global Product Supply, Direct Supply, states that EDA tool providers must think of market opportunities from a perspective of a user, new licenses and pricing model for "mass market", like: low-cost and low-touch technical support, data and IP security. There is a call for companies to go beyond their comfort zone and to exploit new opportunities for providing big data analytics, security stack, etc [14]. Vendors in EDA industry need to unify the solutions and to consider "development of the software stack". Customers have a longer list of requirement and functionality safety comes as an important part where EDA vendors must fulfill it. IoE is a world of hardware and software co-development. Most of the companies must focus on producing new trusted tools for verification and validation, for a secure product. Companies like Synopsys and Cadence are unveiling new simulators for System-on-chip design to be more up to date with the latest verification challenges [15].

Devices produced by the industry will have to fulfill application requirements like, high-power, high performance, application processing, low cost and ultra-power integrated sensors that support sufficient functionality and autonomous device operation. To achieve this level of design flexibility, semiconductor players must rethink their approach to product development model. EDA solutions for designing these devices are critical for the ecosystem [16]. Figure 1 illustrates the approach of system integration in EDA industry which leads to new financial policy for CAD tools providers – costs are mainly taken by device producers instead of users. Further development is focused on CAD tools integration. Research on CAD tools integration through standard passports for CAD tools is presented in [35]. These trends will be reinforced by IoE perspective.

New markets. EDA is well positioned to re-purpose the expertise into new markets, e.g., mechanical design, IP, embedded software and vertical market like automotive and perhaps to seek to expand markets by focusing on design data management, says Laurie Balch, chief analyst at GSEDA [17]. Many companies are working on innovative ideas in home automation, medical devices, automotive and other industries [18]. Thanks to IoT, IP has taken on new importance as a way to deal with complexity and cost [19]. According to Lip-Bu-Tan, CEO and President at Cadence, there is an emerging IP business that is growing rapidly, providing for the company's costumers highly differentiated IP [20] [21]. Other market opportunities include, Mentor Graphics entering the business of selling intelligent IoT gateways, Synopsys focusing on software development and CADENCE getting deeper into embedded vision [9].

The need for collaboration. In its vision EDA360, Cadence, predicted the need for a friendly and cooperative ecosystem inside and outside industry. According to Cadence, the design

Business model canvas for EDA industry businesses.	Designed by: Galia Marinova & Aida Bitri	Date: April 10, 2018.	Version: 1.0
Customer Segments Different sizes of customers, small & big companies. Different industries: Aerospace & Defence, Automotive, industrial, medical, Consumer electronics, telecommunications, commercial electronic industry and manufactures of semiconductors and others. Different customers around the world Different segments for different type of products sensors, microcontrollers, systems, IP, connectivity and memory	**Channels** Direct channels (websites or sales) Partner channels (due to partnerships and merges & acquisitions) Other new sale channels and geographies Academic Network	**Key partners** Foundry Partners Universities Start ups Other industries (automotive, smart homes and cities) Library and Intellectual property partners Industry members	
Value propositions Innovation. Partnerships (Merge and Acquisitions). Systems integration and integrated circuits. Security and privacy protection solutions. Application oriented. Low cost and fast production. Quick access to EDA tools. Building an ecosystem business model. Cloud connectivity and SasS. Low power innovations. Reducing time to market for IoT devices. Software quality and integrity. Smart devices. Brand recognition. Variety of products. Trust. Big data analytics.	**Key activities** Selling EDA tools for semiconductors. Maintain a collaborative approach toward other organizations within industries. Research, development and innovation. New product updates to adjust to changes. Quality & security checks. Production management (cooperation with other start ups). New Marketing, customer and low costs strategies. Problem solving and new design solutions. New Cloud Platform strategies.	**Revenue Streams** New flexible licensing models, different prices from different customers based on their size (seats use at the same time). Intellectual properties, systems and software integrity services. Manufacturing solutions. Leasing emulation hardware technology. Collecting maintenance revenue for software, emulation hardware, and IP Revenues from providing engineering services. Earning royalties from the use of IP. Revenues from Big data analytics. IC physical design & verification. Printed circuit board and multi-chip module revenues.	
Customer Relationships Direct contact with costumers offering support. Analysing data and improving services in real time.	**Cost structure** Research and development Administrative costs Costs on device producers not users.	**Key resources** IP resources New Technology used for designing (cloud-based) Human resources (innovation) Partnerships (academia and start ups)	

Fig. 2: Business model canvas for EDA industry

process is a complex process that needs guidance and companies need new ideas and strategies [1]. Cooperation with academia has been always an important part of the main EDA companies. Cadence, Synopsys, Mentor Graphics, are examples on how industry should do research and exchange ideas with universities to provide new strategies for solving problems [22] [23] [24] [25]. Anupam Bakshi, CEO of Agnisys, InC, says that small companies generate code for the big companies within industry. If big companies want to provide to the customer a product with no mistakes and to fulfill the customer satisfaction, these second-tier companies must have access to tools provided by these big companies. This leads to a strong collaboration between the two of them [26].

Merge & Acquisitions is another technique used by companies to strengthen their position in the industry. The most famous one was the acquisition of Mentor Graphics from Siemens. According to them, Mentor Graphics will help to a new level of offering mechanical, thermal, electrical, electronic and embedded software design capabilities on a single integrated platform [27].

A. Business model canvas for EDA industry

After gathering all the findings, we decided to propose a new business model for EDA. This is the first version and in order to have a clear and concentrated view on our proposal, we have used the business model canvas created by Alexander Osterwalder [12]. There are nine business components in this canvas. These components describe the infrastructure of a company, like key activities, key resources and key partners; the activities and the collection of products and services that bring value to companies in order to distinguish from other companies in the market. Other components like, customer segments, channels, customer relationships describe the type of customers it tries to reach [28] [29] [30] [31] [32] [33] [34]. Figure 2 presents details for each component in the business model canvas for EDA businesses in an IoE future.

IV. Conclusion

At the end we can conclude that IoE is bringing a new wave of changes for EDA. There are a lot of challenges and unknown paths but at the same time huge opportunities. If businesses find ways to cooperate and to use these challenges in their advantage, the profits and the opportunities for the industry will grow and transform the ecosystem. At the business model canvas, we identify the need of organizations to find solutions through discovering new business models, new products and services for their new customers, new strategies on how to handle this huge innovation and how to deal both on the technological and economic aspects and at the same time developing a strong collaborative ecosystem.

References

[1] Cadence, “EDA360: The way forward for Electronic design.” Cadence Design Systems, Inc. 2010. [Online]. Available: http://www.iuma.ulpgc.es/~nunez/IngTelySoc/CDNS%20EDA360%20Vision%20Paper.pdf

[2] G. Moretti, “Evaluating EDA industry using Porter’s Five Forces” Gabeoneda, Sep. 11, 2017. [Online]. Available: http://www.gabeoneda.com/node/5

[3] CISCO, “The Internet of Everything, Global Private Sector Economic analysis.”, CISCO, 2013. [Online]. Available:

https://www.cisco.com/c/dam/en_us/about/ac79/docs/innov/IoE_Economy_FAQ.pdf
[4] M. E. Porter and J. E. Heppelmann (2014, November). "How Smart, Connected Products are transforming competition." *HBR*. [Online]. Available: https://hbr.org/2014/11/how-smart-connected-products-are-transforming-competition [March. 30, 2018]
[5] CISCO, "Securing the internet of everything." CISCO, 2012. [Online].Available:https://www.cisco.com/c/dam/en/us/products/collateral/security/holistic-approach.pdf
[6] M.Sergey, S.Nikolay, E. Sergey, "Cyber security concept for Internet of Everything (IoE)," in System of Signal Synchronization, Generating and Processing in Telecommunications, 3-4 July 2017. [Online]. Available: https://ieeexplore.ieee.org/document/7997540/
[7] Cisco, "Embracing the Internet of Everything". Cisco, 2013. [Online].Available:https://www.cisco.com/c/dam/en_us/buy/cisco-capital/apjc/assets/pdfs/ioe-economy.pdf
[8] M. Gloger, R. Mehrotra, A. Ogrins, A. Sundaram, "Strategies for growth in the IoT era", PwC | Strategy&, 2017. [Online].Available:https://www.strategyand.pwc.com/media/file/Semiconductor-industry.pdf
[9] Ed Sperling, "The great IoE race begins, Ed Sperling", Semiengineering, Nov. 5, 2015. [Online]. Available: http://semiengineering.com/the-great-ioe-race-begins/
[10] V. Umredkar, "Questions and Answers with EDA leaders" June 21, 2014. [Online]. Available: http://electronicsmaker.com/qa-with-eda
[11] E. Worthman, "A new infrastructure for IoE", Semiengineering, March, 9 2015. [Online]. Available: http://semiengineering.com/a-new-infrastructure-for-the-ioe/
[12] T. Greenwald, "Business Model Canvas: A simple tool for designing innovative business models", Forbes, Jan. 31, 2012. [Online].Available:https://www.forbes.com/sites/tedgreenwald/2012/01/31/business-model-canvas-a-simple-tool-for-designing-innovative-business-models/#582e092716a7
[13] A. Osterwalder. "Business model cavas", Strategyzer. Accessed on 30 March 2018. [Online]. Available: www.Strategyzer.com
[14] N. Tyler, "Don't settle for the ordinary, said Tom Beckley at CDNLive" NewElectronics, June 14, 2017. [Online]. Available: http://www.newelectronics.co.uk/electronics-interviews/tom-beckley-spoke-at-cdnlive/156334/
[15] J. Dorsch. "The Growing Importance of Verification in Chip Design" Electronics360, April 26, 2016. [Online]. Available: http://electronics360.globalspec.com/article/6627/the-growing-importance-of-verification-in-chip-design
[16] T. Beckley "The internet of Everything: EDA perspective", in System System on chip Conference, 2-5 Sep. 2014. [Online]. Available: https://ieeexplore.ieee.org/document/6948889/
[17] J. Blyler, "Age of the system comes to EDA", JBSystems, August 31, 2017. [Online]. Available: http://jbsystech.com/age-system-comes-eda/
[18] M. Patel, J. Veira, "Making connections: An industry perspective on the IoT", McKinsey&Company, December2014. [Online]. Available:https://www.mckinsey.com/industries/semiconductors/our-insights/making-connections-an-industry-perspective-on-the-internet-of-things
[19] J. Blyler, "Past, present and future ghosts of EDA-IP for 2017", ChipEstimate, Nov. 30 2017. [Online]. Available: https://www.chipestimate.com/Past-Present-and-Future-Ghosts-of-EDA-IP-for-2017/blogs/2997
[20] N. Tyler, "Lip-Bu talks about the challenges and opportunities facing the EDA sector", NewElectronics, July 14, 2015. [Online]. Available: http://www.newelectronics.co.uk/electronics-interviews/lip-bu-tan-talks-about-the-challenges-and-opportunities-facing-the-eda-sector/87530/
[21] B. Bailey, "IP challenges ahead", Semiengineering, August 10, 2017. [Online]. Available: https://semiengineering.com/ip-challenges-ahead/
[22] R. Goering. "DAC 2015: How academia and industry collaboration can revitalize EDA", Cadence, June 17, 2015. [Online].Available:https://community.cadence.com/cadence_blogs_8/b/ii/posts/dac-2015-how-academia-and-industry collaboration-can-revitalize-eda
[23] N. Steve. "EDA plus academia: a perfect game, set and match", Cadence, July 8, 2014. [Online]. Available: https://community.cadence.com/cadence_blogs_8/b/cic/posts/eda-plus-academia-a-perfect-game-set-and-match
[24] R. Goering. "Cadence aims to strengthen academic partnerships", Cadence, Jan. 12, 2015. [Online]. Available: https://community.cadence.com/cadence_blogs_8/b/ii/posts/cadence-aims-to-strengthen-academic-partnerships
[25] "Synopsys Holds Seminar in Sri Lanka Providing Collaborative Platform for Semiconductor Industry and Academia to Tackle Complex Design Issues", Synopsys, Sep. 2, 2016. [Online]. Available: https://news.synopsys.com/Synopsys-Holds-Seminar-in-Sri-Lanka-Providing-Collaborative-Platform-for Semiconductor-Industry-and-Academia-to-Tackle-Complex-Design-Issues
[26] U. Sarkar, "EDA Companies must collaborate or die", Agnisys, Accessed March 30, 2018. [Online]. Available: https://www.agnisys.com/eda-companies-must-collaborate-or-die/
[27] J. Blyler, "SoS meets SoC as Siemens buys Mentor Graphics", JB Systems, Nov. 28, 2016. [Online]. Available: http://jbsystech.com/sos-meets-soc-siemens-buys-mentor-graphics/
[28] Global Market Insights, Inc. "EDA Market Size By Product (CAE), IC Physical Design & Verification, (PCB) & MCM, SIP, By Application, Industry Analysis Report, Regional Outlook, Growth Potential, Competitive Market Share & Forecast, 2017 – 2024", USA, March 2018, REP: GMI2426. [Online]. Available: https://www.gminsights.com/industry-analysis/electronic-design-automation-eda-market
[29] Global Market Insights, Inc. "EDA Market worth over $14bn by 2014", Global market Insights, Inc. March 9 2018. [Online]. Available: https://www.gminsights.com/pressrelease/electronic-design-automation-eda-market
[30] "Business model of Mentor Graphics", Cleverism, Accessed April 4, 2018. [Online]. Available: https://www.cleverism.com/company/mentor-graphics/
[31] "Business model of Cadence Design systems", Cleverism, Accessed April 4, 2018. [Online]. Avalable: https://www.cleverism.com/company/cadence-design-systems/
[32] "Industry 4.0", Mentor Graphics, Accessed April 3, 2018. [Online]. Available: https://www.mentor.com/embedded-software/iot/
[33] Trac Pharm, "NASDAQ 33rd Inverstor Program", Synopsys, June 16, 2016. [Online]. Avalable: https://www.synopsys.com/content/dam/synopsys/company/investor-relations/34-nasdaq-investor-presentation.pdf
[34] "Business model of Synopsys", Cleverism, Accessed April 4, 2018. [Online]. Avalable: https://www.cleverism.com/company/synopsys/
[35] G. Marinova, O. Chikov, Methodology for tools integration in the Online-assisted Platform for Computer-aided Design in Communications, Proceedings of Int. Conf. ICEST'2015, Sofia, 24-26 June 2015, pp.31-36, ISBN:978-619-167-182-3

Portfolio Risk Optimization Based on MVO Model

Vassil Guliashki[1] and Krasimira Stoyanova[2]

Abstract – **This paper presents a portfolio risk optimization based on Markowitz's mean variance optimization (MVO) model. Historical return data for three asset classes are used to calculate the optimal proportions of assets, included in the portfolio, so that the expected return of each asset is no less than in advance given target value. Ten optimization problems are solved for different expected rate of return. The optimization is performed by a MATLAB solver.**

Keywords – **Portfolio optimization, mean variance optimization model, MATLAB.**

I. INTRODUCTION

Most investors and financial economists acknowledge the investment benefits of efficient portfolio diversification. Markowitz gave the classic definition of portfolio optimality: a portfolio is efficient if it has the highest expected (mean or estimated) return for a given level of risk (variance) or, equivalently, least risk for a given level of expected return of all portfolios from a given universe of securities. The portfolio optimization is a hard optimization problem in the finance area. There are developed different single objective optimization models for different applications in this area. This problem is very important. It is connected with the choice of a collection of assets to be held by an institution or a private individual. The choice should be done in such a way, that the expected return (mean profit) is maximized, while the risk is to be minimized at the same time. Dependent on users preferences, various trade-offs are usually seek.

The relatively low level of analytical sophistication in the culture of institutional equity management is one often-cited reason for the lack of acceptance of MV optimization, along with organizational and political issues.

Formulating this problem in optimization terms, Markowitz [1] states that, ideally, the investor searches for the **optimal portfolio**, i.e., the portfolio that minimizes the risk (within a defined tolerance) while maximizing the return.

II. PROBLEM FORMULATION

The assets S_1 , S_2, … S_n ($n \geq 2$) with random returns are considered. Let a set of $n \in$ N financial assets be given. At time $t_0 \in$ R, each asset i has certain characteristics, describing its future payoff: Each asset i has an expected rate of return μ_i per monetary unit (e. g. dollars), which is paid at time $t_1 \in$ R, $t_1 > t_0$. Let $\mu = [\mu_1, \mu_2,\ldots, \mu_n]^T$. This means if we take a position in $y \in$ R units of asset 1 at time t_0 our expected payoff in t_1 will be $\mu_1 y$ units. Let σ_i be the standard deviation of the return of asset Si. For $i \neq j$, ρ_{ij} denotes the correlation coefficient of the returns of asset S_i and S_j. The correlation coefficient $\rho_{ii} = 1$. Let $\zeta = (\sigma_{ij})$ be $n \times n$ symmetric covariance matrix with $\sigma_{ii} = \sigma_i^2$ and $\sigma_{ij} = \rho_{ij}\sigma_i\sigma_j$ for $i \neq j$, and $i, j \in \{1, ..., n\}$. In this notation σ_{ii} is the variance of asset i-th's rate of return and σ_{ij} is the covariance between asset i-th's rate of return and asset j-th's rate of return.

A portfolio is defined by a vector $x := (x_1, ..., x_n) \in R^n$, which contains the proportions $x_i \in R$ of the total funds invested in securities $i \in \{1, ..., n\}$. The expected return and the variance of the resulting portfolio $x := (x_1, ..., x_n)$ can be presented (see standard Markowitz setting [2 - 7]) as follows:

$$f_1(x) = f_{\text{risk}}(x) = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j \sigma_{ij} \quad (1)$$

$$f_2(x) = f_{\text{return}}(x) = \sum_{i=1}^{n} x_i \mu_i \quad (2)$$

It should be noted that $x^T\zeta x \geq 0$ for any x, since the variance is always nonnegative, i. e. ζ is positive semidefinite. It is assumed here, that ζ is positive definite, which is equivalent to assuming that there are no redundant assets in our collection $S_1, S_2, \ldots, S_n$. Further it is assumed, that the set of admissible portfolios is a nonempty polyhedral set, represented as $X := \{x: Ax = b, Cx \geq d\}$, where A is an $m \times n$ matrix, b is m-dimensional vector, C is a $p \times n$ matrix and d is a p-dimensional vector. One constraint of type linear equality in the set X in the standard problem formulations is:

$$\sum_{i=1}^{n}(x_i) = 1 \quad (3)$$

Simple linear inequality constraints (lower and upper bounds) in the set X are connected with the requirement, that the proportions (weights) of the portfolio should be nonnegative:

$$0 \leq x_i \leq 1,\ i \in \{1, ..., n\} \quad (4)$$

There are several different single objective model formulations of Markowitz'mean-variance optimization (MVO) problem (see [7 - 10]). One single objective MVO model is formulated as follows:

$$\min_x \ \frac{1}{2} x^T \zeta x \quad (5)$$

$$\text{subject to:} \quad \mu^T x \geq T \quad (6)$$

$$Ax = b \quad (7)$$

$$Cx \geq d. \quad (8)$$

This model corresponds to risk minimization. In the first constraint T is a target value, where the expected return is no

[1]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.

[2]Krasimira Stoyanova is with the Institute of Information and Communication Technologies – BAS, "Acad. G. Bonchev" Str. Bl. 2, 1113 Sofia, Bulgaria, E-mail: kr.stoyanova@mail.bg

less than T. In case the problem formulation includes varying T between T_{min} and T_{max}, there will be obtained efficient portfolios.

Defining Efficiency

The notion of defining an optimal set of portfolio weights to optimize risk and return is the basis of Markowitz portfolio efficiency. The efficiency criterion states:

A portfolio P* is MV efficient if it has least risk for a given level of portfolio expected return.

The MV efficiency criterion is equivalent to maximizing expected portfolio return for a given level of portfolio risk.

A portfolio P* is MV efficient if it has the maximum expected return for a given level of portfolio risk.

Which formulation of portfolio efficiency is used is a matter of convenience.

Optimization Constraints

Linear constraints are generally included in institutional MV portfolio optimization. For example, optimizations typically assume that portfolio weights sum to 1(budget constraint eq. 3) and are nonnegative (no-short-selling constraint eq.4). The budget condition is a linear equality constraint on the optimization. The no-short-selling condition is a set of sign constraints or linear inequalities (one for each asset in the optimization) and reflects avoidance of unlimited liability investment often required in institutional contexts. In practice, optimizations often include many additional linear inequality and equality constraints, particularly for equity portfolios.

Computer Algorithms

Several algorithms are available for calculating MV efficient portfolios. Quadratic programming is the technical term for the numerical analysis procedure used to compute MV efficient portfolio in practice. Quadratic programming algorithms allow maximization of expected return and minimization of the variance, subject to linear equality and inequality constraints (see [11 - 13]).

Many algorithms are used for computing MV efficient portfolios. The choice may depend on convenience, computational speed, number of assets, number and character of constraints, and required accuracy.

III. ASSETS DATA USED

Equity portfolio optimization is typical application of MV optimization in asset management. The assets generally include broad asset categories, such as U.S. equities and corporate and government bonds, international equities and bonds, real estate, hedge funds, and venture capital. Sample means, variances, and correlations, based on monthly, quarterly, or annual historic data, may serve as starting points for optimization input estimates. In this study the Markovitz's MVO model is applied to the problem of constructing a long-only portfolio of *French stocks*, *US bonds* and *cash deposit*. Historical return data for these three asset classes are used to calculate the optimal proportions of assets, included in a portfolio, so that the expected return of each asset is no less than in advance given target value. It should be noted that the most MVO models combine historical data with other indicators such as earnings estimates, analyst ratings, valuation and growth metrics, etc. This study differs from the above mentioned approaches. It is focused on the price based estimates for expositional simplicity.

IV. ILLUSTRATIVE EXAMPLE

The 10-year Treasury bond index (CBOE Interest Rate 10 Year T No (^TNX) for the returns on *bonds* and the EURONEXT 100 (^N100) index for the returns on *stocks*, are used. It is assumed that the cash is invested in a money market account whose return is the 1% – deposit *interest rate*. For each asset historic data are used including the annual times series for the "Total Return" from November 1999 through February 2018, e.g. for 220 months period (see [14, 15]).

Let the "Total Return" for asset i = 1,2,3; and t = 0,...,t_f [months], be denoted by I_{it}. Here t = 0 corresponds to November 1999, and $t = t_f$ corresponds to February 2018. For each asset i the raw data I_{it}, t = 0,...,t_f, can be converted into rates of return r_{it}, t = 0,...,t_f [months], by means of the formula:

$$r_{it} = \frac{I_{i,t} - I_{i,t-1}}{I_{i,t-1}} \tag{9}$$

Let the random rate of return of asset i be denoted by T_i. From the historical data we can compute the arithmetic mean rate of return for each asset:

$$\bar{r}_i = \frac{1}{t_f}\sum_{t=1}^{t_f} r_{it} \tag{10}$$

obtaining for the concrete example the result in Table1:

Table 1. Arithmetic mean rate of return for each asset

	Bonds	Stocks	Interest rate
Arithmetic mean $\bar{r}_{it}$	0,02024%	0,13%	0,084545%

Since the rates of return are multiplicative over time, it is preferred the geometric mean to be used instead of arithmetic mean. The *geometric mean* is the constant monthly rate of return, that needs to be applied in months t = 0 through $t = t_f$–1, in order to get the compounded Total Return I_{itf}, starting from I_{i0}. The formula for the geometric mean is:

$$\mu_i = \left(\prod_{t=1}^{t_f}(1+r_{it})\right)^{\frac{1}{t_f}} - 1 \tag{11}$$

The results for the test example are presented in Table 2:

Table 2. Geometric mean rate of return for each asset

	Bonds	Stocks	Interest rate
Geometric mean μ_i	0,36%	0,0123%	0,085%

Further the covariance matrix is computed:

$$\mathbf{cov}(T_i,T_j) = \frac{1}{t_f}\sum_{t=1}^{t_f}(r_{it}-\bar{r}_i)(r_{jt}-\bar{r}_j) \qquad (12)$$

For the test example considered the covariance matrix is presented in Table 3:

Table 3. Covariance matrix for the test example

Covariance	Bonds	Stocks	Interest rate
Bonds	0,0076611701	-0,00011479	-0,000000115
Stocks	-0,00011479	0,0023643199	0,0000000086
Interest rate	-0,000000115	0,0000000086	0,0000000020

The volatility of the rate of return on each asset is computed:

$$\sigma_I = \sqrt{\mathrm{cov}(T_i,T_j)} \ . \qquad (13)$$

The result is presented in Table 4:

Table 4. Volatility of the rate of return for each asset

	Bonds	Stocks	Interest rate
Volatility	0,0875	0,0486	0,0000447

Then the correlation matrix $\rho_{ij} = \frac{\mathrm{cov}(T_i,T_j)}{\sigma_i\sigma_j}$ is also computed (see Table 5):

Table 5. Correlation matrix for the test example

Correlation	Bonds	Stocks	Interest rate
Bonds	1.	-0,026993	-0,029453
Stocks	-0,026993	1.	-0,0039587
Interest rate	-0,029453	-0,0039587	1.

Using the covariance matrix from Table 3 the Quadratic Programming (QP) formulation of the portfolio optimization is:

$$\begin{aligned} \min f = {} & [0{,}0076611701 x_B^2 + 2.(-0{,}00011479)\, x_B.\, x_S + \\ & + 2.(-0{,}000000115) x_B.\, x_I + 0{,}0023643199 x_S^2 + \\ & + 2.0{,}0000000086 x_S.\, x_I + 0{,}0000000020\, x_I^2 \ , \end{aligned} \qquad (14)$$

subject to:
$$x_B + x_S + x_I \geq T$$
$$x_B + x_S + x_I = 1$$
$$x_B \, , x_S \, , x_I \geq 0$$

This problem is solved 10 times, correspondingly for rate of return T = 6%, T = 6,5%, ..., T = 10,5% with increments of 0,5% by means of ***fmincon*** solver of MATLAB "Optimization Toolbox" [16], using the *Interior pint* algorithm.

V. TEST RESULTS

Starting by T = 6%, after 68 iterations the result presented on Fig. 1 obtained.

The optimization results for all ten optimization problems with different T-values by means of ***fmincon*** solver of MATLAB "Optimization Toolbox" are summarized in Table 6.

The curve of efficient frontier is presented on Fig. 2. Every optimal portfolio calculated is presented as a triangle lying on the efficient frontier in the standard deviation / expected return plane.

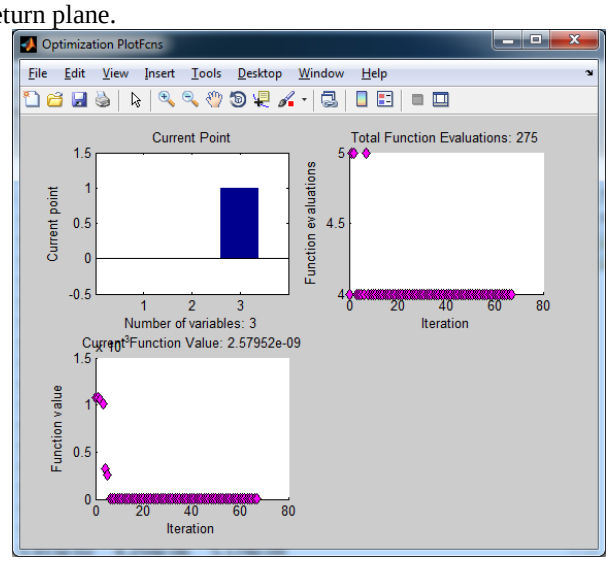


Fig. 1. Optimal portfolio for T = 6%

The optimization results for all ten optimization problems with different T-values by means of ***fmincon*** solver of MATLAB "Optimization Toolbox" are summarized in Table 1 as follows:

Table 6. Optimization results:

Objective function *f*	**Rate of return *T* [%]**	**Iterations**	**Total objective function evaluations**
2,57951566	6.	68	275
2,67812314	6.5	68	275
2,78430648	7.	68	291
2.89806094	7.5	66	270
3.01941047	8.	70	286
3.14832022	8.5	67	273
3.28481381	9.	67	279
3.42888482	9.5	66	271
3.58054119	10.	71	289
3.73975732	10.5	67	274

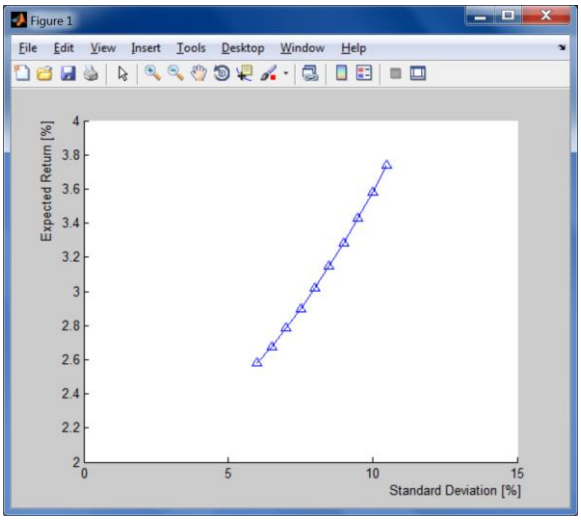


Fig. 2. The Efficient frontier curve

VI. CONCLUSION

Portfolio optimization of three classes of assets (bonds, stocks and cash deposit - interest rate) was performed. Historical return data (220 monthly returns) for these three asset classes are used to calculate the optimal proportions of assets, included in the portfolio, so that the expected return of each asset is no less than in advance given target value of return rate. Fig. 1. shows the results for an optimal portfolio by expected return of 6 %. Table 1 shows simulated optimizations results from 6 to 10,5 with step increase of 0.5% for the rate of return *T*. Through this simulation we get the efficient frontier for the alternative portfolios. The efficient frontier is often helpful in understanding the investments in the portfolio. It serves as useful guidepost for comparing the implications of different portfolios (see [17 - 19]). Fig. 2 shows the relationship between standard deviation and expected return. Logically, as the expected return increases, the deviation increases, too. Because of accepted pessimistic (conservative) strategy **1% positive increasing profitability** for a rate of return, the obtained result is that the bank deposit option is the most secure. We do not comment the security of investment in banks - how sure they are. Interesting investigation would be if a **negative interest rate on the bank deposit is set.** Additional experiments on a representative sample of benchmark problems should be performed for more reliable and precise conclusion about the efficiency and efficacy of the applied approach. Hopefully the mentioned challenges and potential directions for further research will attract more scientists to work in this fruitful area in the future.

ACKNOWLEDGEMENT

This study is realized and partly supported by the CEEPUS network CIII-BG-1103-02-1718.

REFERENCES

[1] Markowitz H., “Portfolio selection”, *Journal of Finance*, 7, March, 1952, pages 77-91.

[2] Schlottmann, F., and D. Seese, "Financial Applications of Multi-Objective Evolutionary Algorithms: Recent Developments and Future Research Directions," in *Applications of Multi-Objective Evolutionary Algorithms*, C. A. Coello Coello and G. B. Lamont, Eds. Singapore: World Scientific, 2004, pages 627-652.

[3] Soleimani, H., H. Golmakani, and M. Salimi, "Markowitz-based portfolio selection with minimum transaction lots, cardinality constraints and regarding sector capitalization using genetic algorithm," *Expert Systems with Applications*, vol. 36, no. 3, pp. 5058-5063, 2009.

[4] Shoaf, J., and J. Foster, "A Genetic Algorithm Solution to the Efficient Set Problem: A Technique for Portfolio Selection Based on the Markowitz Model," in *Proceedings of the Decision Sciences Institute Annual Meeting*, Orlando, Florida, 1996, pp. 571-573.

[5] Yan, W., R. Miao, and S. Li, "Multi-period semi-variance portfolio selection: Model and numerical solution," *Applied Mathematics and Computation*, vol. 194, no. 1, pp. 128-134, 2007.

[6] Radziukyniene, I., and A. Zilinskas, "Approximation of Pareto set in multi objective portfolio optimization," in *World Congress on Engineering (Selected Papers)*, ser. Lecture Notes in Electrical Engineering. Springer, 2008, pp. 551-562.

[7] Lwin, K. T., R. Qu and B. L. MacCarthy (2017), Mean-VaR portfolio optimization: A nonparametric approach, *European Journal of Operational Research*, Volume 260, Issue 2, 2017, ISSN 0377-2217, https://doi.org/10.1016/j.ejor.2017.01.005, pp. 751-766.

[8] Cornuejols, G. & Tütüncü R., *Optmization Methods in Finance*, 349 p., Carnegie Mellon University, Pittsburgh, PA 15213 USA, 2006.

[9] Liagkouras K. and K. Metaxiotis, (2015), Efficient Portfolio Construction with the Use of Multiobjective Evolutionary Algorithms: Best Practices and Performance Metrics, *International Journal of Information Technology & Decision Making*, May 2015, Vol. 14, No. 03 : pp. 535-564

[10] Jin, Y. and Qu, R. and Atkin, J. (2016) *Constrained portfolio optimisation: the state-of-the-art Markowitz models.* In: The 2016 International Conference on Operations Research and Enterprise Systems, 23-25 February 2016, Rome, Italy, Proceedings of 5th the International Conference on Operations Research and Enterprise Systems, ISBN: 9789897581717, pp. 388-395.

[11] Bertsimas D. and R. Shioda, Algorithm for cardinality-constrained quadratic optimization, *Computational Optimization and Applications*, doi: 10.1007/s10589-007- 9126-9. https://pdfs.semanticscholar.org/e68e/7d526372903b4cc510bc1e0e8a3a0b3843ad.pdf

[12] Bienstock D., Computational study of a family of mixed-integer quadratic programming problems, *Math. Programming* 74 (1996), 121-140.

[13] Bomze M., On standard quadratic optimization problems, *J. Global Optim.* 13 (1998), no. 2, 369-387.

[14] EURONEXT 100 (^N100), https://finance.yahoo.com/quote/%5EN100?p=%5EN100

[15] CBOE Interest Rate 10 Year T No (^TNX), https://finance.yahoo.com/bonds

[16] MATLAB, http://www.mathworks.com/products/matlab/

[17] Hochreiter, R., "An evolutionary computation approach to scenario-based risk-return portfolio optimization for general risk measures," in *Proceedings of the 2007 EvoWorkshops 2007 on EvoCoMnet, EvoFIN, EvoIASP, EvoINTERACTION, EvoMUSART, EvoSTOC and EvoTransLog*. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 199-207.

[18] Garkaz M., “The Selection and Optimization of Stock Portfolio using Genetic Algorithm based on Mean Semi-variance Model”, International Conference on Economics and Finance Research. IPEDR, LACSIT Press, Singapore, 4, 2011, pages 379-381.

[19] Di Tollo, G. and Roli, A. (2008). Metaheuristics for the portfolio selection problem. *International Journal of Operations Research*, 5(1):13-35.

Point charge in the air spherical hollow inside bi-isotropic material

Žaklina J. Mančić[1], Zlata Ž. Cvetković[2] and Saša S. Ilić[3]

Abstract –**In this paper, expressions for both electric and magnetic scalar potential of the point charge placed in air spherical hollow, inside bi-isotropic material of Tellegen type, are derived using the image theorem, which represents a contribution to the electromagnetic analysis of bi-isotropic materials.**

Keywords –**electrostatics,bi-isotropic material, image theorem, point charge.**

I. INTRODUCTION

Bi-isotropic materials have been incorporated in various electric devices and the research of specific problems is reported in a number of papers [1−5].In previous years, authors have been reporting solutions of different problems, which incorporate these mediums, on several occasions, using EEM (Equivalent Electrode Method), which indirectly uses Image theorem for obtaining Green's function [6−8], or by applying numerical methods such as FEM (Finite Element Method) [9]. Furthermore, image theorem is exploited in the analysis of bi-isotropic body placed in homogeneous electric field [10] or for field calculation of charged ring near bi-isotropic sphere[11]. Bi-isotropic materials can be described using the following constitutive relations [1,2]

$$\mathbf{D} = \varepsilon\mathbf{E} + \xi\mathbf{H} \quad \text{and} \quad \boldsymbol{B} = \mu\boldsymbol{H} + \xi\boldsymbol{E} \tag{1}$$

where***D*** represents displacement fieldvector, ***E***is electric field vector, ***H*** is magnetizing field vector,***B*** is magnetic field vector, while ε is permittivity of the material, μ is permeability whereas ξ is the parameter which describes bi-isotropic characteristics of the material.

Such materials consist of elements that have permanent electrical and magnetic dipoles, parallel or antiparallel with others according to Tellegen. Thus, electric field in such material simultaneously orients both electrical and magnetic dipoles. Then as well, the magnetic field in such material orients both electric and magnetic dipoles at the same time.
Laplace's equation for the electric scalar potential is obtained starting from Maxwell's equations and constituent relation (1):

$$\Delta\varphi = -\rho_s / \varepsilon_e,\ \varepsilon_e = \varepsilon(1 - \xi^2/\varepsilon\mu), \tag{2}$$

while Laplace's equation for the magnetic scalar potential of free electric charges in the bi-isotropic environment is:

$$\Delta\varphi_m = \xi\rho_s /(\varepsilon_e\mu), \tag{3}$$

where $\varepsilon\mu \neq \xi^2$ [12] and $\Delta(..)$ is the Laplacian function[13].

It is obligatory to integrate the Poisson's equations for the electric and magnetic scalar potential when determining the electric field in such environments, with respecting the existing boundary conditions. Then, field components are obtained using defining relations:

$$\boldsymbol{E} = -\text{grad}\,\varphi \text{ and } \boldsymbol{H} = -\text{grad}\,\varphi_m. \tag{4}$$

In this paper,both theoretical field and potential calculation of the electric and magnetic scalar potential for the case of a pointcharge, placed into the air spherical hollow, is presented. It represents a contribution to the electrostatics of bi-isotropic materials.

II. POINT CHARGE IN THE AIR SPHERICAL HOLLOW IN THE BI-ISOTROPIC MATERIAL- PROBLEM DEFINITION

If we set coordinate system such the charge is placed along *z*-axis while the center of air spherical hollow and the center of coordinate system coincide (Fig. 1), both electric and magnetic scalar potential inside the sphere satisfies Poisson's equation whereas the potentials outside the sphere satisfies Laplace's equation. The system is axially symmetrical and the potentials do not depend on ψ -coordinate of the spherical coordinate system [13].

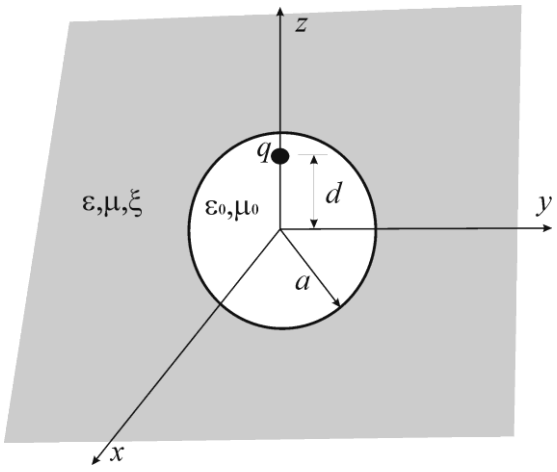


Fig. 1. Point charge in the air spherical hollow which is placed in bi-isotropic material.

[1]Žaklina J. Mančić, Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18106 Niš, Serbia, E-mail: zaklina.mancic@elfak.ni.ac.rs.

[2]Zlata Ž. Cvetković, Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18106 Niš, Serbia,E-mail zlata.cvetkovic@elfak.ni.ac.rs

[3]Saša S. Ilić, Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18106 Niš, Serbia,E-mail sasa.ilic@elfak.ni.ac.rs

The equation (2) for electric scalar potential is defined as:

$$\sin(\theta)\frac{\partial}{\partial r}\left(r^2\frac{\partial\varphi}{\partial r}\right)+\frac{\partial}{\partial\theta}\left(\sin(\theta)\frac{\partial\varphi}{\partial\theta}\right)=\begin{cases}-\dfrac{q\delta(r-d)\delta(\theta)}{2\pi\varepsilon_0}, & r<a\\ 0, & r>a\end{cases} \tag{5}$$

whereas the equation for magnetic scalar potential is:

$$\sin(\theta)\frac{\partial}{\partial r}\left(r^2\frac{\partial\varphi}{\partial r}\right)+\frac{\partial}{\partial\theta}\left(\sin(\theta)\frac{\partial\varphi}{\partial\theta}\right)=0,\ 0<r<\infty. \tag{6}$$

In (5) δ is the Kronecker delta function.This is a function of two variables, usually just non-negative integers. The function isequal toone if the variables are equal $(r=d)$, and equal to zero otherwise.

The solution can be represented in the form:

$$\varphi=\sum_{n=0}^{\infty}R_n(r)P_n(\cos(\theta)), \tag{7}$$

where $P_n(\cos(\theta))$ are Legendre polynomials of the first kind[13].

If we substitute equation (7) in equation (6), using Legendre differential equation:

$$\frac{\partial}{\partial\theta}\left(\sin(\theta)\frac{dP_n}{d\theta}\right)=-n(n+1)\sin(\theta), \tag{8}$$

where $P_n=P_n(\cos(\theta))$,it is obtained:

$$\sum_{n=0}^{\infty}\left[\frac{d}{dr}\left(r^2R_n'\right)-n(n+1)R_n\right]P_n(\cos(\theta))\sin(\theta)=-\frac{q}{2\pi\varepsilon_0}\delta(r-d)\delta(\theta) \tag{9}$$

If we multiply the previous equation with $P_{n'}(\cos(\theta))$, and integrate it with respect to θ, from 0 toπ, we obtain:

$$\frac{d}{dr}\left(r^2R'_n\right)-n(n+1)R_n=-\frac{q}{4\pi\varepsilon_0}(2n+1)\delta(r-d), \tag{10}$$

taking into account that it is valid:

$$\int_0^{\pi}\sin(\theta)P_n(\cos\theta)P_{n'}(\cos(\theta))d\theta=\begin{cases}0, & n\neq n'\\ \dfrac{2}{2n+1}, & n=n'\end{cases} \tag{11}$$

and

$$P_n(1)=1.$$

The equation (8) can be solved using the variation of constant method:

$$R_n(r)=\begin{cases}C_nr^n+D_n/r^{n+1}, & r\le d\\ \left(C_n-\dfrac{q}{4\pi\varepsilon_0d^{n+1}}\right)r^n+\left(D_n+\dfrac{qd^n}{4\pi\varepsilon_0}\right)\dfrac{1}{r^{n+1}}, & d\le r\le a.\\ A_nr^n+\dfrac{B_n}{r^{n+1}}, & a\le r\end{cases} \tag{12}$$

Because the charge point is in the air, the equation (3) for magnetic scalar potential is:

$$\sin(\theta)\frac{\partial}{\partial r}\left(r^2\frac{\partial\varphi_m}{\partial r}\right)+\frac{\partial}{\partial\theta}\left(\sin(\theta)\frac{\partial\varphi_m}{\partial\theta}\right)=0, \tag{13}$$

Analogously to the previous consideration, the solution for Eq (13) can be represented in the form:

$$\varphi_m=\sum_{n=0}^{\infty}R_{1n}(r)P_n(\cos(\theta)), \tag{14}$$

where function $R_{1n}(r)$ is:

$$R_{1n}(r)=\begin{cases}C_{1n}r^n+D_{1n}/r^{n+1}, & r\le a\\ A_{1n}r^n+\dfrac{B_{1n}}{r^{n+1}}, & a\le r\end{cases}. \tag{15}$$

As the electric scalar potential φ in (12) has a finite value in the center of the sphere, as well as in the infinity, it has to be valid:

$$D_n=0 \text{ and } A_n=0. \tag{16}$$

Similarly, magnetic scalar potential φ_m has a finite value in the center of the sphere and in the infinity, thus:

$$A_{1n}=0 \text{ and } D_{1n}=0 \tag{17}$$

in (15).

The remaining constants can be determined from the boundary conditions for the continuity of the electric scalar potential:

$$\varphi(r=a-0)=\varphi(r=a+0), \tag{18}$$

$$\varphi_m(r=a-0)=\varphi_m(r=a+0), \tag{19}$$

$$D_n(r=a-0)=D_n(r=a+0), \tag{20}$$

$$B_n(r=a-0)=B_n(r=a+0). \tag{21}$$

From the condition of electric scalar potential, φ ,continuity, it is:

$$\frac{B_n}{a^{n+1}}=\left(C_n-\frac{q}{4\pi\varepsilon_0d^{n+1}}\right)a^n+\frac{qd^n}{4\pi\varepsilon_0}\frac{1}{a^{n+1}}, \tag{22}$$

while from the magnetic scalar potential (φ_m) continuitycondition:

$$C_{1n}\, a^n = \frac{B_{1n}}{a^{n+1}}. \tag{23}$$

Furthermore, from the continuity condition of D_n on the boundary surface, where D_n is the normal component of vector ***D***, it is obtained:

$$-\varepsilon(n+1)\frac{B_n}{a^{n+2}} - \xi(n+1)\frac{B_{1n}}{a^{n+2}} = \varepsilon_0 n a^{n-1}\left(C_n - \frac{q}{4\pi\varepsilon_0 d^{n+1}}\right) - \varepsilon_0(n+1)\left(\frac{qd^n}{4\pi\varepsilon_0}\right)\frac{1}{a^{n+2}}. \tag{24}$$

Next, from the condition of vector B_n continuityon the boundary surface, where B_n is the normal component of vector ***B***, it can be obtained:

$$-\mu(n+1)\frac{B_{1n}}{a^{n+2}} - \xi(n+1)\frac{B_n}{a^{n+2}} = \mu_0 n a^{n-1} C_{1n}. \tag{25}$$

In the end, it is obtained:

$$C_n = \frac{q}{4\pi\varepsilon_0 d^{n+1}} + \frac{qd^n}{4\pi\varepsilon_0 a^{2n+1}}\frac{(n+1)\{(\varepsilon_0-\varepsilon)[\mu(n+1)+\mu_0 n]+\xi^2(n+1)^2\}}{\{[\varepsilon(n+1)+\varepsilon_0 n][\mu(n+1)+\mu_0 n]-\xi^2(n+1)^2\}} \tag{26}$$

$$B_n = \frac{qd^n}{4\pi}\frac{(2n+1)[\mu(n+1)+\mu_0 n]}{[\varepsilon(n+1)+\varepsilon_0 n]\{[\mu(n+1)+\mu_0 n]-\xi^2(n+1)^2\}} \tag{27}$$

$$B_{1n} = -\xi\frac{qd^n}{4\pi}\frac{(2n+1)(n+1)}{[\varepsilon(n+1)+\varepsilon_0 n]\{[\mu(n+1)+\mu_0 n]-\xi^2(n+1)^2\}} \tag{28}$$

$$C_{1n} = -\xi\frac{qd^n}{4\pi a^{2n+1}}\frac{(2n+1)(n+1)}{[\varepsilon(n+1)+\varepsilon_0 n]\{[\mu(n+1)+\mu_0 n]-\xi^2(n+1)^2\}}. \tag{29}$$

III. NUMERICAL RESULTS

It is of interest to consider parameters which appear in the constants C_n, B_n, C_{1n}, B_{1n} and which depend on $\xi^2/\varepsilon_0\mu_0$. Let's consider the case $\varepsilon = 2\varepsilon_0$ and $\mu = \mu_0$. Furthermore, let's denoted parameters as:

$$p(\varepsilon/\varepsilon_0, \mu/\mu_0, n) = \frac{(n+1)\left\{\left(1-\frac{\varepsilon}{\varepsilon_0}\right)\left[(n+1)\frac{\mu}{\mu_0}+n\right]+\frac{\xi^2}{\varepsilon_0\mu_0}(n+1)^2\right\}}{\left[\frac{\varepsilon}{\varepsilon_0}(n+1)+n\right]\left[\frac{\mu}{\mu_0}(n+1)+n\right]-\frac{\xi^2}{\varepsilon_0\mu_0}(n+1)^2} \tag{30}$$

$$p_1(\varepsilon/\varepsilon_0, \mu/\mu_0, n) = \frac{(2n+1)(n+1)}{\left[\frac{\varepsilon}{\varepsilon_0}(n+1)+n\right]\left[\frac{\mu}{\mu_0}(n+1)+n\right]-\frac{\xi^2}{\varepsilon_0\mu_0}(n+1)^2} \tag{31}$$

$$\alpha(\varepsilon/\varepsilon_0, \mu/\mu_0, n) = \left[\frac{\varepsilon}{\varepsilon_0}(n+1)+n\right]\left[\frac{\mu}{\mu_0}(n+1)+n\right] - \frac{\xi^2}{\varepsilon_0\mu_0}(n+1)^2 \tag{32}$$

and let us show their graphical dependence on bi-isotropy parameter $\xi^2/(\varepsilon_0\mu_0)$, when n is parameter.

From Fig. 2 and Fig.3 it can be seen that parameters $p$and p_1 increase by increasing the bi-isotropy parameterξ, whereas parameter α decreases by increasing bi-isotropy parameter ξ.

In these three figures, n is the parameter.

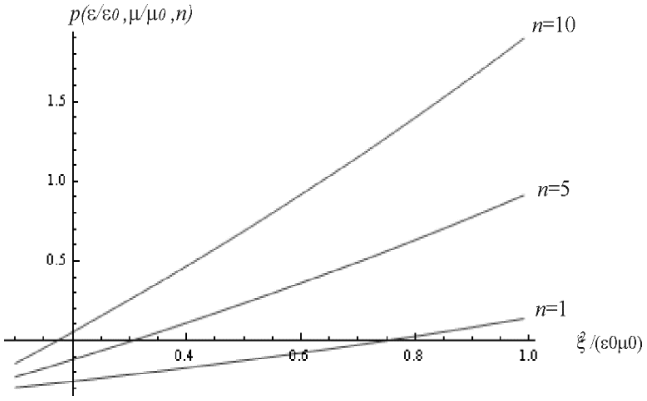


Fig. 2. Dependence $p(\varepsilon/\varepsilon_0, \mu/\mu_0, n)$ onbi-isotropy parameter.

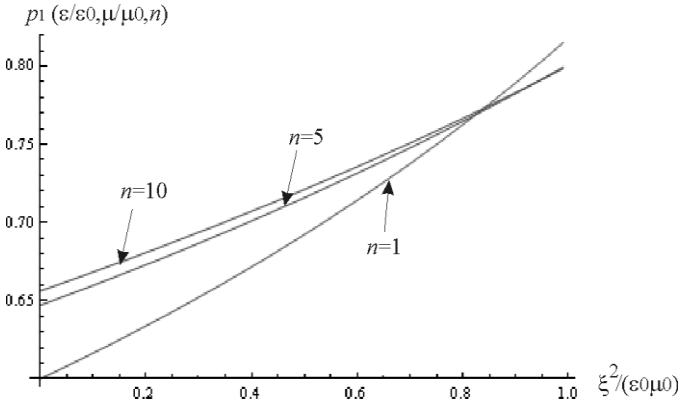


Fig. 3. Dependence $p_1(\varepsilon/\varepsilon_0, \mu/\mu_0, n)$ on bi-isotropy parameter

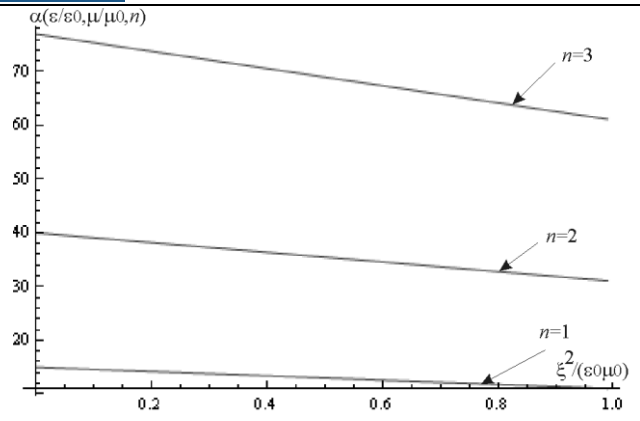


Fig. 4. Dependence $\alpha(\varepsilon/\varepsilon_0, \mu/\mu_0, n)$ on bi-isotropy parameterξ.

IV. CONCLUSION

In this paper, we have derived electric and magnetic scalar potential of point charge which is placed in air spherical hollow, which exists inside bi-isotropic medium of Tellegen type. The expressions are obtained by solving Legendre differential equations for electric and magnetic scalar potential and by satisfying boundary conditions for both potentialcontinuity on boundary surfaces as well as continuity of normal components of vectors $\boldsymbol{D}$ and $\boldsymbol{B}$.

As a consequence of specific constitutive relations for bi-isotropic mediums, constants C_n, B_n, C_{1n}, B_{1n} depend on bi-isotropy parameter, which characterizes the bi-isotropic properties of the surrounding environment outside air spherical hollow.

This further influences both the electrical and magnetic scalar potential, as well as the fields that exist in the vicinity of the point charge, which are obtained by equation (4).In addition, the presence of Tellegen's bi-isotropic environment, although the point charge is placed in the air hollow, leads to the appearance of a magnetic field. When $\xi = 0$, expressions are reduced to those that are derived [13] and magnetic field does not exist.

ACKNOWLEDGEMENT

This work was supported by the Serbian Ministry of Education, Science, and Technological Development under grant TR-32052 and III44004.

.

REFERENCES

[1] A.H. Sihvola, "Electromagnetic modelling of bi-isotropic media", Progress In Electromagnetics Research, PIER no 9, pp.45-86,1994.

[2] W. Ren, "Bi-isotropic layered mixtures", Progress in Electromagnetics research, PIER, 9, pp. 103-116, 1994.

[3] A.H. Sihvola, A.J. Viitanen, I.V. Lindell,S.A. Tretyakov,*Electromagnetic Waves in Chiral and Bi-sotropic media*, (Artech House Antenna Library), ISBN-13: 978-0890066843.

[4] I.V. Lindeli, "Quasi-static image theory for Bi-isotropic Sphere", IEEE Trans. Antennas and Propagat., Vol. 40, No.2, pp. 228-233, Feb.1992.

[5] I. V. Lindell,J. J. Hänninen, "Static image principle for the sphere inisotropicorbi-isotropicspace", Radio Science, Vol. 35, Issue 3, pp.653–660, 2000.

[6] D. Veličković, Ž. Mančić, "Image theorem invarement in biisotropic media", PES 96, Niš, 22-23 May, pp. 59-63, 1996.

[7] N.Raičević, S.Ilić. "One method for chiral and metamaterial strip lines determination", International conference in European Electromagnetics, EUROEM 2008, Lausanne, Switzerland, pp. 44, 21-25 july 2008.

[8] D.Velickovic, B. Milovanovic, Z. Mancic, "New approach for equalising phase velocities in coupled microstrip lines", Electronic letters, Vol. 33, Issue 17, pp. 1444 -1445, 1997.

[9] Z. J. Mancic, V. V. Petrovic, "Quasi-Static Analysis of the Shielded Microstrip Line with Bi-isotropic Substrate by the Strong FEM Formulation" (Proceedings Paper), 2013 11TH International Conference on Telecommunication in Modern Satelite, Cable and Broadcasting Services (TELSIKS) Vols 1 and 2, pp. 513-516, 2013.

[10] Ž. J. Mančić, S. S. Ilić, "Contribution to the Electrostatic Analysis of Bi-isotropic Mediums of Tellegen Type", Proceedings of 13th International Conference on Telecommunication in Modern Satellite, Cable and Broadcasting Services (TELSIKS 2017), Niš, Serbia, October 18–20, pp. 359–362, 2017.

[11] Ž. J. Mančić, Z. Ž. Cvetković, S. S. Ilić, "Contribution to the electrostatics of bi-isotropic materials in the case of lineic ring of charge placed beside a sphere of a bi-isotropic material", The Small Systems Simulation Symposium, CD P10, Niš, Serbia, pp. 59-62, 2018, ISBN 978-86-6125-199-3.

[12] D. Veličković, "Wave equations and its solutions in bianisotropic media", Scientific meeting New sintered materials and their applications, Nis, pp. 40-41, 1990.

[13] D. Veličković, *Calculation methods for electrostatic fields*, Stil, Podvis, Niš, 1982.

Electromagnetic pollution estimation in a communication laboratory

Zdravka Tchobanova[1], Galia Marinova[2], Vera Marković [3] and Aleksandar Atanasković[4]

Abstract – **A case of the electromagnetic pollution in a communications laboratory at the Faculty of Electronic Engineering in Niš, Serbia, is investigated in this paper. In order to estimate whether the EM exposure limits are respected, the EM safety standards both in Bulgaria and Serbia are presented and compared. On the basis of these standards, the analysis of measured data is performed and obtained results are discussed. Finally, some rrecommendations for working with equipment in a communications laboratory are given.**

Keywords – **Electromagnetic pollution, EMF exposure levels, EM safety standards**

I. Introduction

In the large laboratories with many communication devices, the question arises whether it is safe for the health to work in them, i.e. whether the level of EM radiation does not exceed the limit values. The study of electromagnetic pollution in a communications laboratory was done in the Laboratory for the Microwave Technique and Wireless Communications at the Faculty of Electronic Engineering of the University of Nis, Republic of Serbia.

Electromagnetic radiation is classified broadly into ionizing and non-ionizing radiation. Non-ionizing radiation includes the spectrum of ultraviolet (UV), visible light, infrared (IR), microwave (MW), radio frequency (RF), and extremely low frequency (ELF). Microwave and RF frequency ranges are of the particular interest because there are a lot of wireless communication applications in these areas.

The impact of electromagnetic fields (EMF) on the human body has two aspects: direct biological effects and indirect effects [1]. Direct effects occur in the human body as a direct result of its presence in an electromagnetic field. They include thermal effects, such as tissue heating due to absorption of energy from electromagnetic fields, and non-thermal effects like the stimulation of muscles, nerves or sensory organs. The non-thermal effects can harm the mental and physical health and cause symptoms such as dizziness, visual disturbances, annoyance, sleep disorders, etc. [2][3][4]. Indirect effects include interference with medical electronic equipment and devices, such as cardiac pacemakers, metal prostheses, defibrillators and cochlear implants.

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

[2]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.

[3]Vera Marković is with Faculty of Electronic Engineering at University of Nis, Aleksandra Medvedeva 14, 18000 Niš, R Serbia, E-mail: vera.markovic@elfak.ni.ac.rs

[4]Aleksandar Atanasković is with Faculty of Electronic Engineering at University of Nis, Aleksandra Medvedeva 14, 18000 Niš R Serbia, E-mail: aleksandar.atanaskovic@elfak.ni.ac.rs

There are many studies on the effects of electromagnetic fields on human health. The International Agency for Research on Cancer (IARC) scored emission of electromagnetic fields as a possible carcinogen (group 2B). However the available data is not convincing and is insufficient to establish a direct link between possible carcinogenic effects and long-term exposure to electromagnetic fields.

II. Safety regarding the exposure levels of EMF

The quantities which are important from the aspect of EM safety standards are shown in table 1 [5].

Table I
Electric, magnetic, electromagnetic, and dosimetry quantities and corresponding SI units

Quantity	Sym-bol	Unit
Current density	J	Ampere per square meter (A m^{-2})
Electric field strength	E	Volt per meter (V m^{-1})
Magnetic field strength	H	Ampere per meter (A m^{-1})
Magnetic flux density	B	Tesla
Power density	S	Watt per square meter (W m^{-2})
Specific energy absorption	SA	Joule per kilogram (J kg^{-1})
Specific energy absorption rate	SAR	Watt per kilogram (W kg^{-1})

To measure the impact of EMF on the human body at frequencies up to 10 GHz, the quantities: Electric field strength E [V/m], Magnetic field strength H [A/m] and Specific energy absorption rate SAR [W/kg] are most often used. SAR is defined as the rate of power dissipation normalized by the material density. At frequencies higher than 10 GHz, the field penetration depth in the tissues is small. Therefore, the power density S [W/m^2] is a more appropriate quantity than SAR.

Standards related to the limitation of human exposure to EMF determine either the limits of emissions radiating from the devices or the limits of the exposure of people originating from all devices in the environment [6]. They are divided into exposure standards, emission standards and measurement standards. Exposure standards are basically personal protection standards and refer to the maximum levels of human exposure from any number of radiating devices. Emission standards set

different specifications for electrical devices and are based on engineering solutions, such as optimizing performance.

The International Commission on Non-Ionizing Radiation Protection (ICNIRP) issued guidelines on exposures in 1998 covering the frequency range up to 300 GHz [5]. The ICNIRP 1998 guidelines have been endorsed by WHO, the International Labor Office (ILO) and the International Telecommunications Union (ITU), and they have been adopted as their national standard by more than 50 countries worldwide.

In Bulgaria Ordinance No 9 [7] of the Ministry of Health from 14.03.1991, limit levels of strength and power density of EMF in populated territory are defined. In addition to the limit values of EMF in populated areas, there are defined safety zones around the radiating objects. In the presence of more than one radiation source in a frequency range - up to 300 MHz, the geometric sum of the strength of the EMF must be smaller than the limit level. If all emitters operate in the range of 0.3 to 30 GHz, the arithmetic sum of the power density of EMF must be less than the limit permissible level 10 μW/cm2. If there are multiple sources of electromagnetic radiation operating at different frequencies in the area, the condition in Formula 1 must be satisfied:

$$\left(\frac{E_1}{E_{H_1}}\right)^2+\left(\frac{E_2}{E_{H_2}}\right)^2+\ldots+\left(\frac{E_n}{E_{H_n}}\right)^2+\frac{S_{sum}}{10}\le 1 \qquad (1)$$

where E_1, E_2,…, E_n are the electric field strengths, created by the individual emitters in different frequency bands or aggregated strengths emitters of the same band at a transmitting frequency below 0,3 GHz; E_{H1}, E_{H2} ..... E_{Hn} – are the limit levels for the band; S_{sum} is the total power density of emitters with operating frequency exceeding 0,3 GHz.

Limit levels of strength and power density of EMF in Bulgaria are shown in Table II.

TABLE II
LIMIT LEVELS OF STRENGTH AND POWER DENSITY OF EMF IN BULGARIA

№	Transmitter Frequency range	Limit levels of electrical field strength (E) and power density (S) of EMF
1.	30 - 300 kHz	25 V/m (E)
2.	0,3 - 3 MHz	15 V/m (E)
3.	3 - 30 MHz	10 V/m (E)
4.	30 - 300 MHz	3 V/m (E)
5.	0,3 - 30 GHz	10 μW/cm^2 (S)

The provisions of Directive 2013/35 / EU of the European Parliament and of the Council of 26 June 2013 on minimum health and safety requirements regarding the exposure of workers to the risks arising from physical agents (electromagnetic fields) are introduced in the Ordinance No RD-07-5 (2016) in Bulgaria [8]. This Directive specifies the exposure limits for health effects, the exposure limit values for sensitivity effects and the values for action by employers.

In the Republic of Serbia, the conditions and measures for the protection of human health and the environment from the harmful effects of non-ionizing radiation are regulated by the Law on Non-Ionizing Radiation Protection, (Official Gazette of the Republic of Serbia No. 36/2009) [9]. According to Art. 6 (6) (1) of this Law, a Rulebook on the limits of exposure to non-ionizing radiation (Official Gazette of the Republic of Serbia No 104/2009) [10] defined the reference boundary exposure levels of electric, magnetic and electromagnetic fields (0 Hz to 300 GHz) in Republic of Serbia, Table III.

In the case of simultaneous exposure of fields with different frequencies, they must respond the following criteria of Formula 2 in terms of the main limitations.

$$\sum_{i=100kHz}^{1MHz}\left(\frac{E_i}{c}\right)^2+\sum_{i>1MHz}^{300GHz}\left(\frac{E_i}{E_{L_i}}\right)\le 1 \qquad (2)$$

where E_i is the field strength measured at frequency i; $E_{L,i}$ is the reference field of the electric field in Table III; c is 87/f$^{1/2}$ V/m.

III. TESTBED DESCRIPTION AND EXPERIMENTAL SET-UP

For the purpose of the experiment, a simple communication system that simulates the operation of GSM communication was made. A signal generator MXG N5182B KEYSIGHT, which operates at frequencies from 9 kHz to 3 GHz, was used as the transmitter. The generator was set to a frequency of 1200 MHz, with GSM modulation. The output signal was set up to 20 dBm (0,1W), and was emitted by an antenna. The antenna was a Comet SMA-703 - three-band. The antenna bands are 144 MHz, 400 MHz and 1200 MHz, with a gain of 3.4 dBi at 1200 MHz. The measurements are carried out in the antenna far field. The receiver was implemented with USRP NI2920, connected to a PC via Gigabit Ethernet. The strength of the electromagnetic field at various distances from the generator was measured by Selective Radiation Meter SRM 3000 Narda, Fig 1, which works at frequencies 100 kHz to 3 GHz [11].

Fig.1 Selective Radiation Meter SRM 3000 Narda

A. Results

In the "**Spectrum Analysis**" mode of the selective radiation meter, the field strength curves E [V/m] were taken depending on the frequency. The data is recorded in Excel files presented in Table IV for each measurement. File data is processed with MATLAB and displayed in Fig 2. The series of field strength curves show how the signal attenuates with distance.

In the SRM 3000 "**Safety Evaluation**" mode field strength was measured throughout the full operating range of the device: selected from the menu. For each regulated band, the field is from 47 MHz to 2,5 GHz. For this purpose, Full Band EU was

TABLE III
REFERENCE BOUNDARY LEVELS FOR THE EXPOSURE OF THE POPULATION TO ELECTRIC, MAGNETIC AND ELECTROMAGNETIC FIELDS (0 HZ TO 300 GHZ) IN SERBIA

Frequency f	Electric field strength E (V/m)	Magnetic field strength H (A/m)	Magnetic flux density B (μT)	Power density (equivalent plane wave) S_{ekv} (W/m^2)	Time average t (minutes)
< 1 Hz	5 600	12 800	16 000		*
1-8 Hz	4 000	12 800/f^2	16 000/f^2		*
8-25 Hz	4 000	1 600/f	2 000/f		*
0,025-0,8 kHz	100/f	1,6/f	2/f		*
0,8-3 kHz	100/f	2	2,5		*
3-100 kHz	34,8	2	2,5		*
100-150 kHz	34,8	2	2,5		6
0,15-1 MHz	34,8	0,292/f	0,368/f		6
1-10 MHz	34,8/ f$^{1/2}$	0,292/f	0,368/f		6
10-400 MHz	11,2	0,0292	0,0368	0,326	6
400-2000 MHz	0,55 f$^{1/2}$	0,00148 f$^{1/2}$	0,00184 f$^{1/2}$	f/1250	6
2-10 GHz	24,4	0,064	0,08	1,6	6
10-300 GHz	24,4	0,064	0,08	1,6	68/f^{1,05}

measured, the data is automatically summed and saved to a file. If there are levels measured in other, unregulated bands, they are added together and marked on a separate line in the table. The data from the selective radiation meter are presented in Table V. The results are compared with the limit values given in [7] and [10].

TABLE IV
DISTANCE OF THE MEASUREMENT

File	Distance [m]
1,1	0,3
2,1	5
3,1	8, behind one partition wall
4,1	12, behind two partition walls
5,1	15, behind two partition walls
6,1	8, behind two partition walls
7,1	15, behind three partition walls
8,1	20, behind three partition walls

The calculations show that in our experiment, in the situation of simultaneous exposure to fields of different frequencies, the criteria given by equations (1) and (2) are not fulfilled and the total exposure is higher than the limit given in national guidelines. In both cases, sums on the left side shouldn't be higher than 1. For the calculations of Formula 1, the value of the sum is 2,676. If we apply Formula 2, the value of the sum is 1,702. It should also be noted that these values are calculated only for the regulated frequency bands. If we would account the exposure to EMF in the other bands, the results would be even worse. For instance, the sum of the power densities for the regulated frequency bands is 0,134 W/m2 and the power density of other bands is 0,122 W/m2, giving the total value of power density of 0,256 W/m2.

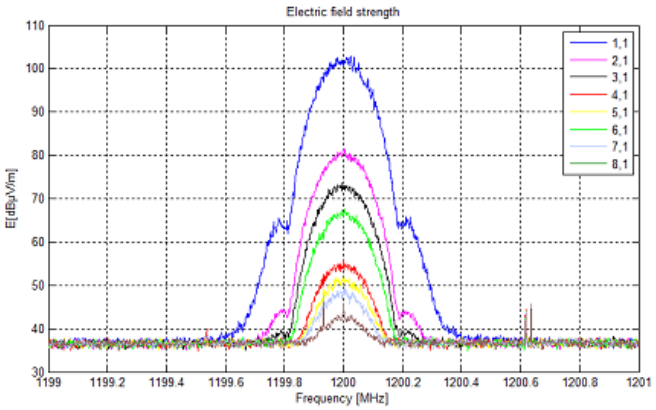


Fig.2 Series of field strength curves on the frequency 1200 MHz for different distances from the generator

B. Recommendations

The recommendations for working with communication equipment, can be given in several directions:

- When installing new equipment, test it for electromagnetic radiation.
- Optimize the operator's time for working with communication equipment as the energy and power limit values absorbed by a unit of body mass in the person body as a result of exposure to electrical and magnetic fields are averaged over 6 minutes, (see Table III). If this is not possible - optimize the location of the equipment inside the laboratory. The relevant distance could be calculated using the Friis Transmission Formula.
- Introduce periodic inspection of the equipment and the levels of electromagnetic pollution inside the laboratory.
- Provide written instructions for work in the laboratory in a visible location.

TABLE V
SUMMARY OF FIELD STRENGTH MEASUREMENT RESULTS IN FULL BAND EU

Service Table Name *Full Band EU*
Total Value [W/m²] *0,25624001*

Measured values		Frequency band of interest				Reference levels Serbia		Calculations	Reference levels Bulgaria		Calculations	
Electric field strength E [V/m]	Power density S [W/m^2]	Name	Low Frequency [MHz]	High frequency [MHz]	Average frequency [MHz]	Electric field strength E [V/m]	Power density S [W/m2]	E_i/E_{Li}	Electric field strength E [V/m]	Power density S [W/m2]	S_{cym}	$(E_i/E_{Hi})^2$
3,3006	0,029377	TV Band I	47	68	57,500	11,2	0,326	0,2946964	3			1,21044
2,0893	0,010888	FM-Radio	88	108	97,750	11,2	0,326	0,1865446	3			0,485019
1,8789	0,0094249	Mid Wave	137	165	151,000	11,2	0,326	0,1677589	3			0,392252
0,9639	0,0025211	Paging	165	174	169,500	11,2	0,326	0,0860589	3			0,103225
2,0743	0,011211	BandIII (DVB-T)	174	230	202,000	11,2	0,326	0,1852054	3			0,47808
0,1398	9,01E-05	Trains	467	468	467,875	11,8967	0,3743	0,0117511		0,1	9,01E-05	
2,8035	0,020927	BandIV (DVB-T)	470	790	630,000	13,8049	0,504	0,2030802		0,1	0,020927	
1,1366	0,0035092	BandV (DAB)	790	862	826,000	15,8071	0,6608	0,0719043		0,1	0,0035092	
0,2605	0,0001619	GSM-R	876	880	878,000	16,2971	0,7024	0,0159826		0,1	0,0001619	
0,9994	0,0027008	GSM 900	890	960	925,000	16,7276	0,74	0,0597432		0,1	0,0027008	
0,8953	0,002097	L-Band (DAB)	1 452	1 492	1 472,000	21,1017	1,1776	0,0424298		0,1	0,002097	
1,8448	0,0090142	GSM 1800	1 710	1 880	1 795,000	23,3021	1,436	0,0791689		0,1	0,0090142	
0,6621	0,0012613	DECT	1 880	1 900	1 890,000	23,9108	1,512	0,0276909		0,1	0,0012613	
1,7669	0,0082797	UMTS-TDD	1 900	2 025	1 962,500	24,3651	1,57	0,0725178		0,1	0,0082797	
1,4478	0,005698	UMTS DL	2 110	2 170	2 140,000	24,4000	1,6	0,0593361		0,1	0,005698	
2,3304	0,014298	W-LAN	2 400	2 484	2 441,750	24,4000	1,6	0,0955082		0,1	0,014298	
1,0456	0,0028911	ISM	2 484	2 500	2 491,750	24,4000	1,6	0,0428525		0,1	0,0028911	
6,7321	**0,12189**	**Others Value**									0,0709	2,669016
	0,2562403							*1,7022*			*2,6761*	

IV. CONCLUSION

The paper describes the measurement of EMF in a communication laboratory. The results show increased levels of EMF power density. In the measurements made, in addition to electric field strength and power density in regulated frequency bands, there is also a significant contribution to the total EM pollution originating from other, unregulated bands. It is difficult to determine the source of this field - it may be due to different sources of signal in the environment. Whether a communication equipment represents a risk to human health depends on the distance between it and the worker and on its time of service. The levels of electromagnetic pollution inside the laboratories should be regularly monitored and, if necessary, the locations of the transmitting equipment could be optimized, in order to avoid a risk for human health.

ACKNOWLEDGEMENT

This research work is supported by the Erasmus+ mobility program and by the project III43012, funded by the Ministry of Education and Science of the Republic of Serbia.

REFERENCES

[1] Directive 2013/35/EU of the European Parliament and of The Council
[2] James C. Lin, Safety Standards for Human Exposure to Radio Frequency Radiation and Their Biological Rationale, IEEE Microwave Magazine, December 2003
[3] Donald Hillman, Exposure to Electric and Magnetic Fields (EMF) Linked to Neuro-Endocrine Stress Syndrome: Increased Cardiovascular Disease, Diabetes, & Cancer, Shocking News, No. 8, November 2005
[4] B. Blake Levitt, Henry Lai, Biological Effects from Exposure to Electromagnetic Radiation Emitted by Cell Tower Base Stations and Other Antenna Arrays, NRC Research Press Web site, 5 November 2010, Environ. Rev. 18: 369–395
[5] International Commission on Non-Ionizing Radiation Protection, Guidelines for Limiting Exposure to Time-varying Electric, Magnetic, and Electromagnetic Fields (Up to 300 GHz), Health Physics April 1998, Volume 74, Number 4
[6] World Health Organization 2006, Framework for Developing Health-Based EM F Standards
[7] Ordinance No 9 of March 14, 1991, on Limitable Levels of Electromagnetic Fields in Populated Territories and the Determination of Hygienic Protection Areas around the Radiation Sites of the Ministry of Health of the Republic of Bulgaria
[8] Ordinance No RD-07-5 of 15 November 2016 on Minimum Requirements for Health and Safety of Workers in Risks Related to Exposure to Electromagnetic Fields of the Ministry of Labor and Social Policy and the Ministry of Health of the Republic of Bulgaria
[9] Law on Non-Ionizing Radiation Protection. ("Official Gazette of the Republic of Serbia", no. 36/09)
[10] The Rulebook on the Limits of Exposure to Non-Ionizing Radiation ("Official Gazette of the Republic of Serbia", no. 104/09)
[11] SRM-3000 Selective radiation meter, Operating Manual

RADIO COMMUNICATIONS, MICROWAVES, ANTENNAS – ORAL SESSION

Experimental Shielding Effectiveness Analysis of Impact Hexagonal Air-vent Distances on a Metal Enclosure

Nataša Nešić[1], Nebojša Dončov[2], Slavko Rupčić[3], Bratislav Milovanović[4] and Vanja Mandrić Radivojević[5]

Abstract – **In this paper the shielding effectiveness of metal enclosure with twelve hexagonal air-vents are experimentally studied. Therefore, two scenarios of hexagonal groups formed in 3x4 and 4x3 order are conducted. The experimental procedure is carried out in a semi-anechoic room. All measurement results are compared to the numerical ones obtained by TLM method with incorporated compact air-vent model and wire model.**

Keywords – **Shielding effectiveness, TLM hexagonal air-vent model, TLM wire model, Experimental analysis.**

I. Introduction

To get protected from external influences, electronic equipment is usually placed inside some protective enclosures. The effect of these enclosures is two-fold. On the one hand, it is necessary to protect the electronic equipment inside enclosure from the outer radiation. On the other hand, the electromagnetic (EM) field which is generated from the electronic equipment inside enclosure should be reduced, as much as it is possible [1]. Commonly, shielding enclosures are made of a highly conductive material in order to reduce not only external EM fields impact on equipment inside the enclosure, but also a level of internal EM interferences emitted from equipment to outside space. Moreover, on the enclosures there usually exist the airflow aperture arrays of different shapes, so-called air-vents, as well as some other slots, which are intended for heating dissipation and airing [2].

To determine the shielding effectiveness (SE) characteristic of a metal enclosure over a frequency range, there exist several different methods, such as the analytical, the numerical and the experimental ones. There are many numerical methods that can be employed, e.g. Finite Difference Time Domain (FDTD) method in [3], Method of Moments (MoM) [4] and Transmission Line–Matrix (TLM) [5], etc. In the experimental methods, an antenna is set inside the enclosure in order to determine its SE. Usually, the monopole [6], and the dipole antenna [7] are used to determine the electric field level as opposite to the loop antenna which is appropriate to determine the magnetic field [6]. In [8] the impact of the physical dimensions of receiving antenna on the resonant frequency is also analyzed by using the TLM method.

The objective of this paper is to experimentally quantify the SE of a metallic enclosure with groups of 4x3 hexagonal air-vents on the removable front panel for one scenario and groups of 3x4 hexagonal air-vents for another scenario. Afterwards, these measurement results are compared to the numerical results obtained by the TLM method with incorporated compact wire model [8], [9] and hexagonal air-vent model [10], [11]. Namely, in the numerical simulations, the physical presence of a monopole-receiving antenna, used in measurements, is taking into account by the wire model while aperture arrays are described by the air-vent model.

The paper is organized as follows. In Section II, experimental procedure of measuring shielding effectiveness with tracking generator and spectrum analyzer is outlined. Section III and IV are engaged to the physical enclosure model and the numerical one. Section V provides discussion of the results. Finally, Section VI summarizes the work.

II. Experimental Procedure

Usually, anechoic rooms are used to measure EM radiation of electronic equipment, as well as to test immunity. For instance, resistance of equipment to external radiation or resistance of equipment under test (EUT) irradiated by a controlled plane wave can be measured. Since interior space in anechoic room is coated with radio-frequency (RF) absorbent material, internal reflections are minimized and the device being tested is exposed to a precisely determined EM field. To measure the SE of enclosure, a spectrum analyzer or a scalar network analyzer can be used.

The measurements presented in this paper are performed in ordinary laboratory space by using RF absorbers. Also, unnecessary procedure of EM shielded chamber against external EMI is carried out. To determine the SE of enclosure a measuring procedure has to be performed twice, without and with enclosure. The SE is measured by the spectrum analyzer and with the s_{21n} and s_{21e}, parameters. The transmission parameters of the measurement without and with an enclosure are marked as s_{21n} and s_{21e}, respectively. The SE can be defined as in Eq. 1:

$$SE = 20\log_{10}\frac{|S_{21n}|}{|S_{21e}|}. \quad (1)$$

In order to find the best possible minimal interference in a measurement chamber, both transmitting antenna and enclosure are rotated. Since when the curve had sufficient low

[1]Nataša Nešić is with the College of Applied Technical Sciences Niš, Aleksandra Medvedeva 20, Niš 18000, Serbia, E-mail: natasa.bogdanovic@vtsnis.edu.rs

[2]Nebojša Dončov is with the Faculty of Electronic Engineering in Niš, University of Niš, Aleksandra Medvedeva 14, Niš 18000, Republic of Serbia, E-mail: nebojsa.doncov@elfak.ni.ac.rs

[4]Bratislav Milovanović is with the University of Singidunum, Danijelova 32, Belgrade 11000, Serbia, E-mail: bmilovanovic@singidunum.ac.rs

[3]Slavko Rupčić and [5]Vanja Mandrić Radivojević and are with Department of Communications Faculty of Electrical Engineering Osijek 31000, Croatia, E-mails: rupcic@etfos.hr vanja.mandric@etfos.hr.

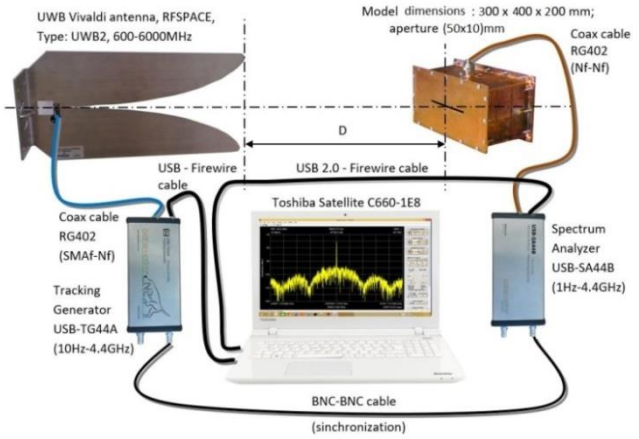
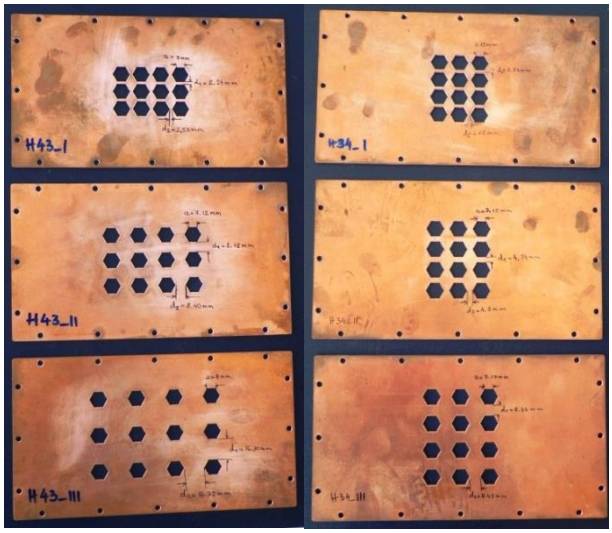


Fig. 1. The measuring configuration used in a semi-anechoic room.



a) b)

Fig. 2. Two sets of front panels with different air-vent distances; *a*) **H43**air-vent layouts, *b*) **H34** air-vent layouts.

ripple of the s_{21} parameter, the measurement process is started. The enclosure is placed in far field against the transmitting antenna, in the whole observed frequency range [12].

The measurement processes are performed by using the spectrum analyzer with tracking generator and the SPIKE Software for PC computer, as shown in Fig. 1 [12]. A Vivaldi dipole antenna was used as a transmitting antenna, while an in-house monopole antenna was employed as a receiving one. The monopole, which is made of a copper material with a length of l = 50 mm and radius of r = 0.1 mm, is placed in the middle of the enclosure.

III. Physical Model

The rectangular metal enclosure with internal dimensions of (100 x 100 x 200) mm and with walls thickness of t = 2 mm is manufactured from copper material. This metal enclosure has removable front wall. Two sets of three panels are made. Each of the front walls is perforated with twelve hexagonal aperture arrays which are placed around the middle of a wall, Fig. 2.

In the first scenario, there are three panels of hexagonal air-vents formed in the group of **4x3**, as depicted in Fig. 2a. These front walls are used in experimental measurements. In order to make it easier to describe distances between any two hexagonal apertures, in both the vertical, d_1, and the horizontal axes, d_2, panels are labeled in the following order:

H43_I – width side of hexagon is a = 7.0 mm, aperture distances d_1 = 2.24 mm, d_2 = 2.53 mm and cov = 0.5348.

H43_II – width side of hexagon is a = 7.12 mm, aperture distances d_1 = 8.18 mm, d_2 = 8.40 mm, and cov = 0.2833.

H43_III – width side of hexagon is a = 7.04 mm, aperture distances d_1 =16.30 mm, d_2 = 16.32 mm and cov = 0.1487.

For the second scenario, the same number of hexagonal air-vents is formed in the group of **3x4**. The panels are shown in Fig. 2b. In the following, more details about parameters of this group:

H34_I – width side of hexagon is a = 7.12 mm, aperture distances d_1 = 2.33 mm, d_2 = 2.62 mm and cov = 0.5108.

H34_II – width side of hexagon is a = 7.15 mm, aperture distances d_1 = 4.24 mm, d_2 = 4.5 mm and cov =0.4250.

H34_III – width side of hexagon is a = 7.17 mm, aperture distances d_1 = 8.33 mm, d_2 = 8.45 mm and cov = 0.2826.

IV. Numerical Simulation

In this paper, the TLM method is employed for numerical simulations. According to the physical enclosure model a numerical one is created. A compact air-vent model was used to model group of hexagonal shaped apertures. A TLM wire model was used to model a monopole-receiving antenna inside the enclosure.

The compact TLM air-vent model consists of two reactive circuits per propagation direction, which are placed between two neighboring TLM cells. The TLM cells coincide with the position of perforated metal wall. Firstly, it was developed for square and circular perforations on thin metal walls and on walls of significant thickness [10]. Afterwards, it was extended for rectangular and hexagonal air-vents [11].

The compact TLM wire model is employed for modeling an antenna inside enclosure. The main purpose of a receiving antenna is to measure the EM field level and its distribution. A receiving antenna is modeled as a wire segment which is incorporated into the Symmetrical Condensed Node (SCN) [8], [9]. The impedances of additional wire network link and short-circuit stub lines depend on the space used and time-step discretization, and also on per-unit length wire capacitance and inductance [11]. Two-way coupling between signal in the wire circuit and external EM field is described by pulses in transmission line network of SCN [10], [11].

In the considered model, the wire is connected to the ground via resistor R. The current which is induced on the wire, due to external EM field, generates voltage on the resistor R, which is loaded at wire base, and this allows measuring the level of EM field.

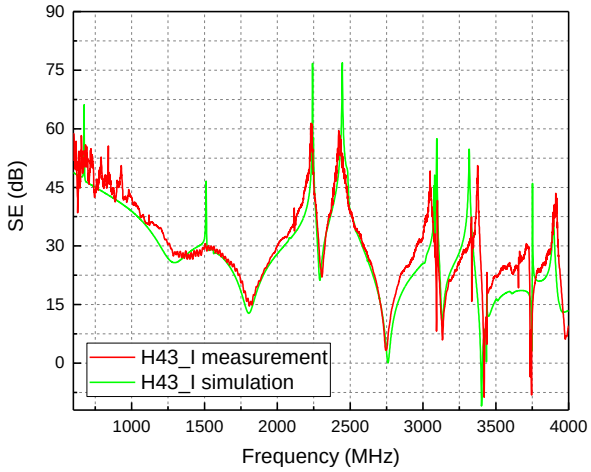


Fig. 3. Compared SE curves of **H43_I** enclosure model.

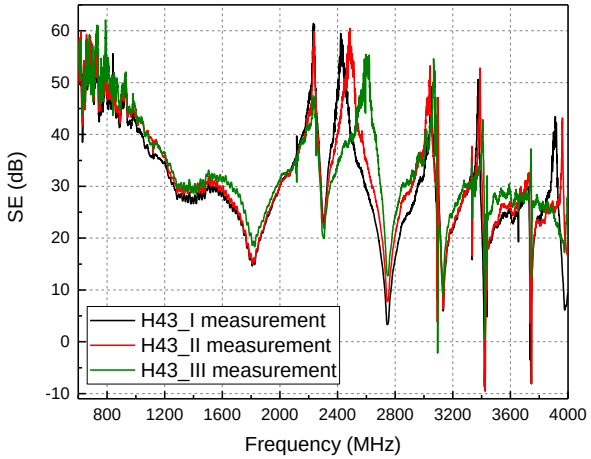


Fig. 4. Compared measurement SE curves for **H43** enclosure scenario against vertical incident plane wave.

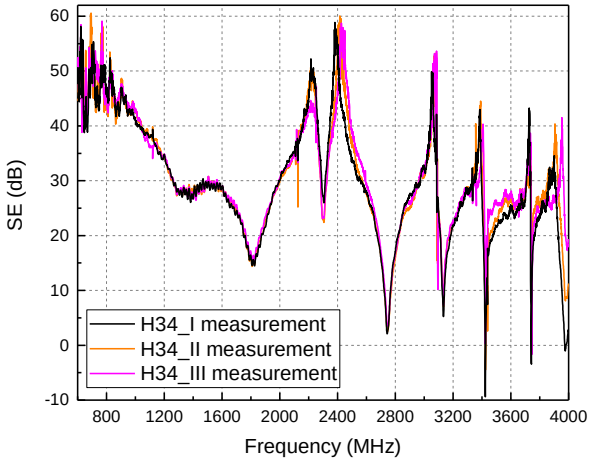


Fig. 5. Compared measurement SE curves for **H34** enclosure scenario against vertical incident plane wave.

The numerical TLM enclosure model is created according to dimensions, thickness, material and distance between air-vents like in physical one. The enclosure is excited by vertically polarized incident plane wave.

The SE characteristic is calculated as logarithmic ratio of the electric field without and with the enclosure, in the same probe point. A monopole-receiving antenna is placed in the middle of the enclosure to detect the electric field level. In the case of monopole antenna in free space, the size of a ground plane is chosen to be as the size of the enclosure wall when monopole antenna is placed inside the enclosure [10].

V. Discussion of Results

In this section, we discuss about the experimental and the numerical SE results obtained for the considered enclosure. The analyses are carried out for two scenarios. In both, the front enclosure wall is perforated with 12 hexagonal air-vents. However, distances between the air-vents differ among panels. Both the numerical and the experimental analyses are conducted in order to improve the effectiveness of enclosure. The excitation toward the enclosure is normal incident plane wave vertically polarized. The frequency range of interest is from 600 MHz up to 4 GHz.

To start with, we consider the same enclosure model with three hexagonal air-vent layouts. It can be seen that the frontal panels differ between each other due to aperture distances, as illustrated in Fig. 2a. In the first scenario, the air-vents are arranged in groups of 4x3 around the center of frontal enclosure wall. Parameters of layouts entitled by **H43_I**, **H43_II** and **H43_III** are given in details in Section III. For instance, the first panel under the title **H43_I** has aperture distances mutually separated by the distance of d_1 = 2.24 mm and d_2 = 2.545 mm in vertical and horizontal directions, respectively. Therefore, the percentage of surface area covered by apertures is equal to 53.48 %. In numerical model, to describe the part of the perforated wall, the compact TLM air-vent model is employed. Figure 3 presents the SE results obtained by the measurement and the numerical simulation of the enclosure model **H43_I**. It can be observed an excellent agreement between compared shielding characteristics. Also, the experimental and the numerical SE curves which are obtained for other layout cases fit very well, so there is no need to display them.

Figure 4 illustrates the comparison of SE measurement curves for three different **H43** enclosure layouts. It can be noticed that all SE curves match very good at resonant frequencies. However, the SE levels are slightly higher with increasing in the distances between air-vents. Although, the resonances are almost the same, the SE peaks are shifted toward the higher frequencies, at frequencies around 2.4 GHz. This effect is occurred due to the reducing the coverage of perforated wall. Namely, the coverage in descending order is considered, from 53.48% up to 14.87 %.

The second scenario presents the air-vents in groups of 3x4 around the center of front wall of considered enclosure. The front panels are depicted in Fig. 2b. In this scenario, the distances between air-vents and the coverage of panels

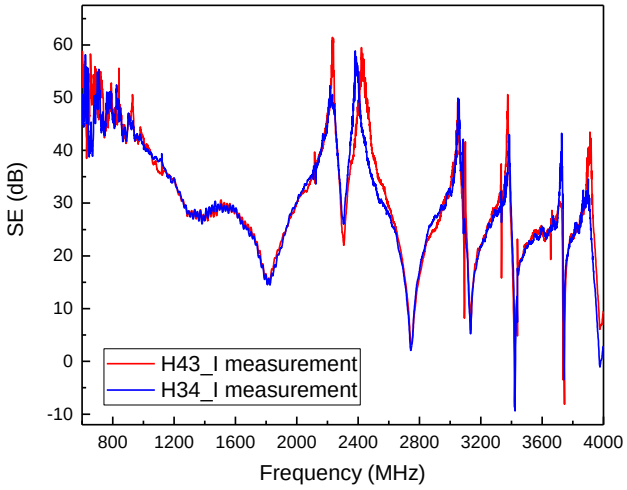


Fig. 6. Compared measurement SE curves for **H43_I** and **H34_I** enclosure models.

entitled **H34_I**, **H34_II** and **H34_III** are given in Section III. Like in the previous scenario, the measurement SE curves fit very well in comparison to the SE simulation results. Therefore, only experimental SE results are mutually compared for three cases, and they are presented in Fig. 5. It can be noticed that the SE characteristics are not significantly changed with increasing in distance between the air-vents. It can be explained that there are not a big difference in value between **H34_I** and **H34_II** coverages. However, due to decreasing in the coverage up to 0.2826 (H34_III), the levels of SE differ a bit at higher frequencies, especially at peaks. Beside this, the SE characteristics have the same values of the resonant frequencies for all three considered **H34** enclosure layouts. Detailed theoretical analysis of these issues is described in [13], [14].

Further, comparison between measurement SE curves of **H43_I** layout and H34_I layout are presented in Fig. 6. It shows excellent matching curves for almost the same spacing between the air-vents and similar coverage, even when the air-vent arrangement is turned; in one case it is 3x4 and the other one is 4x3.

VI. Conclusion

To conclude, the experimental results obtained for the considered enclosure are in an excellent match with the numerical ones. The TLM method incorporated with the TLM wire model and the TLM air-vent model is accurate and fast computer tool for efficient numerical characterization of EM problems on the field of EMC. It can be seen a remarkable matching between compared methods, especially at the resonance frequencies. Also, detailed analysis of the measured and numerical results shows very similar SE characteristics. It can be concluded that for a similar coverage value, even when the arrangement of the air-vents are reversed, the SE curves are quite similar.

Acknowledgement

This work has been partially supported by the Ministry for Education, Science and Technological Development of Serbia; project number TR32052, by the EUROWEB+ project and by the COST IC 1407 Action.

References

[1] C. Christopoulos, *Principles and Techniques of Electromagnetic Compatibility*, 2nd ed., CRS Press, 2007.

[2] M. Li, J. Nuebel, J. L. Drewniak, R. E. DuBroff, T. H. Hubing, T. P. Van Doren, "EMI from Airflow Aperture Arrays in Shielding Enclosures–Experiments, FDTD, and MoM Modelling", IEEE Trans. on EMC, vol. 42, no. 3, pp. 265-2275, 2000.

[3] M. Li, J. Nuebel, J. L. Drewniak, R. E. DuBroff, T. H. Hubing, T. P. Van Doren, "EMI from cavity modes of shielding enclosures – FDTD modeling and measurements", IEEE Trans. EMC, vol. 42, no. 1, pp. 29–38, 2000.

[4] S. Ali, D. S. Weile, T. Clupper, "Effect of near field radiators on the radiation leakage through perforated shields", IEEE Trans. on EMC., vol. 47, no. 2, pp. 367–373, 2005.

[5] B. L. Nie, P. A. Du, Y. T. Yu and Z. Shi, "Study of the shielding properties of enclosures with apertures at higher frequencies using the transmission-line modeling method", IEEE Trans. on EMC, vol. 53, no.1, pp. 73–81, 2011.

[6] P. M. Robinson, M. T. Benson, C. Christopoulos, F. J. Dawson, D. M. Ganley, C. A. Marvin, J. S. Porter, P. W. Thomas, "Analytical formulation for the shielding effectiveness of enclosures with apertures", IEEE Trans. on EMC, vol. 40, no. 3, pp. 240–248, 1998.

[7] J. Shim, D. G. Kam, J. H. Kwon, J. Kim, "Circuital modeling and measurement of shielding effectiveness against oblique incident plane wave on apertures in multiple sides of rectangular enclosure", IEEE Trans. EMC, vol.52, no.3, pp. 566–577, 2010.

[8] N.J.Nešić, N.S.Dončov, "Shielding Effectiveness Estimation by using Monopole-receiving Antenna and Comparison with Dipole Antenna", DE GRUYTER Frequenz, Vol. 70, Issue 5-6, pp. 191–201, April 2016.

[9] N. Nešić, B. Milovanović, N. Dončov, V. Mandrić-Radivojević and S. Rupčić, "Improving Shielding Effectiveness of a Rectangular Metallic Enclosure with Aperture by Using Printed Dog-bone Dipole Structure," 52nd Int. Scien. Conf. ICEST, pp. 97-100, Serbia, Niš, June 28-30, 2017.

[10] N. Doncov, A. J. Wlodarczyk, R. Scaramuzza, V. Trenkic, "TLM modelling of perforated metal screens," 4th Int. Conf. Computational Electromagnetics, Bournemouth, UK, 2002.

[11] N. Doncov, B. Milovanovic, Z. Stankovic, "Extension of compact TLM air-vent model on rectangular and hexagonal apertures," ACES Journal, vol. 26. No.1, pp. 64-72, 2011.

[12] V.Mandrić-Radivojević,, S.Rupčić, N.Nešić, V. Alilović, "The shielding effectiveness measurements of a rectangular enclosure perforated with slot aperture", Proc. of SST 2017, pp.121-126, October 18-20, Osijek, Croatia.

[13] N. J. Nešić / N. S. Dončov, "Analysis of TLM Air-vent Model Applicability to EMC Problems for Normal Incident Plane Wave", *Telfor Journal*, t. 8, br. 2, pp. 104-109, 2016.

[14] N. J. Nešić "Numerical and experimental analysis of aperture arrays impact on the shielding effectiveness of metal enclosures in microwave frequency range", Doctoral thesis, in serbian, Singidunum University, Belgrade, 2017.

EVOLUTION OF MOBILE NETWORKS TO NEXT GENERATION (5G)

Petar Nikolov[1], Desislav Petkov[2], Agata Manolova[3], Nicole Christoff[4], Ilia Iliev[5]

Abstract – **Mobile networks are already an integral part of people's lives. The world is becoming more and more connected and we need a network that will support the development of the technologies of the future - autonomous cars, virtual reality and the Internet of Things (IoT). This article presents a fifth generation (5G) wireless network that will unify all existing standards and allow the development of technologies such as mobile cloud, cognitive radio, virtualization, software-defined networks. The network will be able to transfer more data at higher speeds and will also be able to get through obstacles, which increases scope. The technical characteristics and requirements of the network, the need for resources, frequency spectrum are considered.**

Keywords – **5G network, mobile cloud, cognitive radio, resource, speed.**

I. INTRODUCTION

The 4th generation networks already have been developed in Bulgaria and now the mobile industry points toward the next generation network - 5G. This year the 5G networks and the mobile wireless connections were the accent on the Mobile World Congress held in Barcelona, Spain. Seen on a broader scale every 10 years new generation cellular network has been developed [1]. Following this tendency it is expected that a new generation would show up in 2020 where in comparison with 4G the planned the max peak data rate will be 20 Gbit/s [2]. Technical characteristics comparison of mobile cellular generations related to data speed are shown in Table 1.

TABLE I
TECHNICAL CHARACTERISTICS COMPARISON OF MOBILE CELLULAR GENERATIONS

Generation	Third generation 3G (IMT-2000)	Fourth generation 4G (IMT-Advanced)	Fifth generation 5G (IMT-2020)
Max data rate	2 Mbit/s	1Gbit/s	20Gbit/s
Technology	WCDMA	OFDMA IP-ядро	OFDMA IP-core, combination of LAN and WLAN
Multiplexing	CDMA	CDMA OFDM	CDMA OFDM
Handover	Horizontal plane	Horizontal and vertical plane	Horizontal and vertical plane

[1,2,3,4,5] The authors are with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: nikolovpetar33@gmail.com, desislav_petkov@yahoo.com, amanolova@tu-sofia.bg, nicole.christoff@tu-sofia.bg, igiliev@tu-sofia.bg

The fifth generation mobile system differs from the previous third and fourth generations in that it represents a heterogeneous network integrating 4G, WiFi and other wireless technologies. It combines cloud infrastructure [3], a virtualized network core, intelligent end-to-end services, and a high-performance distribution model designed to transfer data generated by milliard mobile devices, machines, and sensors. The fifth generation will be a prerequisite to transform the mobile network from a man-centric to a much broader network, which will include many machines, sensors and other interactive things that will lead to milliard mobile devices. High data rate and intelligence will be the main features of the 5G network. The quality of service and service priority in the 5G networks will be on an extremely high level and the network itself will determine this automatically. The 5G networks will be implemented with applications for data analysis, traffic prediction and network monitoring to save time, high efficiency and network intelligence. The system will be able to optimize and customize the capabilities of each user application.

II. 5TH GENERATION NETWORK

A. Requirements

Enhanced data rates - The 5th generation of mobile cellular systems will achieve thousands of times higher data rates to current systems. These networks will support broadband channels of several tens of Gbit/s to ensure customers with Gbit service everywhere and any time.

Reduced network latency - Due to inclusion of variety of devices with remote access, such as robots, machines, automotive, and others, will require a highly reliable wireless communication channel with extremely low latency time.

Massive number of connections - With the emergence of the 5th generation, the number of connected devices is expected to increase at a very fast pace. Also, the amount of data transferred over the mobile network will be tremendously increased.

High reliability - The 5th generation mobile systems will be designed to be highly reliable and available with 0% service outage. This is a prerequisite for maintaining a variety of critical services that are not yet available in current networks. The main mobile network goals are to ensure reliable end-to-end service and zero service outages.

High efficiency and performance - One of the key factors for high network performance in 5th generation is the infrastructure of very small cells and high energy efficiency. By definition the small cells are low power wireless access points that operate in the licensed spectrum and are managed by the operator to provide improved cellular coverage, capacity and services for home and business users in urban and suburban areas. Small cells can differ on their coverage areas - the smallest size one is the Fempto-cell, followed by Pico-cell, Micro-cells, and the largest of which is Metro-cell for densely populated urban areas.

B. Network Architecture

Cloud services and shared cloud-based resources in their development period include many different variations of resources that can be shared within a particular cloud or between interconnected clouds. In both cases, we can distinguish resources by the following features - hardware or software, limited or unlimited, case-specific or common. Number of other alternatives exist that can form a group of virtual resources shared in the cloud. Examples of cloud-types are - clouds for so-called computing resources (processor, memory and network), storage clouds, game clouds, etc.

User resources - The user itself is not physically part of the cloud but the resources we look at are owned, managed and controlled by the user. Based on the level of integration in social aspects, the user resources can be delimited as follows:

- Individual, where a single user controls one or more devices and adjusts the operational parameters not only for the one device but for all devices also. This can be called a "personal" cloud. Every interaction here is user dependable, but it may not be a single user, but an organization. Knowing or predicting the individual user behavior is extremely useful as it can be applied to realization of common strategies for interoperability between the sets of devices that constitute this "personal" cloud.
- Group levels include multiple users or operators, and here the social aspect is included which requires additional considerations. Knowing the social or group behavior and along with the individual behavior allows to form cooperative strategies to maximize the utility for the whole group and individual members. Generally, individual members will join the cloud if there are benefits that can be realized.
- Universal resources include the common control capabilities of available resources that are in the cloud range. An axiomatic example of such infrastructure would be an emission measurement sensor led by community-government, but still sending short messages to all devices in the cloud.

Software resources - It is necessary to distinguish the operating system from not-serviced software coming with a mobile device on the one hand and the user-supported software e.g. - mobile applications, from another:

- Operating systems determine the overall cloud node operation. Examples of operating systems - Android, Windows, iOS, etc.
- Usually the unsupported software is present on most mobile devices and works on the background for users. Such types of programs are pre-installed on mobile devices to monitor device usage and return data to the vendor.
- Mobile applications can be installed based on user needs. For individual operating systems, a large number of applications are available which can be downloaded and installed by the users on their devices. User-level apps can be shared in the cloud.

The *hardware resources* available to mobile cloud users represent physical devices or components that build the cloud itself. By definition the available resources can be categorized into different groups:

- Computer or computing – processors (CPU), graphic processors (GPU), or special processors for digital signal processing (DSP), etc.
- Storage - operational energy-dependent memory (RAM), non-volatile memory for long-term storage (FLASH), disk arrays (HDD RAID), etc.
- Detectors or sensors - for temperature, pressure and location. Also microphones, cameras, etc.
- Drive mechanisms - servo motors, display, flash, traffic lights and notification semaphore, etc.
- Power - batteries, mobile batteries, solar panels or uninterrupted power from the power grid.

Network resources - As we look at mobile devices as key players to the methodical research cloud models of the future, connectivity as an opportunity to share resources and connectivity as a resource is significant. The common technologies available on mobile devices are as follows:

- Cellular - from 2G to current 4G technologies.
- Wireless LANs - with the emergence of smartphones, the WLAN interface is additionally used to unload the cellular traffic and to transfer large-sized files in ad hoc mode between mobile devices.
- Bluetooth Interface - today's benefit of this technology is reduced power (Bluetooth Low Energy, BLE) needed to transfer small data files to an almost touch-based network approach.
- Infrared (IR) visible light communications - optical aerial interfaces based on radio frequencies. There are manufacturers of smart devices that incorporate IR transceivers into their devices to allow a remote control.
- Cable interfaces may be present on some mobile devices to allow dial-up communications, either directly or by using dongles.

Network coding - with the introduction of mobile clouds the communication architecture will change radically. At present days, the cellular communications architectures design is still dominated by the centralized management links "point to point". Mobile clouds will interrupt this type of structural design, relying on distributed functionality. The mobile cloud is able to retrieve content from multiple sources at the same time and potentially through several air interfaces. Due to these radical changes, the basic communication technology as well as policies will also be changed. Some of the major challenges in using multiple sources and interfaces include:

- the need to coordinate which data packets to be transmitted from each source and / or air interface, which requires a high signal load and
- the fact that productivity will depend heavily on the changing conditions of these sources / interfaces.

To remove these problems the mobile clouds can use network encryption as a key that will enable this technology. The network encryption cuts off with "Save and Forward" conceptual scheme of current networks, where each node in a packet switched network accepts, stores, and transmits bundles without changing their content and replaces that model with a new "Calculation and forwarding". In this new scheme, the packages that enter a node on the network will be stored, but the packets that go out will be generated as combinations of packets that are already stored in the node buffer. This means that an intermediate network node can work on the content of incoming data. On the one hand, this allows destinations to focus on getting enough combinations to recover original data instead of focusing on individual packets. This means that coordination between multiple sources / interfaces is relieved so that each source / interface can transmit different line combinations to the end recipients. This also allows for more robust mechanisms to deal with system dynamics, where data recovery is no longer dependent on a specific packet that slows down or the interface is interrupted after it gets enough of it. On the other hand the network coding fundamentally changes the resource management across the network. While all packets entered a node will leave this node after a while, as they were stored and forwarded, with that the network coding stops and sends (linear) combinations of received packets, allowing the node to send less packets with the same or higher rate than the incoming rate depending on network conditions and topology. Unlike the existing coding strategies based on the principle for "Erase / error" and encoding of the source or channel, network coding is not limited to end-to-end communications. In this way, these features make network coding an extremely important mobile cloud solution.

Network slicing is an important ability of 5G systems for network resource efficiency, flexible deployment and allows a new generation of mobile applications and services. The network slice purpose is to provide all necessary resources that various services require from a mobile network through different logical subnets determined by the service type. The basic services that 5G networks need to support are presented by the ITU in three main groups (use cases): Enhanced Mobile Broadband (eMBB), Ultra Reliable Low Latency Communications (URLLC) and Massive Machine Type Communications (mMTC). Each of these groups has different service requirements such requirements are the bandwidth, the number of terminals in the system, the time-delay of the signal, the velocity, etc.

The 5G network must be adaptable and adjustable according the terminal applications. Network slices represent stand-alone, end-to-end logical subnetworks utilizing common physical infrastructure. They have direct access to the network resources, services, and functions, and autonomously allocate these resources. Terminal devices will be able to switch from one network slice to another and to stay connected to multiple network slices simultaneously.

C. Efficient radio resource management

The cognitive radio is a new technology that has the potential to meet the stringent requirements for spectrum availability in the 5th generation of networks. Cognitive radio is defined as a radio that can adapt its transmission parameters according to the characteristics of the environment that operates. Cognitive radios are equipped with cognitive capabilities and could be reconfigured. In cognitive radio networks two types of users are presented - primary users who are licensed users and have priority over the frequency spectrum and secondary users who are opportunistic users. They have access to the radio spectrum on a non-interactive or leased basis, according to policies agreed with the main users or designated by the regulatory authorities. Frequency spectrum is one of the most important natural resources that is regulated and exhaustible. The rising number of new applications coupled with the demand for strict requirements and for channel capacity the huge demand to realize the 5G networks, because according to today's standards, bandwidth and power consumption would be a hindrance. That is why efforts are currently being concentrated on new communication and networking models that can intelligently and effectively address these issues. Based on the Shannon theory of information capacity it is obvious that the modern propagation models, the modulation techniques and error correction have improved the capacity of current mobile systems reaching the maximum possible. To increase the available bandwidth for data transmission seems to be the most promising approach to increasing the capacity of future mobile networks, including the fifth generation.

Cognitive radio technology can be presented as an opportunity for effective resource management and utilization due to its intelligent and adaptable nature. Cognitive radio networks will be different from traditional communication models it means that radio stations / devices are capable of adapting their operational parameters, such as frequency, transmission power and modulation types with the variations of their working radio environment. The cognitive radio first acquires knowledge of the state of its radio environment. In the first place, it is aware of spectrum, geographic information, transmitted waves, network protocols types, security policies, local available resources and user needs. Based on this contextual platform the cognitive radio distinguishes the most effective strategy to utilize and adapt the transmission parameters in order to use the available technologies on the best way. The typical cognitive cycle is shown in Fig. 1.

The main cognitive radio functions in the cognitive cycle are:

- Radio Frequency Spectrum Observation and Analysis
- distribute and manage the radio frequency spectrum
- Spectrum handoff and mobility.

Spectrum monitoring and analysis - This feature allows cognitive radio to detect a portion of frequency spectrum that is not used by mainstream users. These unused portions are

called spectral white space. The function also monitors any white space used for secondary transmission to be released in case the primary user reappears. The characteristics of the observed radio channels are calculated on the basis of the collected information from the observed module. An efficient algorithm is then used to extract information on spectral conditions in terms of used time and frequency of the spectrum and provides information on the spatial and time availability of a licensed spectrum.

Radio spectrum allocation and management - After the initial processes of spectrum observation and analysis, spectrum allocation, management and retransmission allows secondary users to have the best bandwidth for data transmission and to hop around multiple radio frequency bands depending on the characteristics of the channels changing in time to meet the quality of service (QoS) requirement. The nature of spectrum mobility in cognitive radio networks can be divided into the following categories:

- Spectrum mobility in a time domain where the cognitive radio adapts its working frequency bands to the new available free bands through different time slots.
- Spectral mobility in the spatial domain, where cognitive radio changes its transmitting frequency based on the current geographic area. This means that when it moves from one place to another the operating frequency changes accordingly.

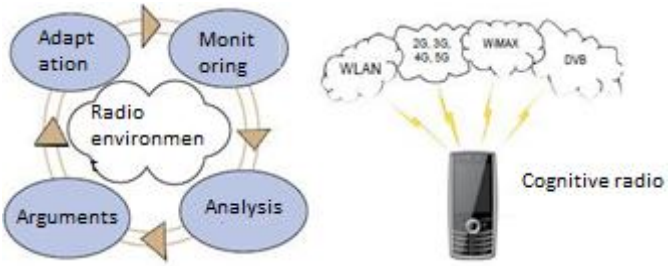


Fig. 1. Cognitive cycle

III. FIFTH GENERATION DEVICES FOR COGNITIVE RADIO

The terminals designed for 5G cognitive radio are intended to be software-defined devices that provide both multi-modes maintenance and effective radio frequency spectrum utilization. The devices will be ready for the fifth generation mobile networks challenges if they have the capabilities to offer the following features: excellent performance, reliability, affordability and ease of personalization and use, ability to deliver a variety of applications. The components of cognitive terminals for 5G are illustrated in Fig. 2. Software-Defined Radio components:

- Geo-locator - the cognitive radio uses the transmitter location provided by a geo-locator (such as a GPS receiver) for making the appropriate decisions.
- Training system - learning capacity is the most important part of cognitive radio. Different justifications and training systems are used to build and apply knowledge about learning states and reactions. In another use of classical artificial intelligence a fuzzy logic system for wave adaptation is used.
- Policy database - The database can be defined and updated by local or global regulatory authorities. It may include the status of the local radio frequency bands and the available channels.
- Sensors measure and monitor radio communications in the environment and collected information is provided to the cognitive core.
- Optimization algorithms - cognitive devices work with different algorithms and technologies to build and adapt radio waves. The sensors collect information from the radio medium that feds to the cognitive radio that is responsible for optimizing and selecting the appropriate frequency, transmit power, routing, performance, and digital error rate.

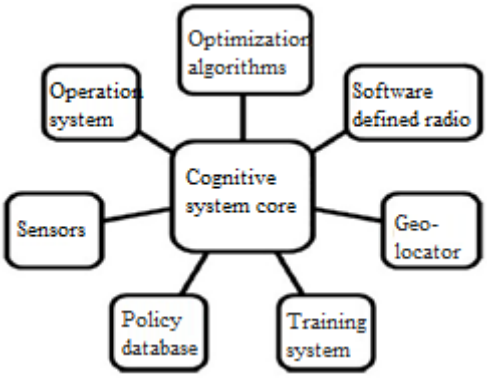


Fig. 2 5G components for cognitive terminals

- Cognitive system (core) - the cognitive core plays an essential role in coordinating and managing the internal parts of the device. This module supports different prediction algorithms and radio schemes for resource allocation.
- Software-defined radio is the main component of the cognitive terminal. This device is software-reconfigurable and designed with programmable components such as digital signal processors (DSPs), field programmable gate arrays (FPGAs), analogue-to-digital and digital-to-analog converters (ADC / DACs), reconfigurable amplifiers, smart antennas and broadband radio frequency (RF) circuits. Software-defined radio is defined as an object that provides software management of various modulation techniques, broadband or narrowband operability, communications security features, and wave requirements to current and emerging broadband standards.

Fig. 3 presents a software-defined radio architecture. The main software-defined radio features are presented as follows:

- Unobstructed ubiquitous communication - As selecting the appropriate wireless network for user's location and requirements.
- Reconfiguration - changing all radio parameters based on internal and external policies.

- Interoperability - capable to explore and communicate with different mobile networks.
- Getting closer to the desired service quality - improving service economy, according to the data rate and cost.

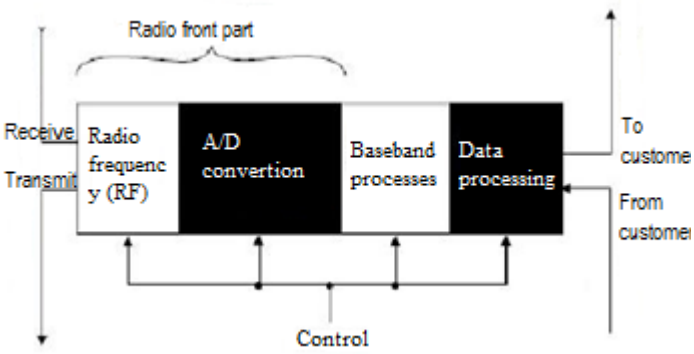


Fig.3 Software-defined radio architecture

IV. Conclusion

The mobile cellular networks are immutable part of modern people's life, and will be tomorrow. Data rates are considered as basic technical feature of current systems. Consumers' demands are increasing and system requirements are increasing as well. The new networks need to be highly reliable, low latent, highly efficient and productive, energy - efficient, to provide massive connectivity and very high data rates. With the introduction of the fifth generation and future technologies as mobile cloud, cognitive radio, virtualization, software-defined networks, our idea of mobile cellular systems will change very significantly.

Acknowledgement

This work was supported by the contract No DN 07/19 15.12.2016 "Methods for Estimation and Optimizing Electromagnetic Emissions in Urban Areas", funded by the National fund for scientific research, Bulgaria, 2016–2018.

References

[1] SK Telecom, 5G architecture design and implementation, 2015
[2] Dohler M., 5G Ultra-High Capacity Network Design, IEEE ComSoc, 2012
[3] Frank H. P. Fitzek, Marcos D. Katz, Mobile Clouds: Exploiting Distributed Resources in Wireless, Mobile and Social Networks, John Wiley & Sons, Inc., 2014
[4] Shi, L., Sung, K.W., and Zander, J., Controlling Aggregate Interference Under Adjacent Channel Interference Constraint in TV White Space, Cognitive Radio Oriented Wireless Networks and Communications, CROWNCOM, 2012
[5] Petrov G., A. Stancheva, V. Kadrev, Specialized Databases for Spectrum Management Research, Analysis and Forecasting of the Effects of Working Electromagnetics Fields, годишник на Департамент "Телекомуникации", НБУ – София, 2014
[6] Пасарелски Р., Универсални мобилни телекомуникационни системи, НБУ, 2013
[7] Petrov G.K., Balabanov B.H., Dimensioning and evaluation of the radio frequency spectrum, Monthly scientific and technical journal Elektrotechnica & Elektronica, The Union of Electronics, Electrical Engineering and Telecommunications /CEEC/, BULGARIA 2016.

ANN based Design of Planar Filters Using Square Open Loop DGS Resonators

Marin Nedelchev[1], Zlatica Marinkovic[2], Alexander Kolev[1]

Abstract – **This paper presents a novel design method for planar defected ground structure (DGS) square open loop resonator filters. The increased complexity of the coupling mechanism between the resonators and the impossibility to analytically calculate the coupling coefficients created the need of accurate modelling of the coupled resonators. This is even more complex when it is necessary to calculate the filter dimension for the given coupling coefficient. A novel method based on artificial neural networks (ANNs) is proposed in this paper. An ANN is used to develop the filter inverse model aimed to calculate the spacing between the resonators for predetermined coupling coefficients. There is a very good agreement between the simulations of the filter designed by the proposed method and the measurements.**

Keywords – **defected ground structure, planar filter, coupling coeffieicent, artificial neural network, inverse model**

I. INTRODUCTION

Microstrip filters are key components in microwave system design and their synthesis is a matter of persistent development and research. They must meet the stringent requirements for low passband loss and high steepness of the response in the stopband. Generally, planar filters are attractive to the designers because of their ease of manufacturing, adjustment and variety of topologies that offer realization of highly customized frequency responses. One of the most adopted microstrip resonators are the halfwave resonators and their derivatives - hairpin resonator, square open loop resonator, miniaturized hairpin resonator [1-3]. Increasing the order of the filter in order to achieve better suppression in stopband leads to increase of the sizes of the entire filter. Consequently, the main purpose is to reduce the size of the filter in order to implement it in modern compact systems in the low microwave band. The benefit of the square open loop filter is the compact topology, but it suffers from wide bandwidths, that require small gaps between the resonators. The synthesis of microstrip filters can be improved by intentionally implementing slots in the ground plane of the microstrip line. These slots are known as defected ground structures (DGS) and can be used as resonators combined to the microstrip line. The advantage is that no manufacturing constraints exist as the DGS and the microstrip line can overlap. The DGS resonators are investigated in [4] as the coupling coefficient is investigated and curve fitting formulas are derived. Also, it is possible to derive the formulas for the inverse relationship, i.e. for calculationg the filter dimensions for the given value of the coupling coefficients. However, the accuracy of these formulas is not satisfactory and additional tuning in a simulator is necessary. Having in mind good fitting abilities of the artificial neural networks (ANNs), which has qualified them as a good modeling tool in the field of RF and microwaves [5-15], this paper presents an alternative approach for design of planar filters using coupling coefficients derivation based on the ANNs. An example filter with DGS square open loop resonators is synthesized using the ANN for calculation of the coupling coefficients and the external coupling factor. The filter response is simulated and the filter is manufactured. The measured and simulated results coincide in order to prove the validity of the proposed approach.

The structure of the paper is as follows. After this introductory section, in Section II the considered model of DGS resonator and coupling structure is given. The ANN based design approach is described in Section III. Section IV contains the numerical results and discussion and the final conclusions are given in Section V.

II. MODEL OF DGS RESONATOR AND COUPLING STRUCTURE

In this paper, all the simulations and design procedures are performed for dielectric substrate FR-4 with height 1.5mm, relative dielectric constant $\varepsilon_r = 4.4$ and loss tangent $tg\delta = 0.02$ and centre frequency $f_0 = 2.4\,\text{GHz}$. The square open loop resonator considered in this paper is etched in the ground plane of the microstrip line and appears to be dual to the standard microstrip square open loop resonator described in [4]. It is shown in Fig.1, where a denotes the side of the square, w is the width of the slot and g is the gap.

The resonator consists of a slot line nearly half wavelength long. The etched resonator is symmetrical around the axis and the open end is in the middle of the main line. The magnetic field is stronger at the both ends of the line and the electric filed is at its maximum near the middle of the resonator.

[1]Marin Nedelchev and Alexander Kolevare with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: mnedelchev@tu-sofia.bg.

[2]Zlatica Marinkovic is with UniversityofNiš, FacultyofElectronicEngineering, AleksandraMedvedeva 14, 18000 Niš, Serbia, E-mail: zlatica.marinkovic@elfak.ni.ac.rs

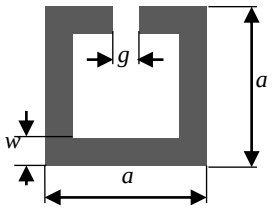


Fig.1. DGS resonator etched in the ground plane

For the further simulations and design procedures, the width of line is equal to the 50Ω microstrip line for FR-4 substrate. For simulations the Ansys Electronics Desktop 2016.2 planar 3D simulator is used. The resonance frequency can be found using the topology shown in Fig. 1 with a feeding line on the top side of the substrate. Once, the resonance frequency is found by the simulation, the filter design process continues with realization of the coupling coefficients with proper coupling topologies.

The most common coupling topology used is shown in Fig. 2. It consists of two closely positioned resonators with their sides.

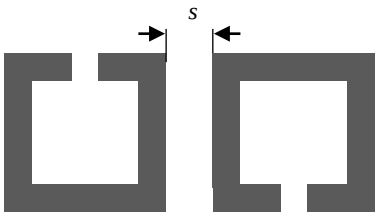


Fig.2. Coupling topology of DGS resonators etched in the ground plane

The nature of the coupling is mixed as neither the electric, nor the magnetic field is dominating over. The sign of the coupling coefficient is positive and this topology can be used in cascade topologies of microstrip filters. The resonance frequency and the coupling coefficients are extracted from the performed simulations in Ansys Electronics Desktop and following the methods described in [1]. The obtained values are used for training and test of the ANN.

III. ANN APPLICATION IN MICROWAVE FILTERS DESIGN AND INVERSE MODELLING

As mentioned in the introductory section a curve fitting technique can be used to extract the relationship between the coupling coefficient and the filter dimensions, i.e. in the particular case the space between the resonators. However, using the extracted dependence results in the dimensions that require additional tuning in order to achieve the desired filter characteristics. In this paper ANNs are proposed to be used to model the dependence of the space between the resonators and the coupling coefficient. For this purpose a multilayered ANN with one hidden layer is proposed, Fig. 3. The ANN has three layers of neurons: input layer (IL), hidden layer (HL) and output layer (OL). In the input layer there is one neuron, having a buffer role, with unitary transfer function. This neuron corresponds to the input parameter, in this case the coupling coefficient. The hidden layer consists of several neurons having a sigmoid transfer function. The number of hidden neurons in not a priori known. In the output layer there is one neuron with the linear transfer functions. In the considered model there is only one neuron and it corresponds to the space between the resonators. Each connection between a neuron and the neurons from the next layer is weighted. The input of the neuron transfer functions is a sum of weighted neuron inputs with added a bias. The ANN learn the dependence between input-output data samples by adjusting the ANN parameters, which are the connection weights and the transfer function biases. This procedure is known as the ANN training and there are several different training algorithms, among them the Levenberg-Marquardt algorithm [5] which is used in the work presented in this paper. The input-output pairs used for the ANN training are obtained in a full-wave simulator.

The trained ANN gives accurate response not only for the input parameters used for the ANN training but also for any other input value from the considered range of values. That means that the space between the resonators for the given coupling coefficient is determined by finding the network response without a need for additional full-wave simulations. It is important to note that the mathematical expressions which describe ANNs are easy to be further implemented in the working environment and therefore easy to be further used.

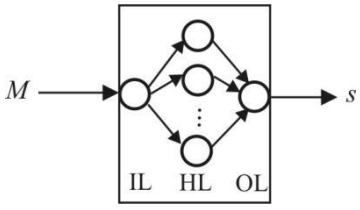


Fig.2. ANN inverse model of the filter coupling

IV. NUMERICAL RESULTS

In order to prove the proposed approach, a third order filter is synthesized. The filter specifications are:

- Centre frequency: $f_0 = 2400$ MHz
- Bandwidth: $\Delta f_0 = 270$ MHz
- Return Loss: $RL = -15$ dB

The design process of DGS square open loop resonator filter is carried out using the method described in [1,4]. It starts with calculation of the coupling matrix $[k]$ for low pass canonical filter topology for Chebyshev approximation. Then

all the coupling coefficients are renormalized to the fractional bandwidth (FBW) and the external coupling factors are calculated as:

$$M_{12} = M_{23} = k_{12}.FBW = k_{23}.FBW \ ,$$

$$Q_e = \frac{k_{S1}}{FBW} = \frac{k_{3L}}{FBW} \qquad (2),$$

where k_{ij} are the coupling coefficients from the approximation and Q_e is the external quality factor.

The calculated values of the coupling coefficients from the Chebyshev approximation are $M_{12} = M_{23} = 0.099$ and the external quality factor is $Q_e = 8.4027$. For the realization of the computed coupling coefficients the topology of mixed coupling was used.

The physical dimensions of the resonator tuned to the center frequency are found to be $a = 14.5\,\text{mm}$, $g = 1\,\text{mm}$, $w = 2.71\,\text{mm}$.

Following the simulations of the coupling topology, the coupling coefficient was extracted. Using curve fitting technique the dependence of the coupling coefficient to the space between the resonators is:

$$s_{mix} = 20.63e^{-25.11M_{mix}} \qquad (1)$$

As can be seen from Fig. 3 this dependence (dashed line) does not fit well the reference values. Therefore, as the next step the ANN inverse filter model was developed. Namely, as the number of hidden neurons is not a priori known, several ANNs with one input neuron, one output neuron and a different number of hidden neurons were trained. Comparing the accuracy of the trained ANNs, the ANN having one hidden layer with five neurons was chosen as the final ANN model. The spacing between resonators obtained by the chosen model is plotted in Fig. 4 with the step of 0.001. It is obvious that much better fitting was achieved. It should be noted that the range of the validity of this model, regarding to the input range, is determined by the range of the values of the training input data.

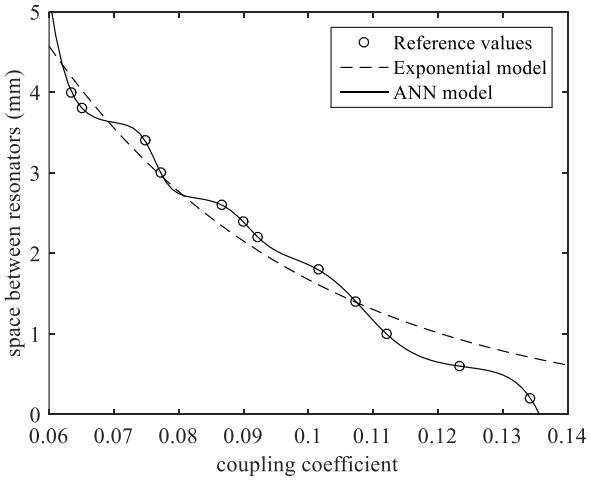
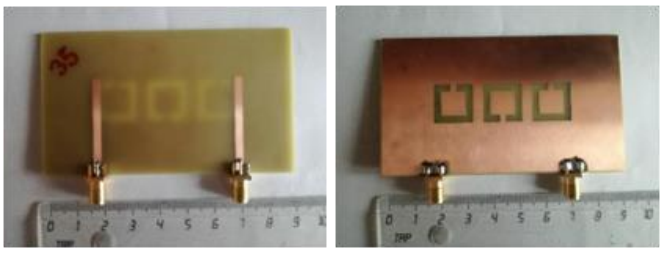


Fig. 3. Spacing between the resonators vs. coupling coefficient

Further, for the calculated coupling coefficient of 0.099 (which is not used neither for exponential fitting nor for the ANN training), the spacing between the resonators was calculated.

The computed distance between the resonators with the ANN is $s_{ANN} = 1.8948$ mm and from the curve fitting is $s_{curvefitting} = 1.7834$ mm.

The designed filter was simulated in Ansys Electronics Desktop with the dimensions computed using the ANN and the all the distances were kept as they are calculated. No further optimizations were performed in order to correctly prove the accuracy and the applicability of the proposed approach for filter design. The synthesized filter was fabricated and the layout (top and bottom side) of the synthesized is shown in Fig. 4.



(a) (b)

Fig. 4. Manufactured and measured slot resonator filter
(a) top layer, (b) bottom layer

The measured and simulated results are presented on a common plot on Fig. 5. As it is seen, there is a very good agreement between the simulated and measured results.

TABLE 1 SIMULATION AND MEASUREMENT PARAMETER COMPARISON

	f_0 [MHz]	f_{low} [MHz]	f_{high} [MHz]	BW [MHz]
design	2400	2265	2535	270
simulation	2438	2272	2552	280
measurement	2402	2228	2525	297

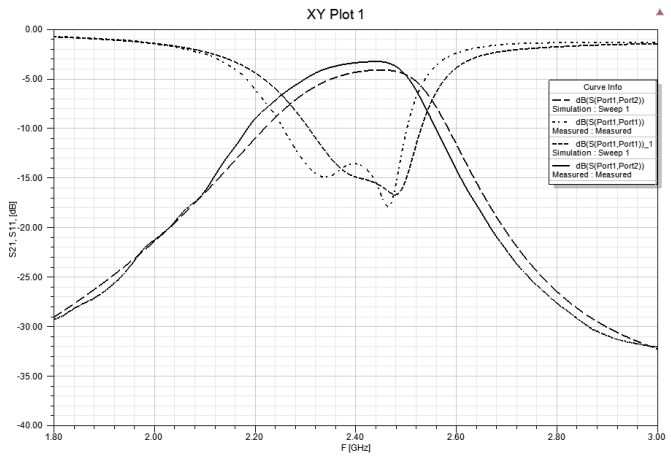


Fig. 5. Measured and simulated response of the third order DGS square open loop resonator filter

Table I summarizes the main parameters of the design requirements, the simulation and measured results. f_{low} and f_{high} denote the low and high cut-off frequency in the filter response.

The minimum measured return loss in the passband is -13.6 dB and the simulated value is -14dB. It is seen from Fig.4 that there is very good agreement between the simulated and measured results. Therefore the proposed method for design of planar DGS resonator filters can be used in the engineering practice.

V. CONCLUSION

A design method for planar DGS square open loop resonator filters is presented in this paper. The method is based on developing the ANN aimed to calculate the spacing between the resonators for predetermined coupling coefficients of the filter. The numerical results showed, that this method enables calculating of the spacing between the resonators which will results in the filter characteristics according to the design requirements, which was not the case when simple curve fitting exponential formulas are used, when it was necessary to perform additional tuning of the spacing value. The filter with the dimensions calculated by the proposed approach was fabricated and the filer was measured. The simulation and measurement results show very good agreement and prove the applicability of the proposed method for the filter dimensions calculation.

ACKNOWLEDGEMENT

The research leading to the presented result are financed and supported by a contract between The Ministry of Education, Republic Bulgaria and Faculty of Telecommunications under number ДН07/19/15.12.2016 „Methods of estimation and optimization of the electromagnetic radiation in urban areas". The work was also supported by the project TR-32052 of the Serbian Ministry of Education, Science and Technological Development.

REFERENCES

[1] J. Hong, M. J. Lancaster, *Microstrip Filters for RF/Microwave Applications*, John Wiley and Sons Inc, 2001.

[2] 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, no. 3, 2004.

[3] S. K. Parui, S. Das, "A New Defected Ground Structure for Different Microstrip Circuit Applications", Radioengineering vol. 16, no 1, 2007.

[4] K. Vagner, "A Novel Bandpass Filter Using a Combination of Open-Loop Defected Ground Structure and Half-Wavelength Microstrip Resonators", Radioengineering, vol. 19, no. 3, pp. 392-396, 2010.

[5] Q. J. Zhang, K. C. Gupta, *Neural Networks for RF and Microwave Design*, Artech House, 2000.

[6] C. Christodoulou, M. Gerogiopoulos, *Applications of Neural Networks in Electromagnetics*, Artech House, Inc. Norwood, MA, USA, 2000.

[7] J. E. Rayas-Sanchez, "EM-based optimization of microwave circuits using artificial neural networks: The state-of-the-art," IEEE Trans. Microw.Theory Tech., vol. 52, no. 1, pp. 420–435, 2004.

[8] H. Kabir, L. Zhang, M. Yu, P. Aaen, J. Wood, and Q. J. Zhang "Smart modelling of microwave devices", IEEE Microw. Mag., vol. 11, pp.105–108, 2010.

[9] J. Michalski "Artificial Neural Networks Approach In Microwave Filter Tuning", Progress In Electromagnetics Research M, vol. 13, pp.173-188, 2010.

[10] Z. Marinković, V. Marković, A. Caddemi, "Artificial Neural Networks in Small-Signal and Noise Modelling of Microwave Transistors," Chapter 6 in *Artificial Neural Networks*. Edited by Seoyun J. Kwon, Nova Science Publishers Inc., pp. 219-236, 2011.

[11] M. Agatonović, Z. Stanković, N. Doncov, L. Sit, B. Milovanović, T. Zwick, "Application of Artificial Neural Networks for Efficient High-Resolution 2D DOA Estimation", Radioengineering, vol. 21, no. 4, pp. 1178-1185, 2012.

[12] M. Agatonović, Z. Marinković, V. Marković, "Application of ANNs in Evaluation of Microwave Pyramidal Absorber Performance," Applied Computational Electromagnetics Society Journal, vol. 27, no. 4, 2012, pp. 326-333

[13] Z. Marinković, G. Crupi, A. Caddemi, G. Avolio, A, Raffo, V. Marković, G. Vannini, Dominique M. M.-P. Schreurs, "Neural approach for temperature-dependent modeling of GaN HEMTs," International Journal of Numerical Modeling: Electronic Networks, Devices and Fields, vol. 28, no 4, pp. 359-370, 2015

[14] Z. Marinković, T. Kim, V. Marković, M. Milijić, O. Pronić-Rančić, L. Vietzorreck, "Artificial Neural network based design of RF MEMS capacitive shunt switches", Applied Computational Electromagnetics Society Journal, vol. 31 no. 7, pp. 756-764, 2016.

V. Đorđević, Z. Marinković, V. Marković, O. Pronić-Rančić, "Extraction of microwave FET noise wave temperatures by using a novel neural approach", The International Journal for Computation and Mathematics in Electrical and Electronic Engineering – COMPEL, vol.35, no.1, pp. 339-349, 2016.

Parametric Analysis of Wideband CPW-fed Bow-Tie Slot Dipole

Marija Milijić[1] and Branka Jokanović[2]

Abstract – **A uniplanar bow-tie slot dipole fed by CPW (coplanar waveguide) is proposed in this paper. Single element is designed for application in 5G frequency range 24.25-27.5 GHz. It is described how dimension parameters of bow-tie slot dipole can be designed to tune the impedance of the antenna over a wide range. The simulated results demonstrate that this structure provides the advantage of wide impedance bandwidth besides small size, low profile, easy and cheap fabrication. This antenna structure is particularly suitable for series-fed array configurations and broad-band design.**

Keywords – **Coplanar waveguide, bow-tie slot antenna, broad-band applications.**

I. INTRODUCTION

The interest for the coplanar waveguide CPW (coplanar waveguide) - fed antennas has increased significantly in recent years [1-4]. Accordingly, many antenna elements suitable for a CPW-fed configuration have been proposed to meet the necessary requirements of future 5G technologies defined as great capacities broadband service and transmission speeds. CPW-fed antennas are preferable for 5G mobile communication applications considering their low cost, light weight, small size, and ease of fabrication.

In order to set a strong foundation for the rapid advancement of the next-generation 5G networks, the International Telecommunication Union (ITU) announced the following spectrum for 5G, which includes the 24.25 – 27.5 GHz, 37 – 40.5 GHz, 66 – 76 GHz bands. Also, in the United States, the Federal Communications Commission (FCC) has announced the spectrum of approximately 11 GHz above 24 GHz for flexible, mobile and fixed wireless broadband, comprising 7 GHz of unlicensed spectrum from 64 to 71 GHz and 3.85 GHz of licensed spectrum in three bands: 27.5 to 28.35 GHz, 37 to 38.6 GHz and 38.6 to 40 GHz [5].

In the interest of meeting the requirements of wide-band applications in 5G frequency range 24.25-27.5 GHz [6,7], where high-gain and unidirectional radiation are crucial specification, this paper has investigated the CPW-fed dipoles with slots of bow-tie shape. The parameters of bow-tie slot are investigated to achieve both the desired operating frequency and impedance.

[1] Marija Milijić is with the Faculty of Electronic Engineering, University of Niš, 18000 Nis, Serbia, E-mail: marija.milijic@elfak.ni.ac.rs

[2] Branka Jokanović is with the Institute of Physics, University of Belgrade, Pregrevica 118, 11080 Pregrevica, Serbia E-mail:brankaj@ipb.ac.rs

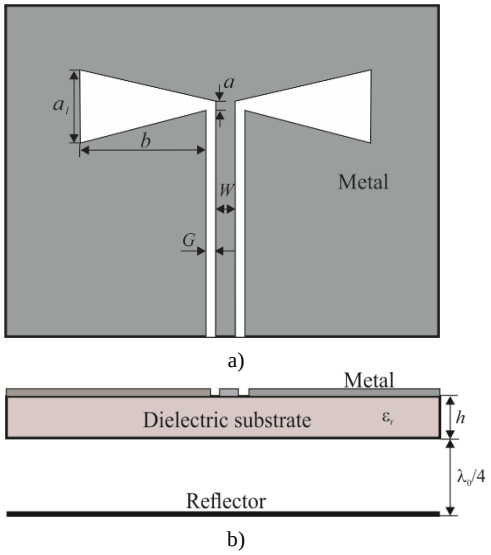


Fig. 1. CPW-fed bow-tie slot dipole: a) Top view b) Side view

II. CPW-FED BOW-TIE SLOT DIPOLE

Proposed CPW-fed bow-tie slot dipole is depicted in Fig. 1. The radiating element and CPW feeding line are positioned as a uniplanar structure on the same side of the substrate with thickness of h=0.508 mm and relative permittivity of ε_r=2.54. The overall size of the substrate is λ_g x λ_g, where λ_g=9.72 mm is CPW line wavelength at the frequency f=25.875 GHz. The bow-tie dipole has length b along which its width increases from a to a_1. The widths of the strip and gap (W and G) of the CPW feeding line, whose impedance is around 100 Ω, are 0.5 mm and 0.3 mm, respectively [8]. The dipole is with planar reflector plate [4] at the distance $\lambda_0/4$, where λ_0=11.6 mm is wavelength in air for the centre frequency of the band f=25.875 GHz.

The antenna is designed in WIPL-D software [9]. The parameters of bow-tie slot dipole are investigated to achieve a wide band coverage of the antenna.

III. SIMULATION RESULTS

First analysis was conducted to determine the advantages of bow-tie slot dipole over standard rectangular slot dipole [3]. Therefore, the initial antenna design is rectangular slot dipole with equal parameters a and a_1 (a=a_1=0.3 mm). The further investigation examines impedance modification for bow-tie slot whose parameter a is constant (a=0.3 mm) when parameter a_1 rises from 0.5 mm to 1.5 mm with step 0.2 mm. Simulated results are presented in Fig. 2-5 and Table I.

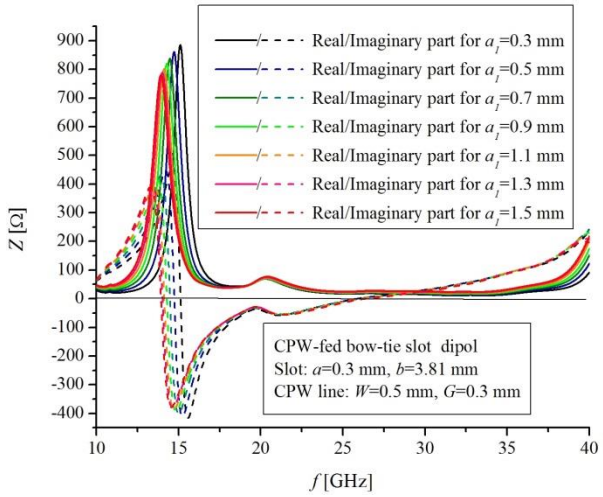


Fig. 2. Dipole's impedance versus frequency for different values of slot parameter, a_1.

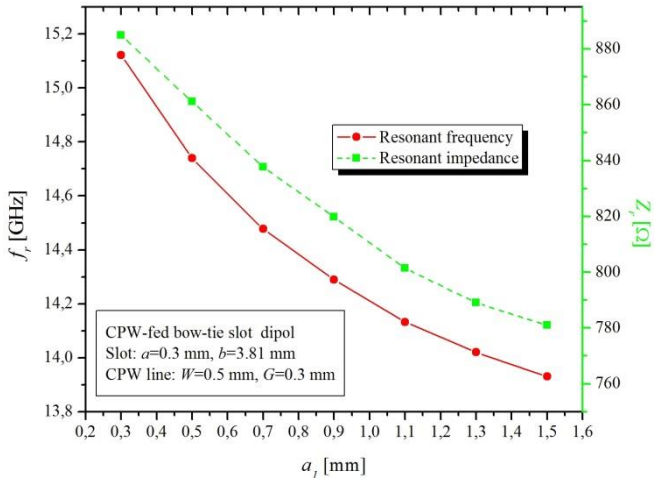


Fig. 3. Resonant frequency and impedance of CPW-fed bow-tie slot dipole for the different values of slot parameter, a_1.

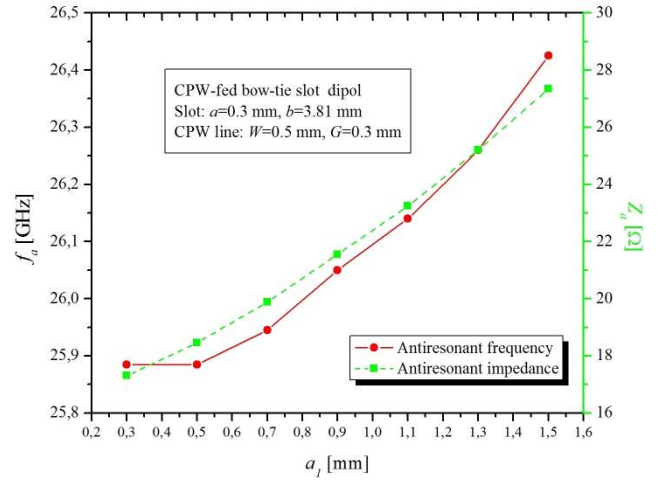


Fig. 4. Antiresonant frequency and impedance of CPW-fed bow-tie slot dipole for the different values of parameter, a_1.

Fig. 2 shows dipole's impedance versus frequency for different values of parameter a_1. It is obvious that proposed dipole has two resonances: the first resonance with high impedance and the second one or antiresonance with low impedance value.

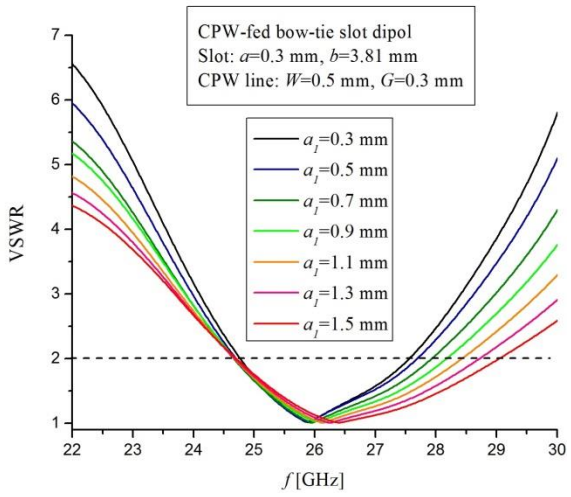


Fig. 5. Dipole's VSWR versus frequency for the different values of parameter, a_1.

TABLE I
BANDWIDTH OF CPW-FED BOW-TIE SLOT DIPOLE FOR THE DIFFERENT VALUES OF PARAMETER a_1

Slot parameter a_1 [mm]	Bandwidth [GHz] In respect to VSWR < 2	Relative bandwidth in respect to the center frequency [%]
0.3	24.775-27.55	10.6
0.5	24.715-27.685	11.3
0.7	24.67-27.925	12.4
0.9	24.7-28.165	13.1
1.1	24.685-28.42	14.1
1.3	24.7-28.705	15.0
1.5	24.715-29.035	16.1

It can be seen that slot impedance at the resonance varies 12.05% when parameter a_1 changes from 0.3-1.5 mm, while at the second resonance the impedance has greater difference from 17 Ω to 27 Ω which makes a relative change of 45%. Considering resonance frequency in Fig. 3, it decreases for 8.2% when slot width a_1 rises from 0.3 mm to 1.5 mm, while at the same time the antiresonant frequency slightly increases from 25.88 GHz to 26.4 GHz (2%) as can be seen in Fig. 4.

Since the slot impedance is very large at the resonance, it is not suitable for series-fed array of slot dipoles to work around the resonance. However, the antiresonance frequency features smaller impedance that can be observed in Fig. 4. Therefore, it can be concluded that antiresonance frequency is more suitable working band for serial connected CPW-fed bow-tie slot dipoles. Moreover, the smaller value of parameter a_1 causes smaller antiresonance impedance.

On the other hand, simulated results show that as the slot width a_1 increases, the bandwidth of the slot grows. Fig. 5 presents VSWR parameter of bow-tie slot dipole for different slot width a_1 matching the antiresonant impedance Z_a. The numerical values from Fig. 5 are presented in Table I together with the bandwidth relative to central frequency.

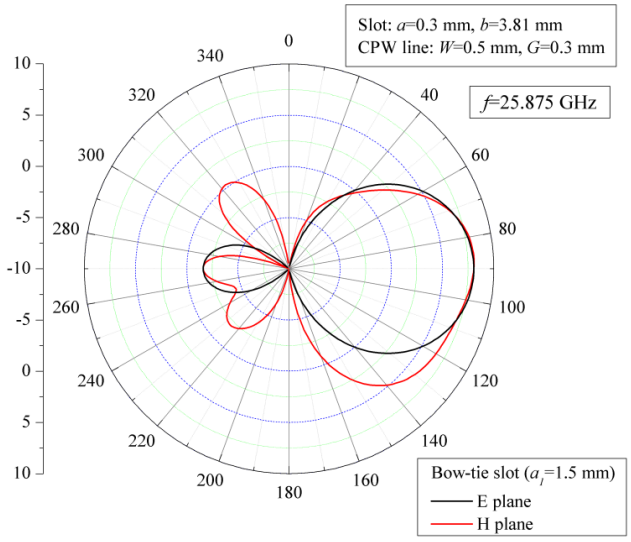


Fig. 6. Radiation pattern of the bow-tie slot dipole in E and H-plane.

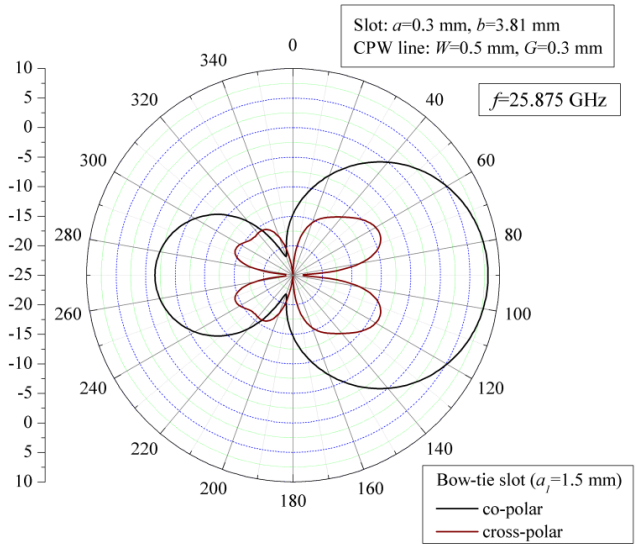


Fig. 7. Co-polar and cross-polar radiation pattern of the bow-tie slot dipole.

Analyzing previous presented results, it can be concluded that greater parameter a_1 leads to bigger bandwidth. Therefore, a_1 is adjusted to biggest examined value (1.5 mm) for further analysis. Fig. 6 presents simulated E-plane and H-plane radiation patterns and Fig. 7 presents simulated co- and cross-polarization of proposed bow-tie slot dipole.

Fig. 8-11 and Table II show the influence of the slot length *b* on resonance and antiresonance frequency bands and impedance of the proposed dipole. Since, the parameter *b* made a primary impact on resonance frequency band when length *b* changes from 3.51 mm to 4.11 mm with step 0.1 mm (Fig. 8). For this change of slot length *b*, resonant frequency has drop for 7.9 % (from 14.5 GHz to 13.4 GHz) while its impedance falls from 925 Ω to 650 (35 %) (Fig. 9). Still, the same change of the slot length *b* causes that the antiresonance frequency f_a decreases for 12.3 % (from 28.5 GHz to 25.2 GHz) while the slot impedance falls from 28 Ω to 23 Ω

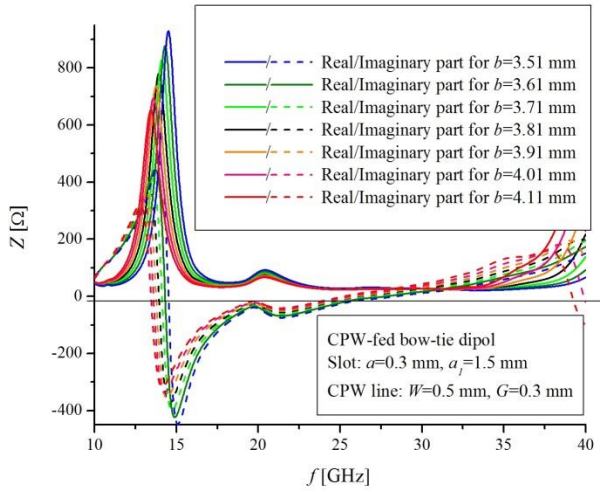


Fig. 8. Dipole's impedance versus frequency for different values of slot parameter, *b*.

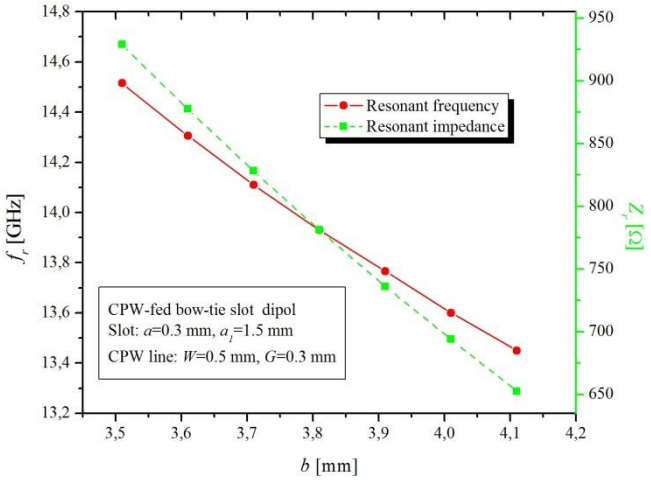


Fig. 9. Resonant frequency and impedance of CPW-fed bow-tie slot dipole for the different values of slot parameter, *b*.

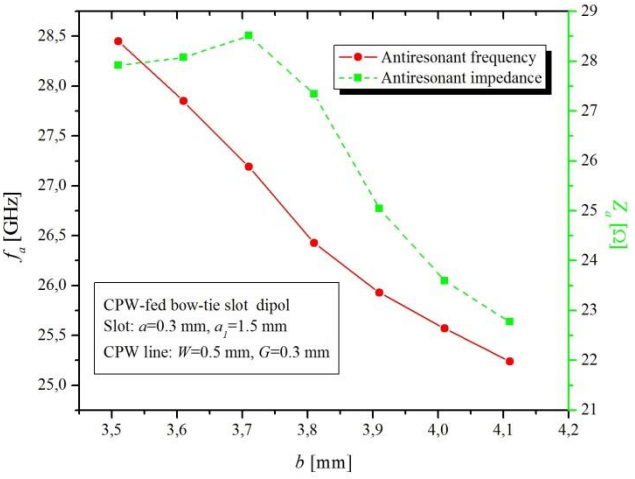


Fig. 10. Antiresonant frequency and impedance of CPW-fed bow-tie slot dipole for the different values of slot parameter, *b*.

(19.6 %) (Fig. 10). Furthermore, the slot impedance has calm development for both, the real and imaginary part, at the dipole second resonance (antiresonant frequency) (Fig. 8).

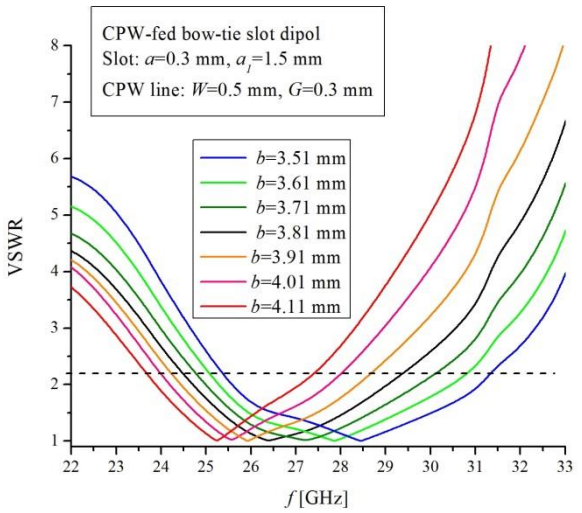


Fig. 11. Dipole's VSWR versus frequency for the different values of slot parameter, *b*.

TABLE II
BANDWIDTH OF CPW-FED BOW-TIE SLOT DIPOLE FOR THE DIFFERENT VALUES OF SLOT PARAMETER *b*

Slot parameter *b* [mm]	Bandwidth [GHz] In respect to VSWR < 2	Relative bandwidth in respect to the center frequency [%]
3.51	25.585-31.09	19.4
3.61	25.3-30.46	18.5
3.71	25-29.725	17.3
3.81	24.715-29.035	16.1
3.91	24.445-28.375	14.9
4.01	24.205-27.73	13.6
4.11	23.89-27.085	12.5

Fig. 11 shows VSWR parameter in wide frequency range matching impedance Z_a at the antiresonant frequency. Table II gives the bandwidth for the different slot length *b*. With respect to presented results in Fig. 11 and table II, it is obvious that the operating bandwidth of a bow-tie slot decreases when the slot length *b* grows.

IV. CONCLUSION

A CPW-fed dipole with slot of bow-tie shape is proposed for working in the frequency band 24.25-27.5 GHz intended for future 5G cellular networks. The dipole design is compact, with dimensions of only 9.72 × 9.72 × 0.508 mm^3. It has a simple geometry and is relatively easy to fabricate because of its single-layer metallic structure. Besides its uniplanar structure that can satisfy the requirements of low profile, the presented simulated results show that proposed dipole features great bandwidth with good impedance matching. A wider bandwidth can be achieved by correctly chosen parameters of a bow-tie slot. It is shown that both parameters a_1 (bigger width of bow-tie shape) and parameter *b* (length of bow-tie slot) effects the bandwidth of the antenna. The demonstrated results indicate that wider bandwidth can be achieve if the antiresonance frequencies are considered, since there is a calm variation of the impedance, for both the real and imaginary part. Furthermore, slot dipole has significantly smaller impedance at antiresonance frequency band.

Presented results have great importance in antenna array design. In order to obtain desired antenna parameters like radiation pattern, gain, side lobe suppression, etc. a large number of dipoles connected in array are required. Considering serial connection of the slot dipoles and CPW feeding line, their impedance should be small enough (around 20 Ω) resulting in acceptable input impedance of the array. Therefore, the adjusting of slot dipole's impedance is a very important step in the beginning of antenna array design.

Further research will be directed to design of a millimeter-wave frequency scanning antenna array for radar sensors [10], which consist of a frequency modulated continuous wave radar in combination with a frequency scanning antenna.

ACKNOWLEDGEMENT

This work was supported by the Ministry of Education, Science and Technological Development of Republic Serbia under the projects No. TR 32052 and TR 32024.

REFERENCES

[1] G. H. Elzwawi, M. Mantash and T. A. Denidni, "Improving the gain and directivity of CPW antenna by using a novel AMC surface," 2017 IEEE International Symposium on Antennas and Propagation, San Diego, CA, 2017, pp. 2651-2652.
[2] A. Nešić, "Slotted Antenna Array Excited by a Coplanar Waveguide," Electron. Lett., vol. 18, no. 6, pp. 275-276, 1982.
[3] A. U. Bhobe, C. L. Holloway, M. Piket-May, R. Hall, "Wide-Band Slot Antennas With CPW Feed Lines: Hybrid and Log-Periodic Designs", IEEE Trans. Antennas Propag., vol. 52, no. 10, pp. 2545-2554, October, 2004.
[4] J.Joubert, J. C. Vardaxoglou, W. G. Whittow, J. W. Odendaal, "CPW-Fed Cavity-Backed Slot Radiator Loaded With an AMC Reflector", IEEE Trans. Antennas Propag., vol. 60, no. 2, pp. 735-742, February, 2012.
[5] M. J. Marcus, 5G and "IMT for 2020 and beyond," IEEE Wireless Commun., vol. 22, no. 4, pp. 2-3, Aug. 2015
[6] Vojislav Milosevic, Branka Jokanovic, Olga Boric-Lubecke, Victor M. Lubecke, "Key Microwave and Millimeter Wave Technologies for 5G Radio," in Powering the Internet of Things with 5G Networks, V. Mohanan, R. Budiatru, I. Aldmour, Eds. IGI Global, July 2017, DOI: 10.4018/978-1-5225-2799-2.
[7] W. Hong, K. Baek, S. Ko "Millimeter-Wave 5G Antennas for Smartphones: Overview and Experimental Demonstration", IEEE Trans. Antennas Propag., vol. 65, no. 12, pp. 6250 - 6261, December 2017.
[8] M. Milijić, A. Nešić and B. Milovanović, "Study of dielectric substrate effect on modelling CPW-fed slot antenna arrays," 2017 13th International Conference TELSIKS, Nis, 2017, pp. 105-108.
[9] WIPL-D Pro v10.0, WIPL-D Team
[10] J. Schafer, B. Gottel, H. Gulan, T. Zwick, " Integrated planar 122 GHz FMCW radar with frequency scanning antenna", Wireless Sensors and Sensor Networks (WiSNet), 2018 IEEE Topical Conference on, pp. 41-43.

Polarization improvement in pin-fed patch antenna associated with lateral displacement of feed point

Yordan Kechev[1], Peter Petkov[2] and Mario Gachev[3]

Abstract – **This paper presents a new technique for polarization improvement of pin-fed patch antennas. In order to obtain higher bandwidth and dual polarization, stacked patch with two orthogonal pin excitations is used. To achieve lower cost and compatibility with standard PCB manufacturing process RO4003 substrates are used. Simulations are made with full wave 3D FEM simulator HFSS.**

Keywords – **Polarization improvement, Dual polarized microstrip antenna, Pin fed patch antenna, Stacked patch antenna, Planar antenna array.**

I. Introduction

Dual polarized patch antennas are widely used in communications and radar systems. Their easy implementation using standard PCB manufacturing process and their low profile make them suitable for many applications. One of the most common issues in dual polarized patch antennas are cross polarization level and isolation between two ports. The insufficient isolation between ports cause degradation of antenna performance and deteriorate signal to noise ratio (SNR) of a system. Cross polarization level depends on mutual coupling between two ports and, in case of antenna array, on mutual coupling between radiating elements. The problem was investigated in case of aperture coupled dual port patches in [1,2] and good isolation between ports and high cross polarization ratio was reported. Another papers [3,4] present balanced pin-fed patch antennas. Low cross polarization levels are reported due to the proper balanced feeding. In this paper a new technique for cross polarization ratio reduction is proposed.

II. Antenna Geometry and Design Technique

Fig. 1 represents geometry of pin-fed dual polarized patch antenna for Ku band. To achieve wider bandwidth stacked patch antenna configuration is used. Simplifying the measurement process, a proper pin to microstrip transitions are designed and entire antenna with transitions is simulated using HFSS full wave 3D simulator. Antenna comprises the following main parts:

- Feed substrate – RO4003 0.020” (0.508 mm);
- Bottom patch substrate – RO4003 0.020’’ (0.508 mm);
- Parasitic patch substrate – RO4003 0.032’’ (0.813 mm);
- Feed microstrip lines width – 1.14 mm;
- Orthogonal pins diameter – 0.6 mm;
- Bottom patch diameter – 3.45 mm;
- Parasitic patch diameter – 3.2 mm;

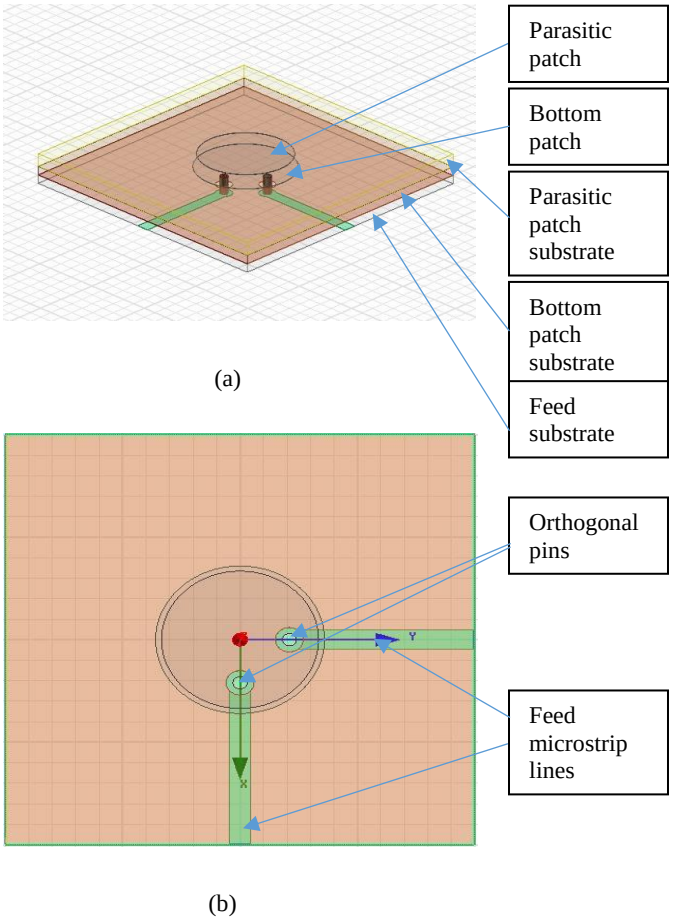


Fig. 1 Structure of dual-polarized patch antenna. (a) Three-dimensional view. (b) Top view.

To obtain two orthogonal polarizations, two orthogonal wave modes need to be excited. This is achieved using two orthogonal pins to feed the antenna. In the commonly used design feeding pins are oriented along the X Y axis (as it is shown on Fig 1 (b) and the polarization ratio of the antenna depends mainly on the mutual coupling (isolation) between two ports. The radiation patterns for co and cross polarization are

[1]Yordan Kechev is a student at Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: jgkechev@gmail.com.

[2]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.

[3]Mario Gachev is CTO at MatriQx Antenna System Ltd., 149 Tsarigradsko shose, Sofia 1784, Bulgaria, E-mail: Mario.gachev@matriqx.bg.

shown on Fig.3 (a) and Fig.4 (a). As it can be easily seen the cros polarization ratio is approximately close to between ports isolation. The proposed antenna configuration in shown on Fig.2. Using the proper feed points displacement a significant cross polarization suppression could be achieved.

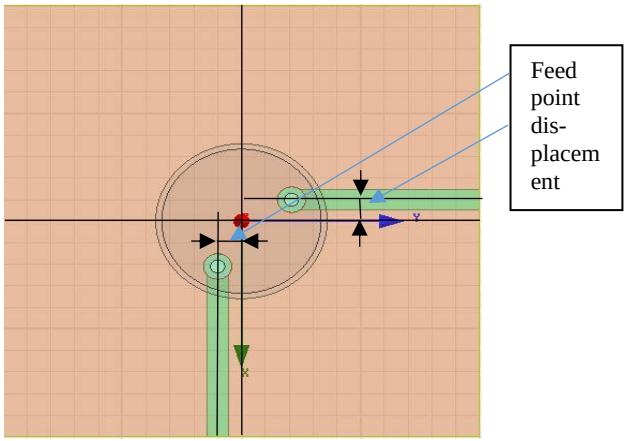
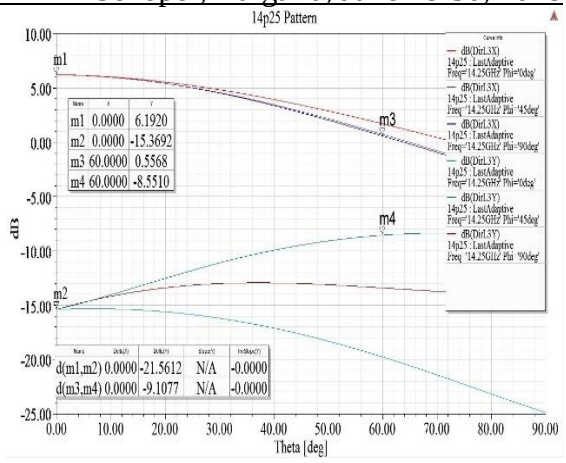


Fig. 2 Top view of patch antenna with lateral displacement.

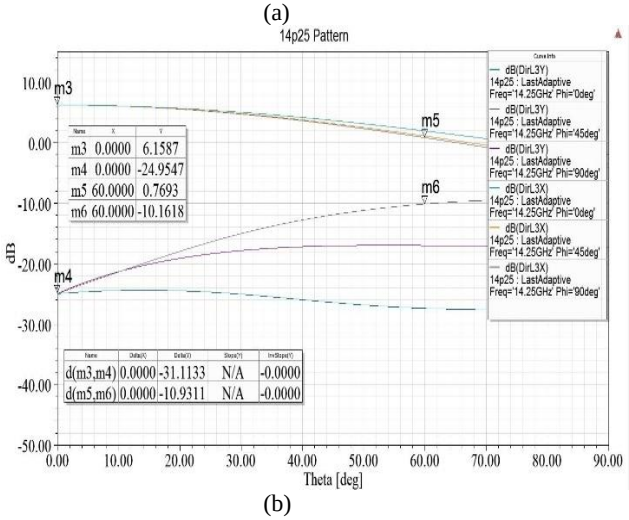
As it is shown both pins have displacement from their nominal position. Pin for "vertical" polarization (X axis) has offset negative to other (Y) axis and pin for "horizontal" polarization (Y axis) has offset negative to X axis or in other words when one of ports is displaced (negative to Y axis) other port is image of first along the same axis (Y) and then second port is rotated on 90 degrees. With this offset degradation of polarization ratio and isolation between ports is expected because of rotation of field vector from nominal coordinate system (0xy on Fig.2) and non-orthogonality between ports. As it is shown on Fig.4 isolation between ports decrease, so mutual coupling is increased and cross-polarization components are transferred. Cause of image copy and rotation of second ports from the first one, these cross – polarization components have 180 degrees phase difference and their sum is zero. Due to this phase difference polarization ratio of antenna is increased without degrading of radiation pattern symmetry and characteristics for the main polarization as it is shown on (Fig.3).

III. Simulation Results

Fig.3 shows radiation pattern for co- and cross – polarizations for nominal design and for the proposed technique. Patches have dimensions close to λ/2 and antenna resonate at its main mode H_{11}. Patterns are symmetrical in different azimuthal cuts and have maximum directivity at boresight about 6 dBi. Fig.3b shows 10 dB increasing of polarization ratio on boresight at 14.25 GHz compared to standard design (Fig.3a). At the same time no significant degradation in return loss between two designs is observed as it can be seen from (Fig.4). Isolation between ports is decreased from 20 dB(Fig.4a) to 17.5 dB(Fig.4b) but the cross-polarization ratio is significantly improved due to the feeding points displacement as it was explained above.

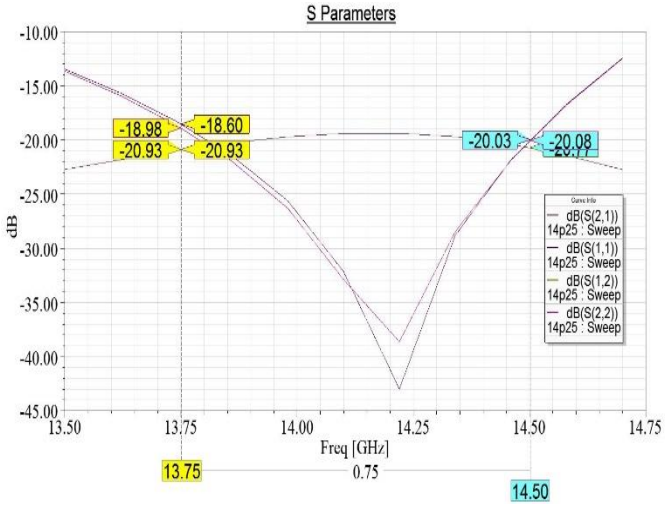


(a)

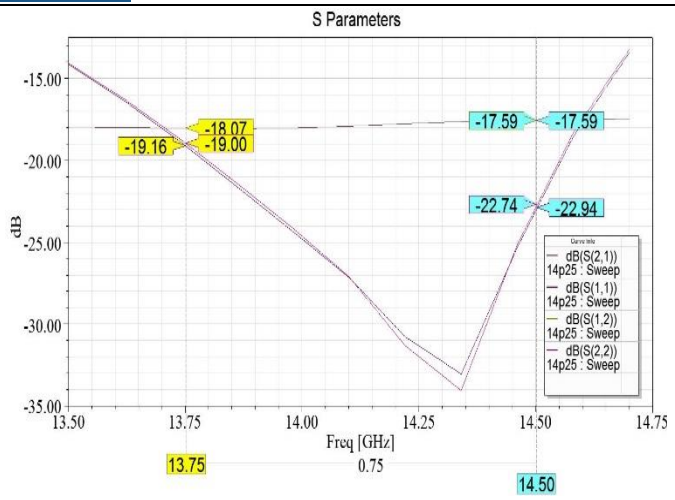


(b)

Fig.3 Radiation pattern for both polarizations. (a) Without lateral displacement. (b) With lateral displacement.

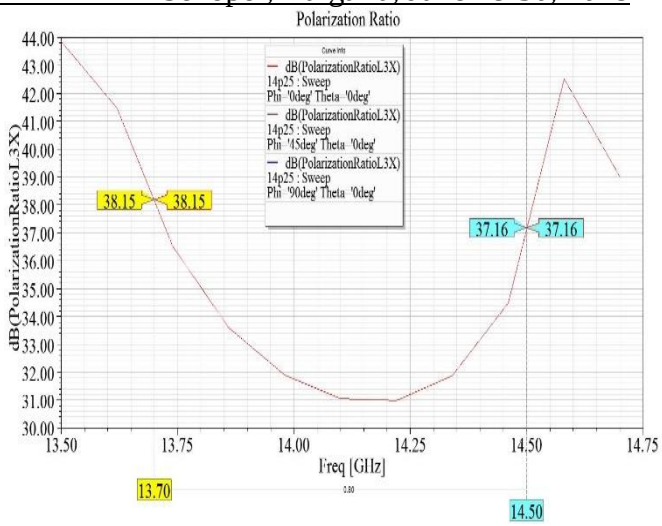


(a)



(b)

Fig.4 Scattering Parameters of the antenna. (a) Without lateral displacement. (b) With lateral displacement.



(b)

Fig.5 Polarization ratio of nominal design and proposed technique. (a) Without displacement. (b) With displacement.

The cross polarization ratio improvement using the proposed techique can be clearly seen from the graphs shown on Fig.5, Fig.6 and Fig.7. They represents two cases:

- (a) standard design;
- (b) design using the proposed technique;

Used offsets are optimized to be 200 um for both ports in order to achieve symmetry. The results from simulations show 10-15 dB higher cross-polarization suppersion in the proposed design compared with the standard one. Improved cross-polarization performance could contribute significantly for better antenna performance, reducing interferience and iproving of the communication link signal-to-noise ratio .

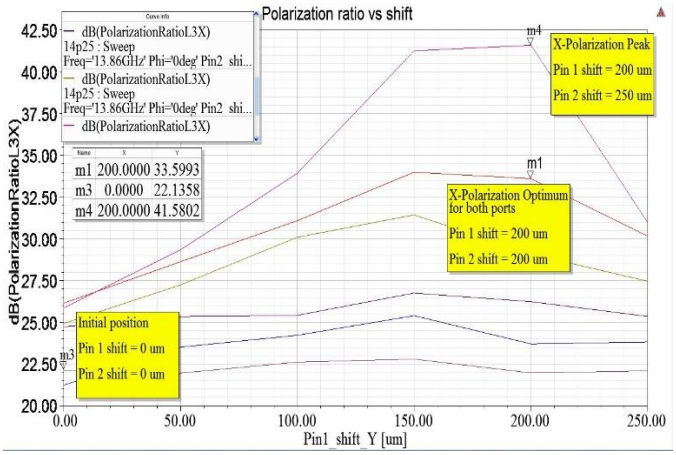
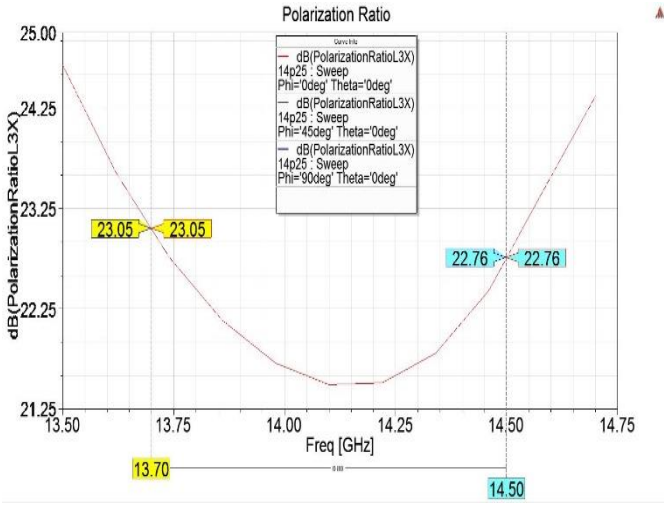
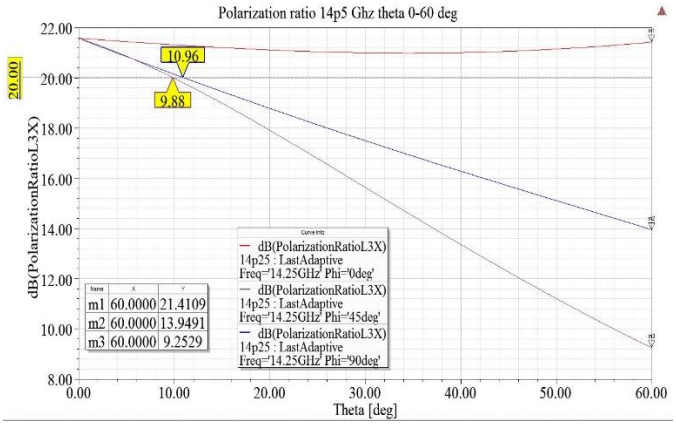


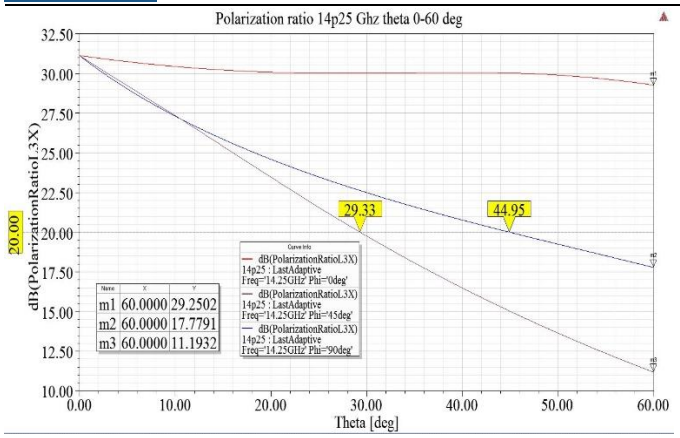
Fig.6 Polarization ratio with different ports shift.



(a)



(a)



(b)

Fig.7 Polarization ratio for 0 to 60 degrees elevation (Theta) and in three azimuthal cuts (Phi). (a) nominal design. (b) With optimum offset.

IV. CONCLUSION

A new technique for polarization ratio improvement of two port patch antenna is presented. It allows to improve significantly cross polarization ratio using proper displacement of antenna feeding points. Simulation shows 10-15 dB improvement of cross polarization isolation at boresight compared to the between port isolation. Some improvement can also be achieved in case of using such antenna in scanning arrays for different antenna beam tilts. From another side no changes in S- parameters and Directivity of the antenna were observed. The prosed design of dual port patch antenna can be successfully implemented in antenna arrays for communication and radar applications.

REFERENCES

[1] C. H. Tsao, Y. M. Hwang, F. Kilberg and F.Dietrich, " Aperture-coupled Patch Antennas with Wide-bandwidth and Dual polarization Capabilities", Antennas and Propagation Society International Simposium,1988. AP-S. Digest.

[2] S. LI, X. Cao, J. Gao , P. Gao, " High_Isolation Dual Polarized Microstrip Antenna via Substrate Integrated Waveguide Technology", RADIOENGINEERING, vol. 23 , NO. 4, DECEMBER 2014.

[3] C. Fulton and W. Chappel, " A Dual-Polarized Patch Antenna for Weather Radar Applications", Microwaves, Communications, Antennas and Electornic Systems (COMCAS), 2011 IEEE International Conference.

[4] D. Sun, Z. Zhang, X. Yan, and X. Jiang, " Design of Broadband Dual-Polarized Patch Antenna with Backed Square Annular Cavity", IEEE Transactions on Antennas and Propagation, Volume: 64, Issue 1, Jan. 2016.

Franklin Monopole Antenna with Different Iteration Fractal Elements

Kliment Nikolaev Angelov
Department of Radiocommunications and Videotechnologies
Technical University of Sofia
Sofia, Bulgaria
e-mail: kna@tu-sofia.bg

***Abstract* — In this study a possibility of antenna implementation with elements in shape of fractal curves that form different iterations has been considered. The Franklin antenna as a monopole is the root form in which the main arms are replaced with shapes of first and/or second iteration Koch curves. An optimization, using the design of experiment theory has been made in order to discover the optimal dimensions of the different antenna structures from matching at chosen frequency perspective. The results from simulations with different parameters, such as antenna gain, HPBW, frequency bandwidth, common dimensions etc. have been compared.**

Keywords—Fractal Antenna; Koch's curve; Diferent Iteration Shapes; Matching Optimization; Design of Experiment;

I. INTRODUCTION

Nowadays different approaches to implement antennas with varied parameters and characteristics have been examined. There are many options in this aspect, one of which is to use the so-called fractal antennas in order to achieve beneficial improvements, such as multiband features, decreased dimensions, bigger bandwidth, and others [1][2][3].

It is attractive to include fractal elements in the structure of more complex antennas, such as antenna arrays [4][5], and observe the overall behavior. In this study a linear antenna array with series excitation, better known with the popular name Franklin Antenna, has been chosen for making research of its parameters and characteristics in the case of replaced dipole elements with fractal structures. A good shape for this purpose has the Koch's curve [6] as it is possible to replace different elements with shapes from different iteration. It is possible also to achieve axial radiation, instead of omnidirectional, with excitation of individual element with the appropriate phase of the wave. This can be done with the help of adjusting the phasing stubs that separate the elements.

In this article the results from simulations of antennas with different combination types of fractal elements have been presented. For further experimental research the antennas' matching is optimized in advance to ISM frequency of 433.92 MHz but the attained results can be applied for other frequencies due to the principle of electromagnetic similarity. Also, the antennas are observed as planar structures in order to make possible their accomplishment in printed circuit board (PSB) and/or other planar implementations.

II. ISSUE DESCRIPTION

A. *Fractal Franklin antenna structures*

The main Franklin antenna consists of several radiating dipole elements, separated with phasing stubs. An overview of the antenna can be seen on Fig. 1.

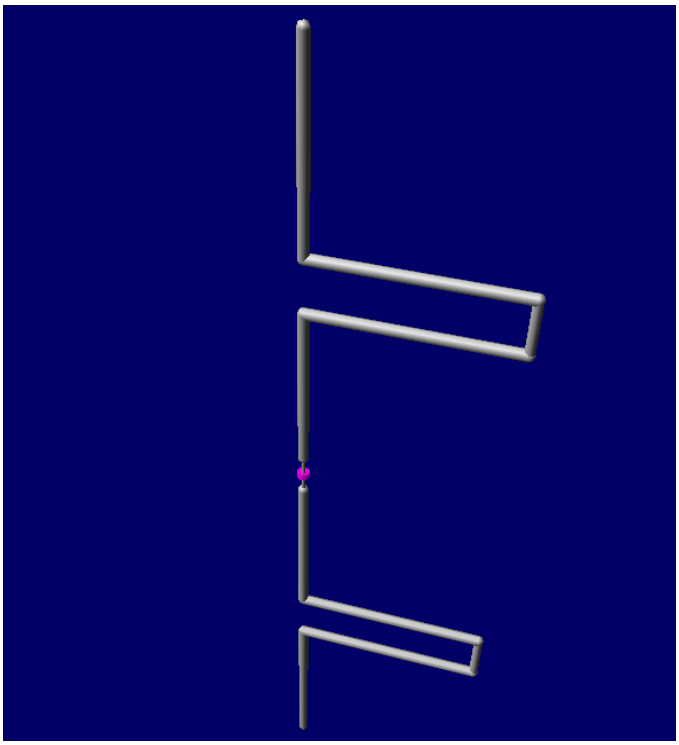


Fig. 1. Overview of a linear antenna array with series excitation (Franklin antenna)

In case of axial radiation pattern the antenna has two maximums of the radiation pattern in direction of the two arms of the antenna. If the antenna is fulfilled as a monopole antenna, as shown on Fig.2, the maximum of the radiation

pattern is one. This type of structure is called with code name ’00’ for easier recognition further in the article.

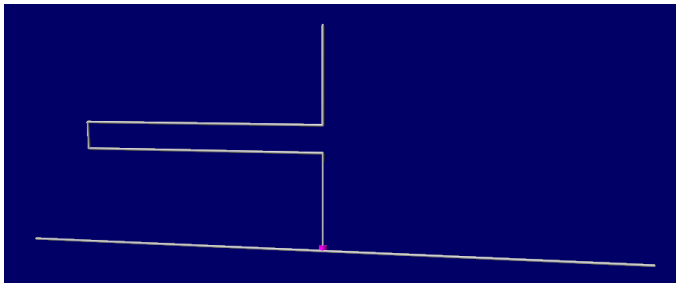


Fig. 2. Franklin antenna monopole (Type code 00)

Fig. 3 and Fig. 4 show the observed antennas with substituted elements with fractal structures.

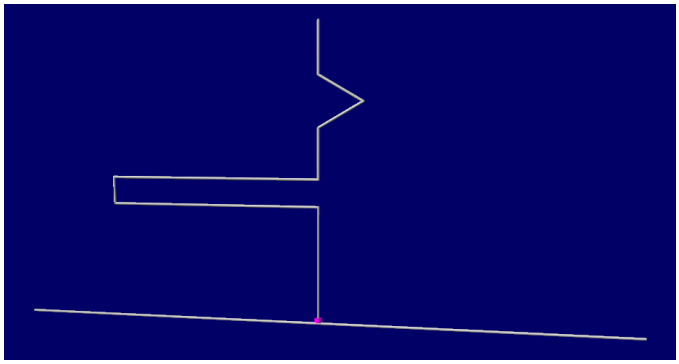


Fig. 3. Antenna with elements from zero and first iteration (code name 01)

The structure on Fig. 3 contains one straight element, which can be observed as a zero iteration of a Koch’s curve and one element from first iteration (structure code name ‘01’).

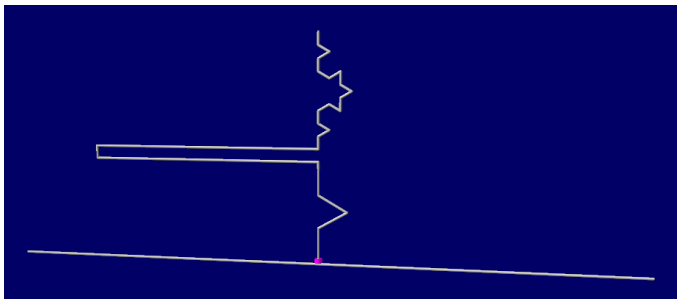


Fig. 4. Antenna with elements from first and second iteration (code name 12)

The structure on Fig. 4 contains one element from first and one from second iteration of Koch’s curve and has a code name ‘12’.

B. Matching the antennas

The antenna structures shown in Fig. 2, Fig. 3 and Fig. 4 need to be matched on the same frequency in order to compare their parameters and characteristics. Such optimization can be achieved with the help of the design of experiment theory. This analytical description can present the behavior of a given parameter in relation to the factors from which it depends. In this case the factors ***x1*** and ***x2*** are the dimensions of the different fractal elements, represented by coefficients ***k11*** and ***k12***. It is necessary to conduct full factorial experiment type 3^2, in which two factors are varied on three levels and the parameter against which the optimization is made has been read. In this case it is appropriate for the parameter ***Y*** to be the antenna matching represented by the standing wave ratio (SWR) at 433.92 MHz. Based on the data generated in nine attempts in a planned experiment type 3^2 this analytical dependence has the form [7]:

$$Y = b_0 + b_1.x_1 + b_2.x_2 + b_{12}.x_1.x_2 + \\ + b_{11}.x_1^2 + b_{22}.x_2^2 \quad , \tag{1}$$

The coefficients ***b***, where ***i,k = 1,2***, are defined according to the following laws:

$$b_0 = \frac{5}{9}\sum_{j=1}^{9} x_{0j}.\bar{y}_j - \frac{1}{3}\sum_{i=1}^{2}\sum_{j=1}^{9} x_{ij}^{\ 2}.\bar{y}_j \quad , \tag{2}$$

$$b_i = \frac{1}{6}.\sum_{j=1}^{9} x_j.\bar{y}_j \quad , \tag{3}$$

$$b_{ik} = \frac{1}{4}.\sum_{i=1}^{2}.\sum_{j=1}^{9} x_{ij}.x_{kj}.\bar{y}_j \quad , \tag{4}$$

$$b_{ii} = \frac{1}{2}.\sum_{i=1}^{2}\sum_{j=1}^{9} x_{ij}^{\ 2}.y_j - \frac{1}{3}\sum_{j=1}^{9}\bar{y}_j \quad , \tag{5}$$

Fig. 5 shows the structures of the observed antennas with marked dimensions for better illustration. In fact, the optimization through the described above planned experiment has been made on the coefficients ***k11*** and ***k12***, which represent the relative segment length of the fractal elements referred to the quarter wavelength ***λ/4***.

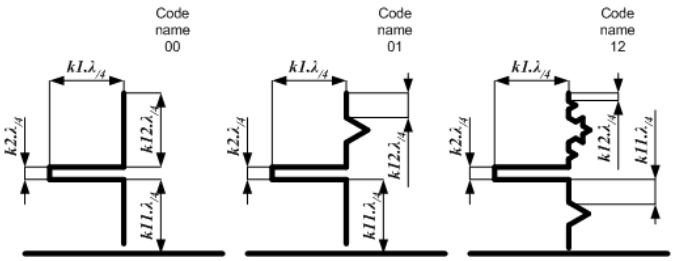


Fig. 5. Dimensions of antenna structures

III. RESULTS

A. Matching optimization

By means of previous basic tests and taking into account that the length of the individual radiating elements should be in the order of $\lambda/4$ the values for the variation of the coefficients ***k11*** and ***k12*** into the factor space are defined in advance. The exact values vs. coded values of ***x1*** and ***x2*** are given in Table I.

TABLE I. FACTOR VARIATION LEVELS

Code Variation Level	Antenna code name 00		Antenna code name 01		Antenna code name 12	
	k11	*k12*	*k11*	*k12*	*k11*	*k12*
+1	0,40	0,40	0,48	0,22	0,132	0,050
0	0,38	0,38	0,45	0,20	0,131	0,049
-1	0,36	0,36	0,42	0,18	0,130	0,048

In Table II are given the values for the optimization parameter (in this case SWR) for the nine different samples of each antenna structure.

TABLE II. EXPERIMENTAL MATCHING VALUES

N	Coded Place in factorial space		SWR		
	k11	*k12*	*Antenna code name 00*	*Antenna code name 01*	*Antenna code name 12*
1	+1	+1	1,57	2,41	1,25
2	+1	0	1,31	1,26	1,11
3	+1	-1	1,09	1,75	1,08
4	0	+1	1,31	1,86	1,20
5	0	0	1,09	1,10	1,06
6	0	-1	1,12	2,40	1,11
7	-1	+1	1,10	1,41	1,18
8	-1	0	1,11	1,49	1,09
9	-1	-1	1,36	3,23	1,16

Through the data from Table II the models according to (1) are estimated. The coefficients of these estimations are given in Table III.

TABLE III. COEFICIENTS FOR THE PLANNED EXPERIMENT'S MODELS

	Antenna code name 00	Antenna code name 01	Antenna code name 12
b_0	1,11444	1,19111	1,07222
b_1	0,06833	-0,28333	0,04667
b_2	0,06667	-0,11833	0,00167
b_{12}	0,18500	0,62000	0,03750
b_{11}	0,08833	0,89333	0,07667
b_{22}	0,08333	0,13833	0,02167

On Fig. 6, Fig. 7 and Fig. 8 is presented the overview of the parameter behavior, estimated with MatLab [8] in the considered factor space for each type antenna structure.

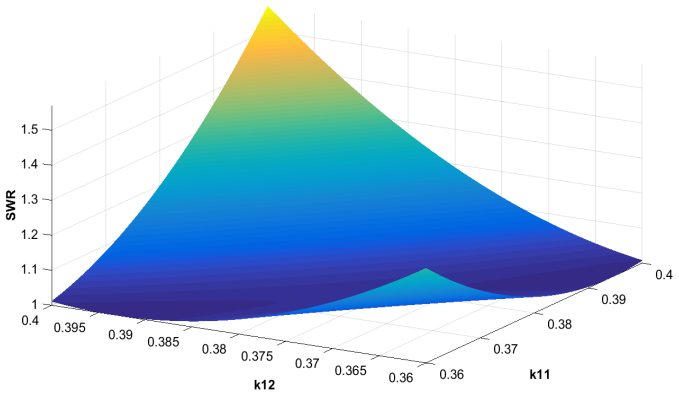


Fig. 6. Model for antenna code 00

Based on the behavior of parameter in the factorial space the optimal points have been found. This defines the values for coefficients ***k11*** and ***k12*** for each antenna structure.

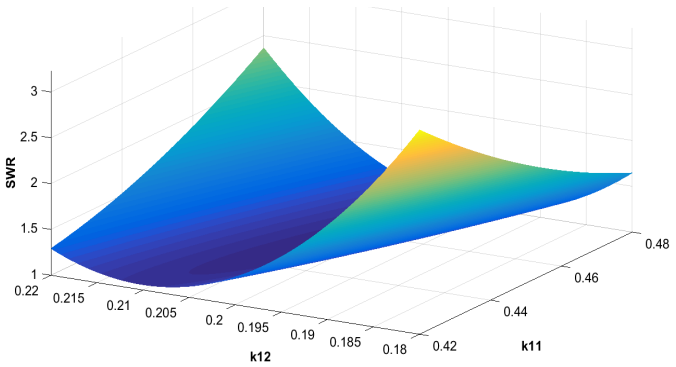


Fig. 7. Model for antenna code 01

The exact values, received through the optimization, are given in Table IV.

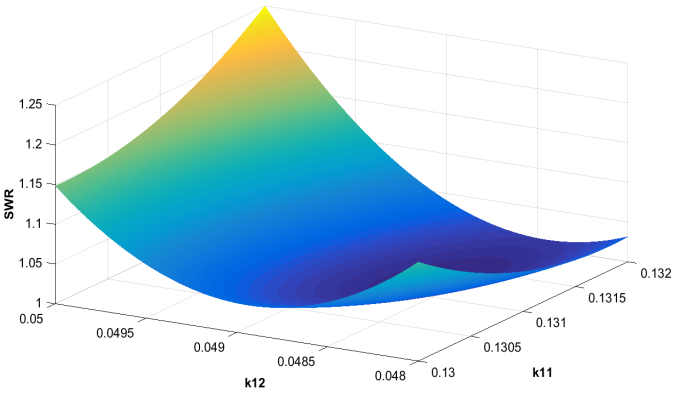


Fig. 8. Model for antenna code 12

Table IV also shows the parameters of the phasing stubs, which are defined again in relative units in reference to the quarter wavelength $\lambda_{/4}$. These are the coefficients ***k1*** and ***k2***, which define the length and the width of the stubs respectively (see Fig. 5). Their values are optimized for the observed frequency too.

TABLE IV. OPTIMISED ANTENNA DIMENSIONS

	Antenna code name 00	**Antenna code name 01**	**Antenna code name 12**
k11	0,37912	0,45559	0,13082
k12	0,37912	0,19905	0,04917
k1	0,98273	0,84551	0,91278
k2	0,10904	0,10904	0,05000

B. Antenna properties

With the optimized antenna structures a number of simulation has been ran in order to determine the antenna parameters, such as gain, half power beam width (HPBW), frequency bandwidth (BW), dimensions, etc. Table V contains the values of these parameters.

TABLE V. OPTIMISED ANTENNA DIMENSIONS

Values @ 433,92 MHz	**Antenna code name 00**	**Antenna code name 01**	**Antenna code name 12**
SWR, -	1,01	1,11	1,09
BW, % *(@ SWR<2)*	3,11	3,34	2,88
G, dBi	3,67	3,42	3,56
HPBW, deg *(@ -3dB)*	75	75	70
Front to Back, dB	3,38	3,77	3,51
Higth of the Antenna, mm	149,9	200,8	152,9

From the data it is visible, that the antennas have similar behavior, but by increasing the curve grade, the beam width goes narrow. The dimensions of antenna type with code names 00 and 12 are almost equal (the difference is just 3 mm), but the radiation pattern of the second antenna is more focused (75 vs. 70 deg) The frequency bandwidth is also narrower (3,11 vs. 2,88 %). Fig. 9, Fig. 10 and Fig. 11 show the 3D radiation pattern for each antenna.

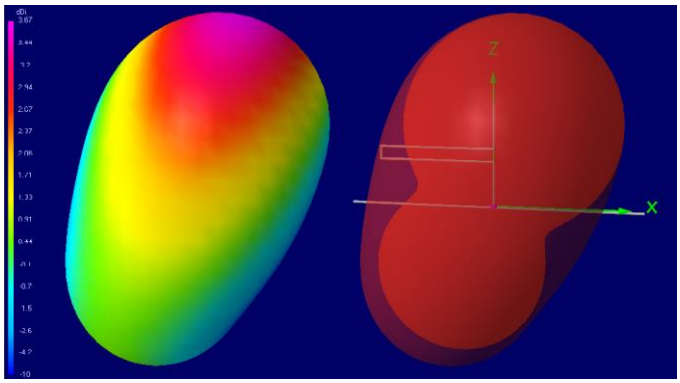


Fig. 9. 3D Radiation pattern for antenna code 00

Antennas with this type of parameters can be used in narrow band frequency communications with applications with high integration grade, such as RFID and Access Control systems. The specific radiation patterns can be useful in case of spot orientated short distance data transfer and to minimize the interference between devices operating at the same frequency in little space area.

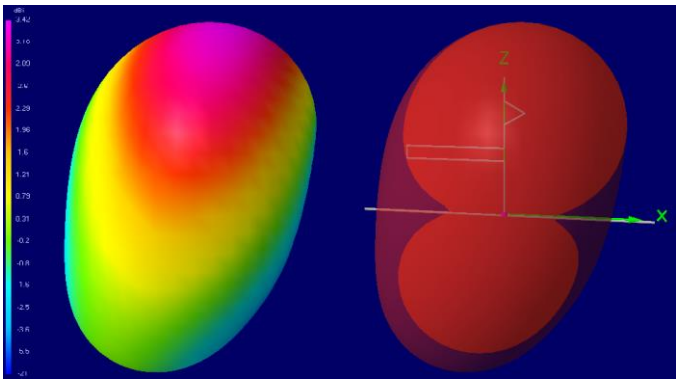


Fig. 10. 3D Radiation pattern for antenna code 01

From fig. 9 – 11 can be seen, that the radiation pattern for the antenna type with code name 01 and 12 have less radiation in the X-Y plane, which is better in case of implementing a spot zone in the direction of Z-axis.

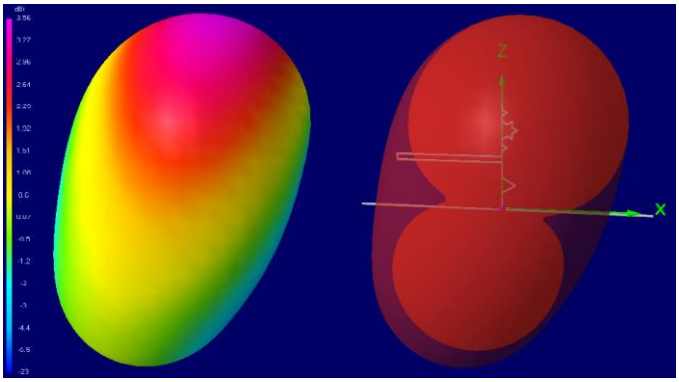


Fig. 11. 3D Radiation pattern for antenna code 12

IV. CONCLUSION

In this study an observation of antennas contained different iteration fractal elements has been made. From the results can be seen, that:

- the proposed structures can be precisely matched and achieved similar results as Gain, HPBW, F/B and dimensions;
- the planar type of this antennas gives the opportunity to implement them in PCB or other flat structures;
- it is appropriate to study furthermore this type of structures in order to achieve other useful parameters.

ACKNOWLEDGEMENT

This work was supported in part by the Grant DN07/19/15.12.2016 "Methods for Estimation and Optimization of Electromagnetic Radiation in Urban Areas" of the Bulgarian Science Fund.

REFERENCES

[1] Adnan S.M., P. Asthana, O.P. Singh, S. Mishra, M.M. Siddiqui, Introduction to fractal antenna, International Journal of Innovative Research in Elacrtical, Electronics, Instrumentation and Control Engineering, Vol. 4, Issue 2, Feb 2016;

[2] Cohen N., Fractal antennas: Part I, Communications Quarterly, Summer, pp-7-22, 1995;

[3] Werner D. H., S. Gangul, An overview of fractal antenna engineering research, IEEE Anlennasand Propagation Magazine. Vol. 45, No. I, February 2003;

[4] El-Khamy S.E., M.A. Aboul-Dahab, M.I. Elkashlan, A simplified Koch multiband fractal array using windowing and quantization techniques, IEEE international Symposium on Antennas and Propagation Digest, Salt Lake City, Utah, vol. 3, pp.1716-1719, 2000;

[5] Gianvittorio J.P.,Y. Rahmat-Samii, Fractal elements in array antennas: Investigating reduced Mutual Coupling and tighter packing, IEEE international Symposium on Antennas and Propagation Digest, Salt Lake City, Utah, vol. 3, pp.1704-1707, 2000;

[6] Singh K., V. Grewal, R. Saxena, Fractal antennas: A novel miniaturization technique for wireless communications, International Journal of Recent Trends in Engineering, Vol 2, No. 5, November 2009;

[7] Vuchkov I.N., Bojadjieva L.N., Quality Improvement with Design of Experiments. A response Surface Approach, "Kluwer Academic Publishers", Dordrecht, Netherland, 2001.

[8] MATLAB IMAGE PROCESSING TOOLBOX. User's Guide, "The Math - Works Inc.", 2000. www.mathworks.com.

Blockage Arrangements Modelling in Indoor Wireless Networks: A Comparative Study

Viktor Stoynov[1] and Zlatka Valkova-Jarvis[2]

Abstract – **In this work, the performance of indoor users' throughput is studied through the abstract modelling of walls. A more realistic blockage object arrangement is compared with existing abstract-modelled blockage arrangements to demonstrate its improved representation of real-life scenarios. Ten different scenarios are defined, combining different wall layouts and transmitter dispositions. The scenarios represent possible real working environment sites and are experimentally tested for different numbers of receivers.**

Keywords – **Indoor environment, wireless communications, blockage modelling, user throughput**

I. Introduction

One of the main problems facing telecommunication networks is how to provide an excellent service to users located at the periphery of the serving cell. These users are subject to significant interference from neighbouring base stations and, in the case of wireless indoor environments, from adjacent transmitters (*Tx*). One major obstacle in enclosed spaces is the wall layout: walls, as blockage objects, mitigate interference but also cause the signal to deteriorate, thus worsening the quality of the mobile services provided. Good indoor coverage depends to a large extent on the femtocells' location. The signals from non-serving transmitters permeate the wall and fade easily, hence they are not a source of interference.

The modelling of indoor obstacles is often neglected and the enormous influence of wall layout on signal propagation is thus overlooked in the research. Some recent works [1], [2], [3] have presented and investigated different wall generation models of indoor communication environments, and considered the major parameters of signal obstacles, such as length, attenuation level, density allocation, etc. Analytical expressions of the average attenuation of signals passing through walls are derived and system-level simulations performed to demonstrate the impact of the walls and transmitter devices arrangement on the Signal-to-Interference Ratio (SIR) and users' throughput.

In this paper a new, more realistic, abstract-modelled wall layout is investigated. It demonstrates clear advantages when compared experimentally to previously-developed similar abstract models. Furthermore, ten scenarios, consisting of four different types of wall distributions and free indoor space propagation, are evaluated and compared. The abstract wall generation methods use the same wall density, aiming to achieve clear conclusions after the direct comparison of users' throughputs. The experimental set uses the same enclosed space (Region of Interest - RoI), number of *Tx*, transmitter distance and power, while the number of users (receivers - *Rx*) varies. Using the same amount of physical resources in each scenario while increasing the number of users leads to a decrease in average user throughput.

Viktor Stoynov[1] and Zlatka Valkova-Jarvis[2] are with the Faculty of Telecommunications at Technical University of Sofia, Bulgaria, E-mails: {vstoynov, zvv}@tu-sofia.bg.

II. System Model

A. Abstract wall layouts generation methods

In this work, four methods for wall arrangement are considered (Fig. 1). The first wall generation method is based on a Boolean scheme, where the positions of the centre points of the walls are randomly distributed according to a Poisson Point Process (PPP) of density λ. The lengths of the walls follow arbitrary distribution $f_L(l)$. The disposition of the walls is either parallel or at right-angles, which defines a two-state wall layout, realised when the angle between any two walls is a binary choice – $\{0; \pi/2\}$. This abstract wall generation method is denoted as **[binary]** (Fig. 1a).

The wall distribution, generated by a Manhattan grid of equidistantly-spaced walls is named **[regular]** (Fig. 1b). It is assumed that the walls are oriented perpendicularly to the coordinate axes. The space between every two adjacent parallel walls is set to a constant Δ. This distance is calculated based on the dimensions of the considered RoI and is related to the average wall length $E[L]$ and wall density parameter λ: $\Delta=2/\lambda E[L]$. Geometrically, the RoI is a rectangle or square with sides whose length is both an integer and a multiple of Δ. In order to achieve different realisations of the **[regular]** wall layout, Δ might be randomly shifted by δ_x in the *x*-axis and by δ_y in the *y*-axis.

The third abstract, but more realistic, wall generation method is obtained by two Manhattan Line Processes (MLP) and is named after it - **[MLP]** (Fig.1c). This method is very similar to the **[regular]**, but differs in that the wall distance Δ is not a constant but a variable.

The fourth method of blockage arrangement named Realistic Indoor Environment Generator - **[RIEG]** (Fig.1d) is designed as a simple generation tool of indoor environments. It consists of positioning rectangles in a predefined area – RoI. The coordinates of the starting point of each rectangle are selected randomly. The sides of the rectangle are then plotted, ensuring that the rectangle remains within the RoI. The total length of the walls takes into account the value of the wall density parameter. Hence, different methods for wall arrangement can be compared. The confined spaces, resulting from the rectangles' distribution, achieve a more realistic indoor design compared to the other three methods.

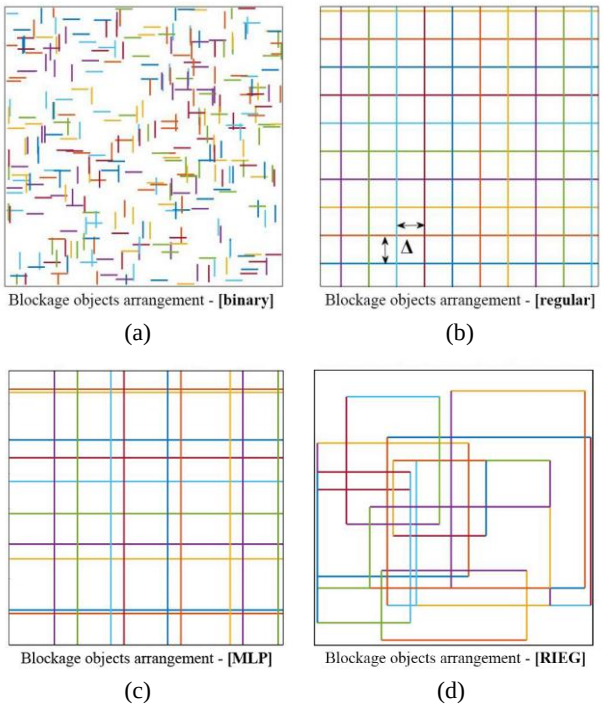


Fig.1. Generated wall maps for (a) **[binary],** (b) **[regular],** (c) **[MLP]** and (d) **[RIEG]** cases

There are also scenarios where no walls are distributed – **[free space]**. Thus, the experimental *results* can show how the existence of walls affects the level of user throughput.

B. Transmitter and receiver location

Four *Txs* in the indoor system model are located in the vertices of a square with side-length *R*, denoted as **[square]** in the scenarios' descriptions (Fig. 2a). When the **[square]** constellation of transmitters is rotated by an angle of $\pi/4$, the alternative transmitters' disposal is obtained. It is labelled as **[rhomboid]** and is shown in Fig. 2b. These two transmitter setups simplify the theoretical analysis of the system interference.

The receivers are located at the cell edge, at a constant distance of *R*/2 from the closest transmitter. This transmitter is denoted as the desired transmitter – (*dTx*). The other three transmitters are assumed to be sources of interference ($iTx_{1\div3}$). The position of each receiver is determined by its polar coordinates ($R/2$, Φ), measured against the nearest transmitter. The angle Φ ranges from 0 to $\pi/2$ (Fig. 2).

III. ANALYTICAL MODEL

The downlink signal is assumed to experience attenuation due to the wall blockages, distance-dependent path loss and small-scale fading. The path loss law *l(d)* is defined in [4]. The attenuation due to the walls is determined by summing the attenuation of each wall.

In the models considered in this work, the blockages are defined as two-dimensional objects, and the investigated wireless network is designed to be interference limited.

One of the most important parameters is the average number of blockages $E[K]$ that obstruct the path with length d between the *Tx* and the *Rx*. For the **[binary]** case, this can be expressed as:

$$E[K] = \lambda E[L] d \frac{\left(|\sin(\phi)| + |\cos(\phi)|\right)}{2}. \quad (1)$$

where ϕ denotes the angle of the link between transmitter and receiver against the *x*-axis.

It is clear that the average number of blockages $E[K]$ located between a *Tx* and a *Rx* is directly proportional to the average length of these wall objects $E[L]$.

For the **[regular]** case, $E[K]$ is calculated as:

$$E[K] = N_x + N_y + p_x + p_y, \quad (2)$$

where N_x and N_y denote the number of walls without random shifts δ_x or δ_y, while p_x and p_y are the number of new walls, required to preserve the average wall density, after a random shifting is performed.

When the number of walls is set to K_i, the total attenuation of the signals in the area will be $\omega_i=\omega^{K_i}$. Although each wall may have a different attenuation, the experiments conducted here consider 10 dB fixed attenuation. Thus the SIR for one indoor user will be:

$$\gamma = \frac{P_0 h_0 l(d_0) \omega_0}{\sum_{i=1}^{3} P_i h_i l(d_i) \omega_i}, \quad (3)$$

where d_o is the distance between the receiver *Rx* and its serving (desired) transmitter *dTx*, P_0 is the transmit power of *dTx*, while P_i (i = 1, 2, 3) is the power of the interfering transmitters iTx_1, iTx_2 and iTx_3, respectively. h_0 and h_i denote the small-scale fading, d_i is the distance between the receiver and the $i^{\text{-th}}$ interfering transmitter and $l(d_0)$ and $l(d_i)$ are the path losses.

In [1] an analytic expression to approximate *geomean* (γ) for the **[binary]** case is derived. The average SIR is calculated by:

$$geomean(\gamma) \approx \int_{-\infty}^{\infty} \left(\frac{d}{d\delta} \left(1 - \prod_{i=1}^{3} \frac{1}{1 + \delta \frac{\overline{\omega}_i'}{\overline{\omega}_0} \frac{l(d_i)}{l(d_0)}} \right) \Bigg|_{\delta=t} \right) t \, dt. \quad (4)$$

$\overline{\omega}_i'$ provides an accurate approximation for *geomean* (ω_i) and is called effective wall attenuation.

For **[MLP]** the SIR is calculated by splitting the line-processes into the horizontal and vertical processes (Fig. 2a) [2]. For the **[square]** case, the SIR is calculated by:

$$\gamma = \frac{h_0 d_0^{-\alpha} \omega^{K_v + K_h}}{h_1 d_1^{-\alpha} \Omega_{sq\_1} + h_2 d_2^{-\alpha} \Omega_{sq\_2} + h_3 d_3^{-\alpha} \Omega_{sq\_3}}, \quad (5)$$

where

$$\Omega_{sq\_1} = \omega^{K_v' + K_h}; \quad \Omega_{sq\_2} = \omega^{K_v' + K_h'}; \quad \Omega_{sq\_3} = \omega^{K_v + K_h'}. \quad (6)$$

K_v and K_h are the wall counts between any user and the *dTx* and K'_v and K'_h are the wall counts between the user and the interfering transmitters iTx_1, iTx_2, iTx_3 (Fig. 2a). They all are Poisson Random Variables.

For the **[rhomboid]** case (Fig. 2b), the SIR is defined by the equation:

$$\gamma = \frac{h_0 d_0^{-\alpha} \omega^{K_v + K_h}}{h_1 d_1^{-\alpha} \Omega_{rh\_1} + h_2 d_2^{-\alpha} \Omega_{rh\_2} + h_3 d_3^{-\alpha} \Omega_{rh\_3}}, \quad (7)$$

where

$$\Omega_{rh\_1} = \omega^{K'_v + S(-\cos(\phi))K_v + K'_h}; \quad \Omega_{rh\_2} = \omega^{K_v + K'_h + K''_h}; \quad \Omega_{rh\_3} = \omega^{K''_v + S(-\cos(\phi))K_v + K'_h}. \quad (8)$$

K''_v and K''_h denote the wall counts between any user and the interferers iTx_1, iTx_2, iTx_3 as shown in Fig. 2b.

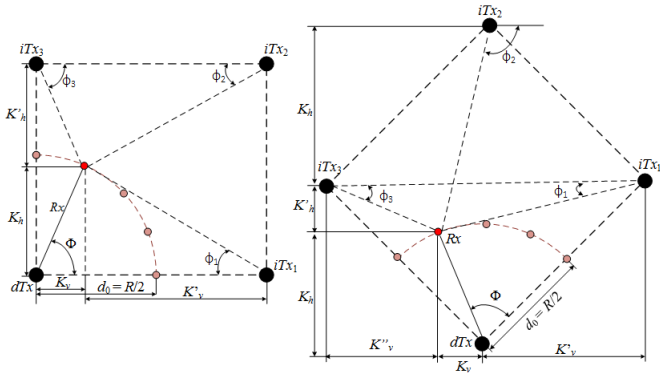


Fig.2. Transmitters (*Tx*) and receivers (*Rx*) location (a) **[square]** and (b) **[rhomboid]**

IV. System-level simulations and results

A. Network design

The Vienna LTE-A system-level simulator [5] is an appropriate tool for the abstract modelling of the scenarios considered in this work. The results for users' throughput are obtained after 500 simulation runs, each with 200 identical transmission time intervals. The wall attenuation research conclusions are the result of 5000 simulations. The four transmitters are spaced a distance of R=40m apart, regardless of the pattern of their arrangement - **[square]** or **[rhomboid]**. Each *Tx* is defined as a femtocell with a transmit power of 100mW. The distance between the *dTx* and the receivers is set to 20m. The wall density is λ=0.05m^{-2}, the average wall length is $E[L]$=5m and the wall attenuation is set to 10 dB for each simulation. The experiments are conducted under the same conditions - the average number of walls, receivers and transmitters remaining constant for each simulation. The defined ten scenarios are simulated for two distinct numbers of users – 5 and 15.

Combining the location of the transmitters and wall layouts, the following ten scenarios are defined: **S1 = {[binary], [square]}**, **S2 = {[binary], [rhomboid]}**, **S3 = {[regular], [square]}**, **S4 = {[regular], [rhomboid]}**, **S5 = {[MLP], [square]}**, **S6 = {[MLP], [rhomboid]}**, **S7 = {[free space], [square]}**, **S8 = {[free space], [rhomboid]}**, **S9 = {[RIEG], [square]}**, **S10={[RIEG], [rhomboid]}.**

B. Average wall attenuation and user throughput results

The average wall attenuation and SIR can be analytically calculated only for the **[binary]** wall distribution. The **[regular]** wall pattern scenarios can only be examined using simulations. In [1] and [2] the wall attenuations for the different scenarios are determined. The **[binary]**, **[regular]** and **[MLP]** scenarios are analytically verified and it is shown that analytical results and simulation curves for the average attenuation level per transmitter match perfectly for both **[square]** and **[rhomboid]** transmitter layouts.

Fig. 3 and 4 show the average throughput of each user when two different numbers of *Rx* positions are used – 5 and 15 respectively. Investigations using a larger number of users are performed in [6] and clear conclusions are drawn.

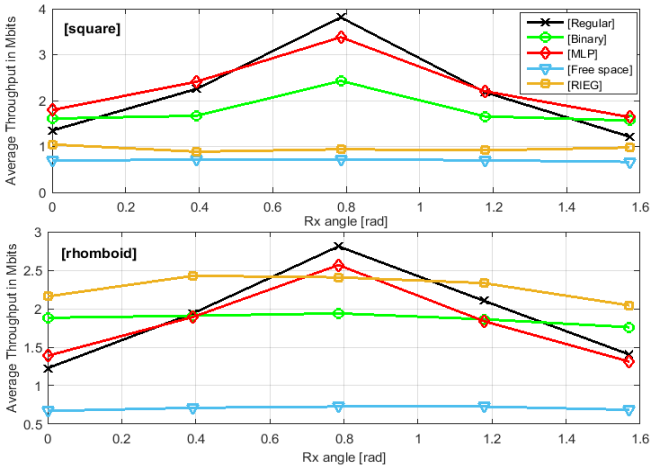


Fig.3. Average user throughput results for 5 users and **[square]** and **[rhomboid]** transmitter layouts

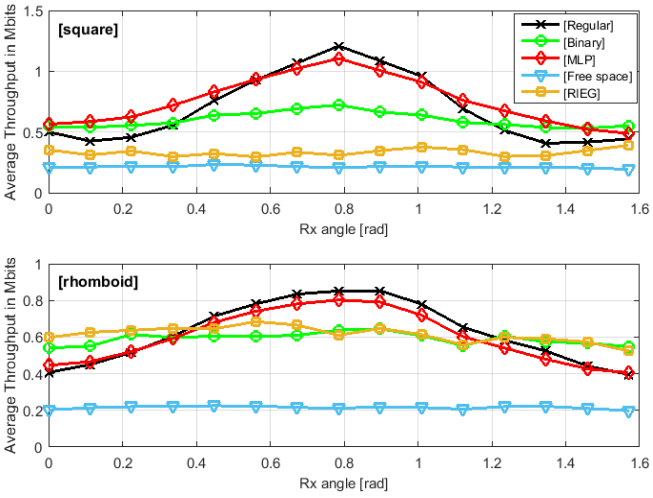


Fig.4. Average user throughput results for 15 users and **[square]** and **[rhomboid]** transmitter layouts

As anticipated, when **[free space]** scenarios are considered, the average throughput for each user dramatically decreases, which proves the important role of obstacles in interference mitigation. When comparing the simulation results for **[square]** and **[rhomboid]** transmitter layouts and an equal number of users, the same values of the throughputs are obtained. The location layouts of transmitters and receivers lose relevance when there are no walls – scenarios **S7** and **S8**. The angular collocation of the receiving and transmitting devices affects the wireless network performance only when direct visibility is impaired due to wall obstruction.

Apart from the users in the periphery of the femtocell, the graphs follow the same trend results regarding SIR as reported in [2]. The Manhattan grid-like wall arrangements, denoted as **[regular]** and **[MLP]**, show better performance compared to the **[binary]** and **[RIEG]** cases, no matter how many *Rx* positions are explored. This is caused due to the walls located along the *x*- and *y*-axis affecting the signal propagation between the *dTx* and the users to a much lesser extent. These walls, however, are able to suppress the interference from other transmitters ($iTx_{1\div3}$). Being a stochastic method, the distribution of the walls in **[binary]** is much more difficult to predict compared to **[regular]** and **[MLP]**.

In the case of **[MLP]** and **[regular]**, the **[square]** layout is characterised by lower interference compared to the **[rhomboid]** transmitters location, due to the different angular co-location of the walls to the desired transceiver. Walls parallel to the x- and y-axis remain invisible to *dTx*, which increases the throughput of the users located in positions $\Phi=0$ and $\Phi=\pi/2$. At these positions, the main sources of interference are the nearest distributed transmitters – iTx_1 and iTx_3 respectively (Fig. 2). The **[binary]** arrangement may provide fewer blockages to signal propagation, especially for user positions in $\Phi=0$ and $\Phi=\pi/2$, in contrast to the **[regular]** and **[MLP]** wall arrangements, which always contain obstacles between the *Rx* and the *dTx*. An increase of throughput is observed for *Rx* positions around $\Phi=\pi/4$, where the three interfering transmitters have an equally strong impact on the total interference. Clearly, the **[regular]** grid offers the best protection for users against interference.

It turns out that in the **[RIEG]** wall arrangement and **[square]** *Tx* locations, users receive nearly equal quality of service. Unlike the **[regular]** and **[MLP]** methods providing strictly determinate locations for the walls, when the **[RIEG]** layout is used the walls between users and transmitters are always random in number. In the **[rhomboid]** case, each wall can suppress the interference caused by $iTx_{1\div3}$, resulting in better coverage especially for users in the furthest positions compared to **[square]**. Logically, we may conclude that the more realistic the distribution of the walls used, the harder it becomes to predict their impact on user throughput. Obviously, due to the stochastic nature of wall arrangement the performance of **[RIEG]** is very similar to **[binary]**.

An interesting fact is that the **[MLP]** curves are very closely located to those for **[regular]** - something which is not observed in the figures for SIR. The reason for such behaviour is the identical number of walls for both generation methods although their wall layouts appear to be different. The reduction of the users' throughput which was observed in a part of the simulations for the **[MLP]** wall layout is due to the greater probability that fewer obstacles will be located between the *Rx* and $iTx_{1\div3}$ compared to those between *Rx* and the *dTx*, which may result in increased interference (Fig. 1c). Conversely, the [**regular**] wall generation method provides roughly the same number of obstacles between users and the desired transmitter, and users and the interfering transmitters for each iteration (Fig. 1b).

V. Conclusion

In this work, simulation and analytical results for average wall attenuation and system-level simulation results evaluating the average users' throughput are presented. Ten scenarios are composed and investigated determined by four different wall layouts and two different transmitter locations.

The experimental results for average user throughput obtained via system-level simulations show the same trends as the SIR performance [2]. It transpires that, due to different wall arrangements, the average wall attenuation and average throughput are angularly dependent. When no walls are used, i.e. free space of signal propagation is taken into account, the specific transmitter arrangement does not influence the average throughput of the users.

The wall distribution models can be used to achieve more realistic indoor environments in order to test different techniques for interference mitigation. The closest to a real-world scenario is the **[RIEG]** wall arrangement method, which is realistic enough to be used as an indoor environment to test interference suppression techniques and algorithms for realistic human mobility and intelligent resource allocation.

References

[1] M. Müller, M. Taranetz, V. Stoynov, M. Rupp. "Abstracting Indoor Signal propagations: Stochastic vs. Regular", 58th International Symposium ELMAR-2016, Zadar, Sept. 2016, pp. 249-252.

[2] M. Müller, M. Taranetz, and M. Rupp, "Analyzing Wireless Indoor Communications by Blockage Models," in IEEE Access, Volume PP, Issue 99, Dec. 2016, .

[3] V. Stoynov. "Investigation of Indoor Wireless Communication Environment Using Abstract Modelling", E+E, vol. 52 No 1-2/2017, pp. 26-31.

[4] 3rd Generation Partnership Project (3GPP), "Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA physical layer aspects," 3rd Generation Partnership Project (3GPP), TR 36.814, Mar. 2010.

[5] M. Rupp, S. Schwarz, and M. Taranetz, The Vienna LTE Advanced Simulators: Up and Downlink, Link and System Level Simulation, 1st ed., ser. Signals and Communication Technology. Springer Singapore, 2016.

[6] V. Stoynov, Z. Valkova-Jarvis, "The Impact of Abstract-Modelled Wall Layouts on User Throughput in Wireless Indoor Environments", First International Balkan Conference on Communications and Networking (BalkanCom 2017), Tirana, Albania, May 30-June 2, pp. 1-2.

DIGITAL IMAGE AND SIGNAL PROCESSING, AND ELECTRONICS – ORAL SESSION

Examination of the polar response of a circular piston acoustic transducer with a cone form of the membrane

Eng. Georgi Markov[1]

Abstract – **In this paper the problem about polar field of a circular piston transducer with a cone form of the membrane is considered. Theoretical expression is proposed for calculating sound pressure level (SPL) and plot polar diagram of a circular piston transducer with a cone form of the membrane. Conclusions with practical value have been made. Experimental measurements of the polar response diagram have been made.**

Keywords – **polar response, polar diagram, sound pressure level, circular piston transducer, conical membrane.**

I. INTRODUCTION

The acoustic transducer's radiation pattern is of great interest recently due to the increased demand of quality radiators for echography, geology and architectural acoustics[1]. Furthermore requirements for the design and modeling of loudspeakers and microphones for audio systems are increased[2]. Therefore, a theoretical analysis of the expressions for calculating the directional diagram of a circular piston acoustic transducer with conical form is made. Analytical expression for calculating polar diagram of circular piston with conical shape radiator is proposed. An experiment has been conducted to measure a cone shape loudspeaker polar diagram to demonstrate the practicability of the proposed expression.

II. SOLUTION OF RESEARCH PROBLEM

A. Theoretical background

The purpose of this study is a circular piston transducer with a conical shape of the membrane. Because of its symmetry with respect to the z axis (fig. 1) it can be assumed that conical piston polar pattern will be one with rotation symmetry. In order to determine this symmetry, it is sufficient to determine the sound pressure level at a distance *d* in a plane whose norm is perpendicular to the z-axis.

B. Aanalytical presentation of the problem

The acoustic pressure generated by an emitter in the environment can be calculated by assuming that its surface consists of a plurality of elementary sections *dS* that can be obtained by the solution of a differential equation (fig.1).

$$dS = \sqrt{E \cdot G - F^2}\, d\alpha d\rho \qquad (1)$$

where

$$E = \left(\frac{\partial x}{\partial \alpha}\right)^2 + \left(\frac{\partial y}{\partial \alpha}\right)^2 + \left(\frac{\partial z}{\partial \alpha}\right)^2$$

$$G = \left(\frac{\partial x}{\partial \rho}\right)^2 + \left(\frac{\partial y}{\partial \rho}\right)^2 + \left(\frac{\partial z}{\partial \rho}\right)^2$$

$$F = \frac{\partial^2 x}{\partial \alpha\, \partial \rho} + \frac{\partial^2 y}{\partial \alpha\, \partial \rho} + \frac{\partial^2 z}{\partial \alpha\, \partial \rho}$$

where

$x = r \cdot \rho \cdot \cos\alpha$
$y = r \cdot \rho \cdot \sin\alpha \quad 0 \le \alpha \le 2\pi \quad u \quad 0 \le \rho \le 1$
$z = \rho \cdot h$

which are the coordinates of the point of the elementary emitter E.

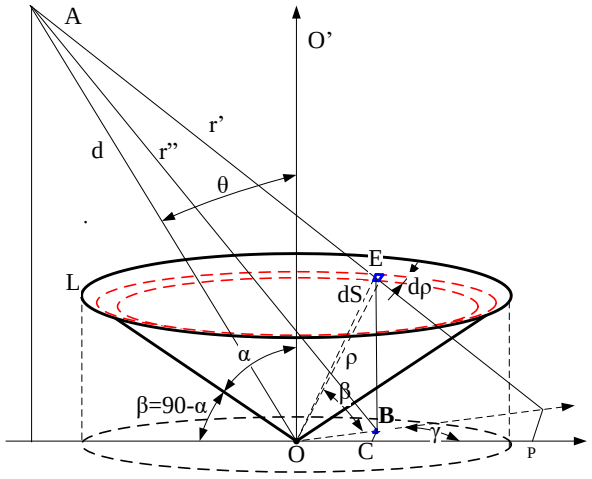


Fig. 1. Geometry of conical piston acoustic transducer.

As a result of the solution of the equations, the following result was obtained:

$$dS = \rho \cdot r\sqrt{r^2 + h^2}\, d\gamma d\rho \qquad (2)$$

where

- ρ and γ are current radius vector and angle
- h – height of the cone transducer
- r – radius of the piston

Each elementary section dS creates a sound pressure at the point of observer A. Point A is located at a distance r' from the point of elementary radiator E and angle θ relative to axis z and has the following coordinates:

$x_1 = d \cdot \sin\theta \cdot \cos\varphi$

[1] eng. Georgi Markov is PhD student with the Department of Communication Equipment and Technologies at Technical University of Varna, 1 Studentska str, Varna, 9000, Bulgaria, E-mail: georgi.s.markov@mail.bg.

$y_1=d\cdot sin\theta\cdot sin\varphi$, $-\pi/2\leq\theta\leq\pi/2$ $0\leq\varphi\leq\pi$
$z_1=d\cdot cos\theta$

where

- d – distance from the center of the cone to A.
- θ and φ current angles on the z and x axes.

C. Aanalytical solution

From analytical geometry the distance r' (from E to A) is determined by the expression:

$$r' = \sqrt{(d\cdot sin\theta\cdot cos\varphi - r\cdot\rho\cdot cos\gamma)^2 + (d\cdot sin\theta\cdot sin\varphi - r\cdot\rho\cdot sin\gamma)^2 + (d\cdot cos\theta - \rho\cdot h)^2} \quad (3)$$

Due to the rotational symmetry, r' is not dependent on the angle φ along the x - axis and it can be assumed that φ = 0 ° and for r' is obtained:

$$r' = \sqrt{(d\cdot sin\theta - r\cdot\rho\cdot cos\gamma)^2 - (r\cdot\rho\cdot sin\gamma)^2 + (d\cdot cos\theta - \rho\cdot h)^2} \quad (4)$$

The elementary section *dS* creates a sound pressure at a point A which is at a distance r' and at an angle θ relative to the OZ axis. This elementary sound pressure is [3]:

$$dp_\theta = \frac{r_0\cdot p_m}{r'} e^{j(\omega t - kr')} \quad (5)$$

where

- r' is the distance from the elementary emitter to the point of observer A.
- p_m - the amplitude of the sound pressure
- r_0 distance OE which is distance from the center of the cone to emitting section *dS*.

$$p_m = \rho_s\cdot c_0\cdot k\cdot r_0\cdot v_m \quad (6)$$

where

- c_0 - the speed of the sound
- $k = \frac{2\pi}{\lambda}$ - a wave number
- ρ_s - density of the environment
- λ - wavelength

After multiplying both sides of equation (6) with and also multiplying the right side with $\frac{2\pi}{2\pi}$, taking into account that the elementary area of the emitter is $dS = 2\pi r_0^2$ therefore:

$$r_0\cdot p_m = \frac{\rho_s\cdot c_0\cdot k\cdot v_m}{2\pi} dS \quad (7)$$

If the expressions (2) and (7) are replaced in (5) the elemental sound pressure in the direction determined by the angle θ will be:

$$dp_\theta = \frac{\rho_s\cdot c_0\cdot k\cdot v_m}{2\pi\cdot r'} e^{j(\omega t - kr')}\rho.r\sqrt{r^2+h^2}\,d\gamma\,d\rho \quad (8)$$

After replacing an expression (4) with an expression (8) and integrating the piston area, an expression proposed by the author is obtained for sound pressure level (SPL) determination of a circular piston acoustic transducer with conical shape of membrane:

$$p_\theta = \frac{\rho_s\cdot c_0\cdot k\cdot v_m}{2\pi\cdot r'} e^{j\omega t}\int_0^r\int_0^{2\pi} \frac{e^{-jk\sqrt{(d\cdot sin\theta - r\cdot\rho\cdot cos\gamma)^2 - (r\cdot\rho\cdot sin\lambda)^2 + (d\cdot cos\theta - \rho h)^2}}}{\sqrt{(d\cdot sin\theta - r\cdot\rho\cdot cos\gamma)^2 - (r\cdot\rho\cdot sin\lambda)^2 + (d\cdot cos\theta - \rho\cdot h)^2}}\rho.r\sqrt{r^2+h^2}\,d\gamma\,d\rho \quad (9)$$

Radiation pattern G(θ) is derived from the ratio[3,4]:

$$G(\theta) = \frac{p_\theta}{p_{0^\circ}} \quad (10)$$

where

- p_θ is the sound pressure amplitude in the direction determined by the angle θ;
- p_{0° is the sound pressure on the acoustic axis.

Therefore, in order to obtain the radiation pattern, it is necessary to determine the SPL for θ = 0 °[5,6]. Due to the zero value of the angle θ the expression (9) yields the type:

$$p_{0^\circ} = \frac{\rho_s\cdot c_0\cdot k\cdot v_m}{2\pi\cdot r'} e^{j\omega t}\int_0^1 2\pi\cdot r\cdot\sqrt{r^2+h^2}\cdot\frac{e^{jk\sqrt{(\rho\cdot r)^2 + (d-\rho\cdot h)^2}}}{\sqrt{(\rho\cdot r)^2 + (d-\rho\cdot h)^2}}\,d\rho \quad (11)$$

In general, after replacing an expression (9) and an expression (11) an expression (10) produces [7]:

$$G(\theta)=\frac{\int_0^{r}\int_0^{2\pi}\frac{e^{-jk\sqrt{(d\cdot\sin\theta-r\cdot\rho\cdot\cos\gamma)^2-(r\cdot\rho\cdot\sin\lambda)^2+(d\cdot\cos\theta-\rho\cdot h)^2}}}{\sqrt{(d\cdot\sin\theta-r\cdot\rho\cdot\cos\gamma)^2-(r\cdot\rho\cdot\sin\lambda)^2+(d\cdot\cos\theta-\rho\cdot h)^2}}\rho.r\sqrt{r^2+h^2}\,d\gamma\,d\rho}{\int_0^{1}2\pi\cdot r\sqrt{r^2+h^2}\cdot\frac{e^{jk\sqrt{(\rho r)^2+(d-\rho h)^2}}}{\sqrt{(\rho\cdot r)^2+(d-\rho\cdot h)^2}}\,d\rho} \quad (12)$$

Expression (12) is proposed by the author for calculating a radiation pattern of a circular cone acoustic transducer.

In the case of a distant zone, the distance is much larger than the diameter of the radiator (d >> 2r), it can be assumed that with respect to the amplitudes, the distances are commensurate with $d = r' = r''$[3,8,9], and we can assume that a known expression is applied to the distances for a flat piston which is a projection of the cone on the xOy plane [10] fig. 2:

$$G_f(\theta)=2\cdot\frac{I_1(k\cdot r\sin\theta)}{k\cdot r\cdot\sin\theta} \quad [10](13)$$

where I_1 is Bessel's function in first-order about argument ($krsin\theta$).

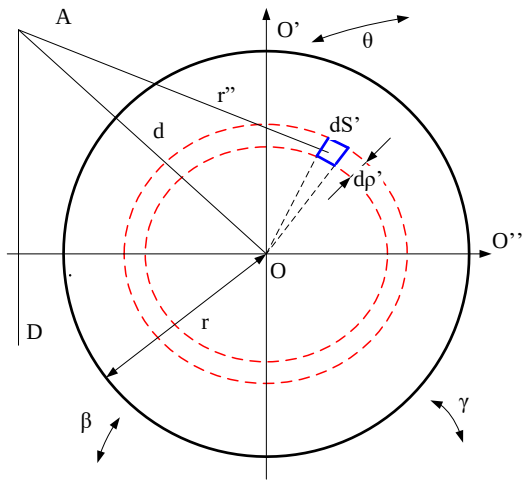


Fig.2. Flat piston transducer geometry [9,10] which is a projection of cone in xOy plane.

D. Experimental solution

To demonstrate the applicability in practice of the proposed expression, an experiment has been performed with measurement an radiation pattern of Selenium 10MB3P loudspeaker. Which is with the following geometric characteristics:

- diameter - 2r = 229 mm = 0.229 m
- height of conical membrane - h = 71 mm = 0.071 m.

The results of the measurement were compared with the analytical solutions of the known and proposed expression. Calculations and graphical depiction of analytical expressions and measurements were presented using Mathcad[11]. The measurement of the radiation pattern was performed in the TU-Varna anechoic chamber, using the Robotron Präzisions microphone MK301, the generator used is Realtek High Definition Audio and Matlab® sound test program [12] at frequency *f = 2000Hz* distance *d = 1m*. by using time scavenging [13] to obtain the following result on fig. 3:

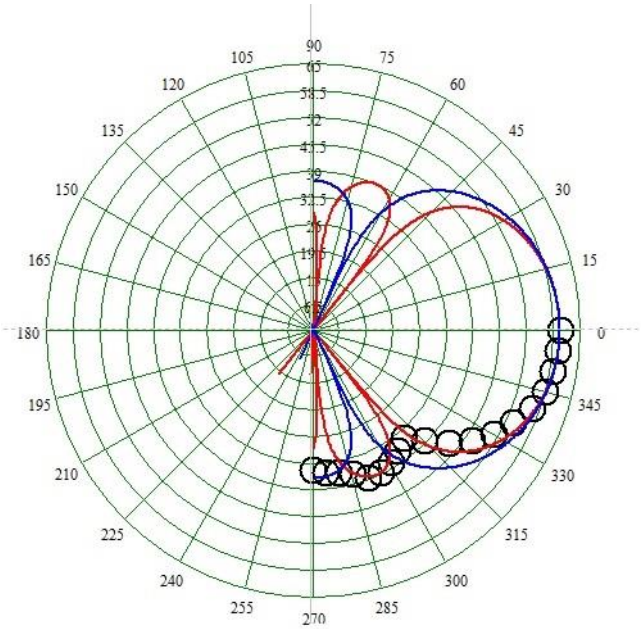


Fig. 3 Experimental result: Measured radiation pattern (black circles), known expression (13)(blue dense line) and proposed expression (12)(red dense line).

According to the result demonstrate in Fig. 3 the applicability and the higher precision of the proposed expression (12) were proved valid for the MidBass - type loudspeaker polar calculations with a pronounced cone shape of the acoustic transducer.

III. Additional Remarks

In addition the experiment was carried out at distances: d = 0.2m, 0.6m with the same frequency, no difference in the

position of the SPL minimum in measured polar diagrams was notified.

IV. CONCLUSION

The object of this paper is the problem of calculating the polar diagram of a circular piston acoustic transducer with conical membrane shape after the analysis were reached the following results:

- an alytical expressions for determining the sound pressure on the axis of radiation expression (11) and in the area in front the acoustic transducer expression (9) are proposed.
- an analytical expression is proposed for determining the radiation characteristic applicable to cone-shaped circular piston acoustic transducers expression(12);
- experimental and analytical comparison of expressions (12) and (13) was carried out by demonstrating the practical applicability of the proposed expression (12) for cone-shape acoustic transducers.
- important conclusions were drawn for the dependence of the polar characteristic on the geometric characteristics of the transducer (diameter and height of the cone) and the frequency *f* which opened a field for further studies.
- there are practically no clear minimum values in the space in front of the loudspeaker along the radiating axis, but in the far field, minimum values that can be measured with the measuring system appear.

ACKNOWLEDGEMENT

The author expresses his deep gratitude to Assoc. Prof. Ekaterinoslav Sirakov for his recommendations and advices.

The research and the results presented in this paper, were carried out under the PD 2/2017 project within the framework of the scientific research activity inherent in TU-Varna, financed by the state budget.

REFERENCES

[1] T. D. Rossing, M. Schroeder, W. Hartmann, N. Fletcher, F. Dunn, D. Campbell, A. Pierce, Handbook of Acoustics, Springer, New York, 2007.

[2] L. E. Kinsler, A. R. Frey, A. B. Coppens, J. V. Sanders, Fundamentals of Acoustics 4th ed., John Wiley&Sons, New York, 2000;

[3] I. Valchev, "Electroacoustics", „Technics", Sofia, 1975 (In Bulgarian);

[4] T. Nimura, E. Matsui, K. Shibayama, and K. Kido, "Study on the cone type dynamic loudspeakers," J. Acoust. Soc. Jpn., vol. 7, no. 2.

[5] T. Ueno, K. Takahashi, K. Ichida, and S. Ishii, "The vibration analysis of a cone loudspeaker by the finite element method," J. Acoust. Soc. Jpn., vol. 34, no. 8, 1978

[6] N. Kyouno, T. Usagawa, T. Yamabuchi, and Y. Kagawa, "Acoustic response analysis of a cone-type loudspeaker by the finite element method," J. Acoust. Soc. Jpn., vol. 61, no. 6, 2005.

[7] L. F. Lependin, Acoustics (in Russian), Moscow, Vysshaya Shkola, 1978;

[8] F. J. Frankfort, "Vibration and Sound Radiation of Loudspeaker Cones," Thesis, Delft, 1975 Networks", TELSIKS'99, Conference Proceedings, Nis, Yugoslavia, 1999.

[9] Ek. Sirakov. *"Electroacoustics"*. Varna, Tachnical University - Varna, 2009.

[10] I. Iliev, „Polar Response of a Circular Piston", TEM Journal, vol. 3, no.3, ISSN: 2217-8333, 2014.

[11] Y. Hu, X. Zhao, T. Yamaguchi, M. Sasajima, Y. Koike, Excitation Experiments of a Cone Loudspeaker and Vibration-Acoustic Analysis Using FEM, World Academy of Science, Engineering and Technology International Journal of Mechanical, Aerospace, Industrial, Mechatronic and Manufacturing Engineering Vol:9, No:12, 2015

[12] G. Markov, Ek. Sirakov, "Generation of measurement signals whit different enveloping form using matlab implementation" (in Bulgarian), Conference "Acoustics 2016", vol. 18, No. 15, ISBN: 1312-4897, Sofia, 2017.

[13] Klippel W., Schlechter J. Measurement and Visualization of Loudspeaker Cone Vibration. Presented at the 121st Convention of the Audio Eng. Soc., New York, USA, 2006;

Delay Locked Loop Clock Skew Compensator for Differential Interface Circuit

Goran Jovanović, Mile Stojčev, Tatjana Nikolić and Goran Nikolić

Abstract – **Skew is a big problem for sending parallel data and its clock across cables or printed circuit board (PCB) traces. The problem is that the phase relation of the data and clock can be lost due to different travel times through the link. The problem can be efficiently solved when source-synchronous differential point-to-point parallel link interface architecture is used. In this paper an efficient clock de-skew structure based on Delay Locked Loop (DLL) is described. It is implemented in IHP 0.13 μm BiCMOS technology and is characterized with: worst case locking time 40 ns (20 cycles @ 500 MHz), wide lock frequency range from 470 MHz up to 870 MHz, and static phase error of 7 ps.**

Keywords – **Differential interface, Clock skew, Transmission line, DLL.**

I. Introduction

With the rapid advances in modern VLSI IC technology synchronous data transfer between PC boards, at rates of order up to several gigahertzes, is required. The need for high input-output bandwidth has led to widespread use of differential signaling. In essence, a differential pair is realized with two transmission lines that have equal and opposite polarity signals propagating on them, so that the positive path and the negative path (of a differential pair) are tightly timed [1]. However, in practical realizations there are differences between two signal paths. The differences are primarily imposed by unequal signal path length routing, local variation of the epoxy laminate dielectric constant, unmatched twists or kinks in the link connections, etc. As a consequence, all these differences can cause positive and negative paths to deliver their signals at different moments at the end of a differential channel, i.e. to provide signal skew, also called within-pair skew or intra-pair skew, what at different locations may degrade system performance, and even cause system malfunction. To remedy for within-pair skew in PCB traces, system design engineers use dielectric materials construction with a mixture of fiber and non-fiber, while for de-skew compensation (correction) in cables they choose customized cables and connectors. However, the system cost of those solutions is very high [2-4].

In this paper we propose an electronic within-pair de-skew solution intended for automatic align of two signals in a differential pair at the receiver. This electronic is part of the source-synchronous point-to-point parallel link interface intended for high-speed differential data transfer over PCB. It is DLL based circuit which samples the delay differences from both the positive and negative edges of the reference clock signal and is characterized with fast locking time, wide bandwidth, and small static phase error.

Goran Jovanović, Mile Stojčev, Tatjana Nikolić and Goran Nikolić, are with University of Niš, Faculty of Electronic Engineering, Niš, Serbia, E-mail: goran.jovanovic@elfak.ni.ac.rs.

II. Problem Definition

During the last three decades, with the incredible rapid advance of MPU's operating frequency, memory and input-output subsystems are required to keep-up by increasing clock-distribution and data-transfer speed. The need for high memory/input-output (I/O) bandwidth has led to the widespread use of point-to-point parallel links. Conventional parallel links are generally source-synchronous with a clock sent along with data signals for receiver timing recovery, i.e. high speed system interfaces usually transmit a high speed clock synchronously with a parallel data stream [4].

A. Transmission line delay, theory:

We will explain now, in short, the used theoretical aspect for skew estimation (timing difference between signals) in terms of signal path difference and dielectric constant [5]. Phase velocity of signals on transmission line, v_p, have a form:

$$v_p = \frac{1}{\sqrt{\mu \cdot \varepsilon}} = \frac{c}{\sqrt{\mu_r \cdot \varepsilon_r}} \cong \frac{c}{\sqrt{\varepsilon_r}} \tag{1}$$

where c is the speed of light, μ_r relative magnetic permeability, ε_r relative dielectric permittivity. On standard PCB, based on FR4, as most commonly used composite material, μ_r is 1 and ε_r is in the range from 4 up to 4.4. When the signal path on the PCB changes for

$$\Delta l = l_1 - l_2 \tag{2}$$

at transfer speed

$$v_p = \frac{\Delta l}{t_{delay}} \tag{3}$$

produces a delay that can be estimated based on the effective dielectric constant (for the FR4-based PCB typically $\varepsilon_{r\,eff} = 3.4$). The delay calculated as

$$t_{delay} = \frac{\Delta l}{v_p} = \frac{\Delta l}{c} \cdot \sqrt{\varepsilon_{r\,eff}} = \frac{1\text{m}}{3\,10^8 \frac{\text{m}}{\text{s}}} \cdot \sqrt{3.4} \tag{4}$$

and the resulting value is t_{delay} = 6.15 ns/m, i.e. t_{delay} = 61.5 ps/cm. This implies that Δl = 1 cm can cause serious problem in clock synchronization.

B. Source-synchronous point-to-point parallel link interface

Typical source-synchronous unidirectional and differential point-to-point parallel link interface architecture [3], is presented in Fig. 1.

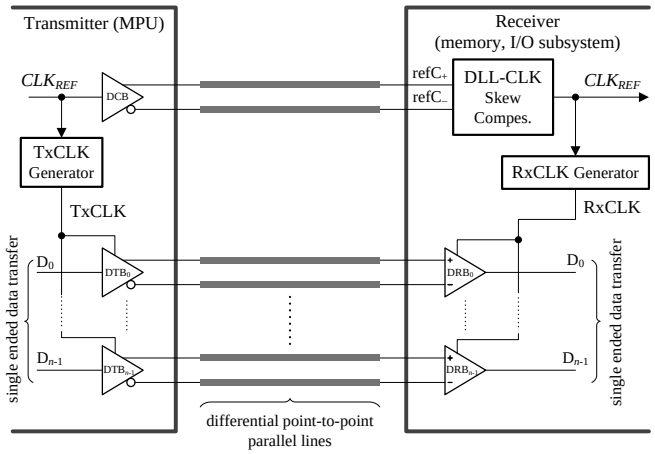


Fig. 1 Source synchronous simultaneous unidirectional and differential point-to-point parallel link interface architecture

Notice: CLK_{REF} stands for referent clock signal; TxCLK (RxCLK) *–global transmitter (receiver) clock signal;* TxCLK_Gen (RxCLK_Gen) *– transmitter (receiver) clock generator;* DCB - *differential clock buffer;* $DTB_0,\ldots,$ DTB_{n-1} ($DRB_0,\ldots,DRB_{n-1}$) - *differential transmitter (receiver) data buffer; DLL-CLK-Skew_Compes. – delay locked loop skew compensator;* $D_0, \ldots, D_{n-1}$ *– data signals*

All data signals ($D_0,\ldots, D_{n-1}$) and a referent clock signal CLK_{REF} are transmitted synchronously. Data rate of signals $D_0,\ldots, D_{n-1}$ is determined by TxCLK (RxCLK). At the receiver, a delay locked loop skew compensator (DLL-CLK-Skew_Compes) generates referent clock signal CLK_{REF}, while the receiver clock generator RxCLK_Gen generates a global receiver clock RxCLK. The RxCLK is used to sample all incoming data signals $D_0,\ldots, D_{n-1}$. Correct sampling is achieved when TxCLK = RxCLK.

Here in focus of our interest is the design of DLL-CLK-Skew_Compes as constituent of the system presented in Fig. 1. Design of other building blocks (sketched in Fig. 1) is currently under further investigation and is directed towards development of bidirectional parallel link interface.

III. Delay Locked Loop Clock Skew Compensator Architecture

The timing difference between signals is called skew. It mainly arises from signal trace incongruity such as trace-length difference in PCB or on the memory and I/O modules, unequal parasitic elements of the packages, etc. Data bus skew have critical impact on the whole performance of the memory and I/O subsystems at over 100 MHz clock period [1-4]. To solve this problem in an efficient manner, clock de-skew electronics is used. In general, Phase Locked Loops (PLLs) and Delay Locked Loops (DLLs) are broadly used in high-speed digital systems, clock synchronization and data recovery systems. Fig. 2 shows the architecture of the DLL-CLK Skew Compensator. As can be seen in Fig. 2 this system building block consists of a phase-frequency detector (PD), a loop filter (LF), and a voltage controlled delay line (VCDL), and inverter. The PD is an important component in designing DLL based structure [6]. It detects the phase error information between input reference signal refC+ and generated signal refC- generated output by VCDL. Phase error information is generated in the form of UP and DOWN signals. The produced signal by PD is sent to CP. To adjust the delay of VCDL, CP and LF (integrator) generate appropriate value for control voltage of VCDL (marked as V_{ctrl}) based on the phase difference of *refC+ and *refC- signals.

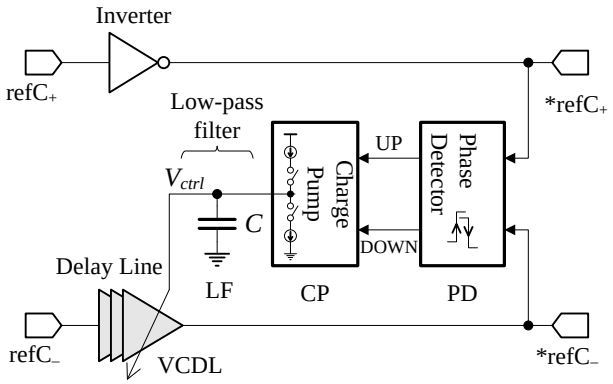
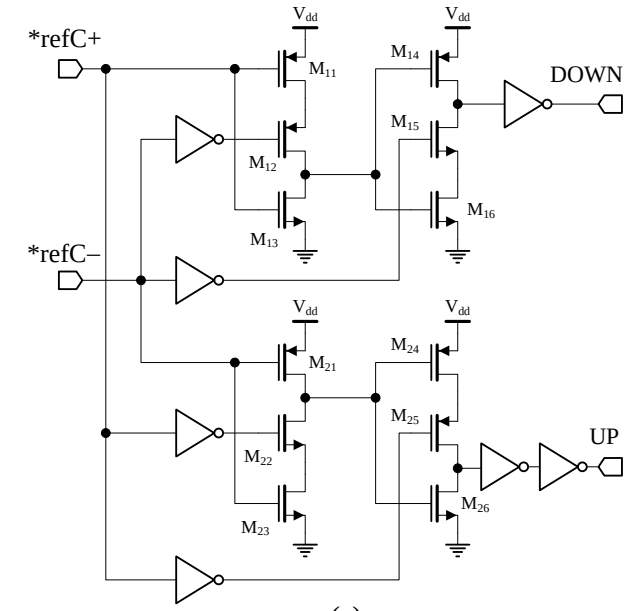
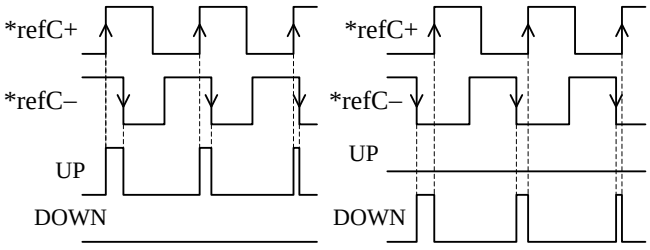


Fig. 2. Architectural structure of the delay locked loop clock skew compensator



(a)



(b)

Fig. 3. PD transistor schematic (a), waveforms of PD (b)

C. The proposed phase detector architecture

The PD sketched in Fig. 3 represents crucial component of the DLL-CLK Skew Comp building block. The transistor schematic of a phase detector sensitive both to positive and negative clock edge transitions is presented in Fig. 3(a). The PD consists of two blocks intended to handle the input signals *refC+ and *refC–. Each block is composed of two stages, *p*-precharge (*n*-precharge) and *n*-precharge (*p*-precharge), connected in cascade, followed by an inverter (buffer). The PD operation principle is depicted in Fig. 3(b). As can be seen from Fig. 3, the widths of UP and DOWN signals are proportional to the phase difference between the *refC+ and *refC– signals. Waveforms in the left side of Fig. 3(a) represent the case *refC+ leading *refC–, while those given in right side of Fig. 3(b) to the case *refC– leading *refC+.

D. Charge Pump

In the DLL structure given in Fig. 4, the phase error between the input reference clock and the VCDL output clock is sensed by the PD and transferred to the CP in the form of voltage pulses. The CP performs the function of adjusting the voltage of the LF and thereby altering the VCDL delay according to the phase error information from the PD. In principle, the CP simply consists of two controlled switches, one current source (transistor M3), and one current sink (transistor M2), as shown in Fig. 4. Transistors M1 and M4 are used as switching elements. The LF is realized with the capacitor *C* which acts as an integrator.

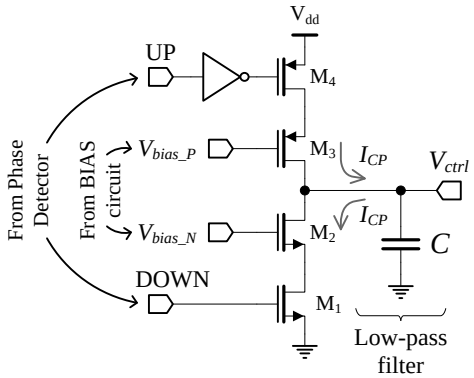


Fig. 4. A simplified charge pump schematic

E. Voltage controlled delay line

The VCDL is also one of the crucial DLL's building block. The output signal of the DLL is directly taken from the VCDL. The proposed VCDL consists of six delay stages (cells) which are connected in series. The total delay of the VCDL is equal to one clock period CLK_{REF} (or a phase shift of 360°) in the locked state. Theoretically, all the delay stages in the VCDL are identical, and each delay stage contributes a time delay of CLK_{REF} / 6 for six-stage VCDL. Let note that, the number of delay stages is adjusted in accordance with the operating frequency. Using more stages increases the phase resolution, but also increases the minimum VCDL delay. The structure of the VCDL delay stage is presented in Fig. 5. At the left part of Fig. 5, the bias circuit of VCDL is presented, and on the right side circuit structure of a single delay stage is given [7].

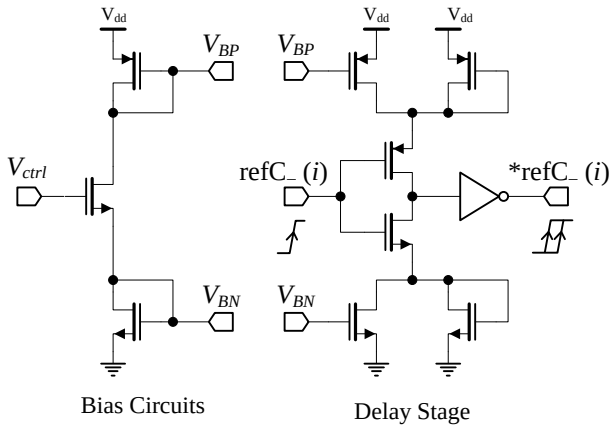


Fig. 5. VCDL delay stage

Operation of the six stage delay line is simulated and the obtained result which corresponds to the total time dependency in terms of control voltage is sketched in Fig. 6.

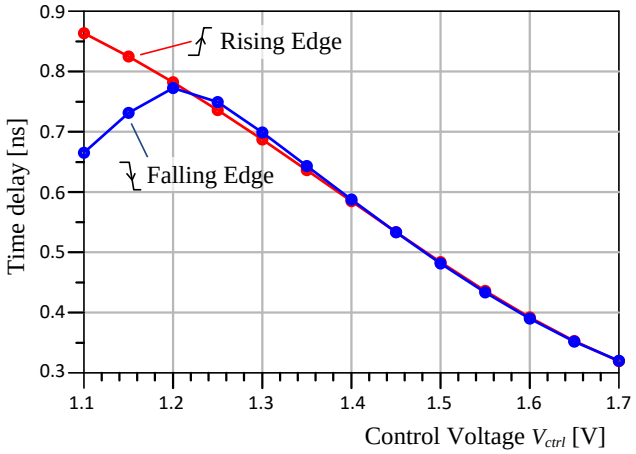


Fig. 6. Time delay variation in terms of control voltage

F. The PD & CP characteristic

There are several important characteristics concerning PD & CP, but one having the largest impact is the output phase characteristic, i.e., the PD & CP output current vs. phase error [6]. Fig. 7 illustrates the phase characteristic for the proposed PD & CP architecture. As can be seen from Fig. 7 the phase characteristic is dominated by two issues, dead-zone and blind-zone. In our design solution the dead-zone is minimized thanks to the fact that the proposed PD architecture does not utilize intermediate signals for reset operation, as is the case in conventional PDs, but rather generates UP and DOWN signals directly. With the proposed circuit topology, the PD architecture achieves a small blind-zone close to the limit imposed by dynamic characteristics of MOS transistors and parasitic capacitances of internal nodes.

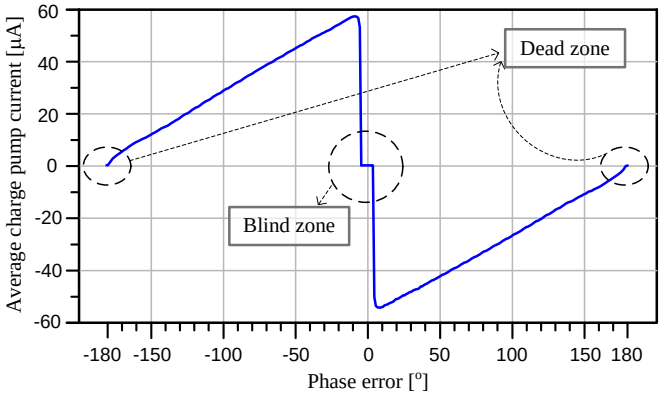


Fig. 7. Phase characteristics of PD

The maximal operating frequency of PD was determined by measuring rising-, falling-, and hold-time of the UP (DOWN) signal obtained at the output of the circuit. Rising- (falling-) time deals with time period during which the pulse amplitude variation are within a range from 10% up to 90% (from 90% down to 10%). Hold time correspond to the needed time for charging (discharging) parasitic precharge capacitors. During this, the worst case technology corner was selected for performance evaluation. By summing the obtained parameters, the maximum operating frequency was determined. In our case it was 8 GHz.

IV. Simulation Results

In Fig. 8 the transient response of the DLL-CLK Skew Compes (from Fig. 2) constituent is given. As can be seen from Fig. 8. after powering-up the DLL-CLK Skew Compes enters into locking state after 40 ns what correspond to 28 clock cycles at operating frequency of 660 MHz. In the locked state, the voltage (charge) of the loop filter is kept constant. It is possible that equal charging and discharging will still happen in the locked state. In fact, it is desired to have such activities to minimize jitter. However, the charging and discharging currents must be identical as well as very narrow so that the voltage of the loop filter will not be disturbed.

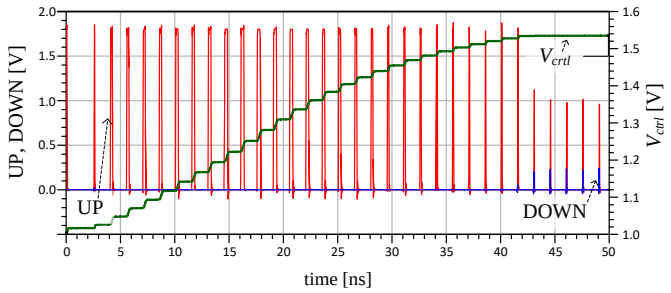


Fig. 8. Transient response of DLL-CLK Skew Compes

In Fig. 9, waveforms at the inputs refC+ and refC– and outputs *refC+ and *refC– of DLL-CLK Skew Compes building block are presented. As can be seen in Fig. 9, in locking state, the estimated static phase error is very small (~7 ps).

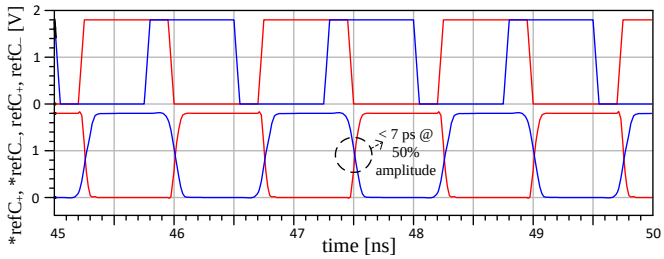


Fig. 9. Waveforms at inputs and outputs of DLL-CLK Skew Compes

The presented results are obtained by using Advance Design System software tool with IHP design kit for 0.13 μm BiCMOS technology [8].

V. Conslusion

Clock skew represents crucial problem in designing source-synchronous unidirectional and differential point-to-point parallel link interface architecture. In this paper an efficient Delay Locked Loop clock de-skew compensation based architecture is described. The proposed circuit is implemented in IHP 0.13 μm BiCMOS technology [8]. It has fast locking time 40 ns (20 cycles @ 500 MHz), wide lock frequency range from 470 MHz up to 870 MHz, and static phase error of 7 ps.

Acknowledgement

This work was supported by the Ministry of Education and Science of the Republic of Serbia, as a part of the project “Low-Power Reconfigurable Fault Tolerant Platforms”, TR 32009.

References

[1] Lee W. Ritchey, A Treatment of Differential Signaling and its Design Requirements, Speeding Edge, Glen Ellen, CA. USA, 2008.

[2] Yuxiang Zheng, Jin Liu, A 5 Gb/s Automatic Within-Pair Skew Compensator for Differential Data in 0.13 mm CMOS, IEEE Transactions on Circuits and Systems—I, Vol. 58, No. 6, pp. 1191-1202, 2011.

[3] Kyungho Ryu, Dong-Hoon Jung, Seong-Ook Jung, A DLL with Dual Edge Triggered Phase Detector for Fast Lock and Low Jitter Clock Generator, IEEE Transactions on Circuits and Systems—I, vol. 59, No. 9, pp. 1860-1870, 2012

[4] Evelina Yeung, Mark Horowitz, A 2.4 Gb/s/pin Simultaneous Bidirectional Parallel Link with Per-Pin Skew Compensation, IEEE Journal of Solid-State Circuits, vol. 35, No. 11, pp. 1619-1628, 2000.

[5] Reinhold Ludwig, Gene Bogdanov, RF Circuit Design: Theory and Applications, sec. ed., Prentice Hall, New Jersey, 2009.

[6] Goran Nikolić, Goran Jovanović, Mile Stojčev and Tatjana Nikolić, Precharged Phase Detector with Zero Dead-Zone and Minimal Blind-Zone, Journal of Circuits, Systems, and Computers, vol. 26, No. 11, 1750179 (16 pages), 2017.

[7] G. S. Jovanović, M. K. Stojčev, Current starved delay element with symmetric load, International Journal of Electronics, pp. 167-175, vol. 93, No 3, 2006.

[8] IHP-Microelectronics, SiGe:C BiCMOS Technologies for MPW & Prototyping, https://www.ihp-microelectronics.com/en/services/mpw-prototyping/sigec-bicmos-technologies.html.

Medical Images Watermarking using Wavelet Transform and DCT

Rumen P. Mironov[1], Stoyan Kushlev[2]

***Abstract* – An algorithm for digital watermarking of medical images using Wavelet transform and DCT is presented. The developed algorithm ensures high transparency of the watermark and is resistant to various types of malicious attacks. The obtained experimental results for some attacks over the test medical images are made on the base of mean-squared error and signal to noise ratio of the reconstructed images.**

***Keywords* – Medical Image Watermarking, Wavelet Transform, DCT, Unitary Transforms.**

I. INTRODUCTION

Recent technological advances in Computer Science and Telecommunications introduced a radical change in the modern health care sector, including: medical imaging facilities, Picture Archiving and Communications System (PACS), Hospital Information Systems (HIS), information management systems in hospitals which forms the information technology infrastructure for a hospital based on the DICOM (Digital Imaging and Communication in Medicine) standard. These services are introducing new practices for the doctors as well as for the patients by enabling remote access, transmission, and interpretation of the medical images for diagnosis purposes [1], [2], [3].

Digital watermarking has various attractive properties to complement the existing security measures that can offer better protection for various multimedia applications [4]. The applicability of digital watermarking in medical imaging is studied in [5] and a further justification of the watermarking considering the security requirements in teleradiology is discussed in [2].

The new medical information systems required medical images to be protected from unauthorized modification, destruction or quality degradation of visual information. The other problem is a copyright protection of disseminated medical information over Internet. In this regard three main objectives of watermarking in the medical image applications: data hiding, integrity control, and authenticity are outlined in [5], which can provide the required security of medical images.

[1]Rumen P. Mironov is with the Faculty of Telecommunications, Technical University of Sofia, Boul. Kl. Ohridsky 8, Sofia 1000, Bulgaria. E-mail: rmironov@tu-sofia.bg

[2]Stoyan Kushlev is with the Faculty of Telecommunications, Technical University of Sofia, Boul. Kl. Ohridsky, 8, Sofia 1000, Bulgaria. E-mail: skushlev@mail.bg

Every system for watermarking can be characterized with invisibility of the watermark, security of the watermark, robustness of the watermark and the ability for reversible watermarking. The importance of each depends on the application and how it is used [6], [7]. For the needs of medicine the main watermarking characteristics are:

- Invisibility of the watermark – the embedded watermark should be invisible without reducing the quality of the original images;
- Security of the watermark – secrecy to unauthorized persons of the information for the embedded watermark;
- Robustness and fragility of the watermark – robust watermarking is resistant to possible attacks such as image processing and on the other hand fragile watermarks will allow high detection of unauthorized access or attacks on the watermark;
- Reversibility of the watermark – removing of the embedded watermark should not reduce the quality of the original images.

Based on processing domain, watermark techniques can be separated as watermarking in spatial domain, watermarking in frequency domain and watermarking in phase domain of the input signal. According to the way of watermark preprocessing, discern two groups of methods: the first one is when the watermark is transformed in the domain of the input image and the second one is when the watermark is not transformed in the domain of the input image. Another classification is based upon the transparency of the watermark into the input images - the watermark is transparent or non- transparent.

Watermarking in spatial domain allow easy realization of the algorithms for watermarking. The disadvantage of using the spatial domain is that the watermarks have low efficiency and robustness. Using frequency and phase domain allow watermarks with high transparency and robustness. On the other hand using transformations on the watermarks themselves assures high security agents against unauthorized attacks.

The best way to test the watermark robustness is by simulating of unauthorized attacks. Unauthorized attacks are attacks against the integrity of the watermark. The most used attacks are unauthorized removal, adding or detection of watermark. The removal and adding of watermarks are active attacks while the detections of watermarks are passive attacks.

An outline of the medical image watermarking field that uses various techniques to embed watermark data and utilize various functions to detect tampered regions is given below in the paper [8].

In the present work an algorithm for digital watermarking of medical images using Wavelet transform and DCT is described. The developed algorithm ensures high transparency of the

watermark and is resistant to various types of malicious attacks. The obtained experimental results for some simulated attacks over the three test medical images are made on the base of mean-squared error and signal to noise ratio of the reconstructed images. The robustness of the watermark against some attacks are tested with the post processing of watermarked images by adding of Salt and Pepper noise, Gaussian noise, filtration whit median filters and average filters.

II. Mathematical Description

The common block scheme of the developed algorithm for watermarking is shown on Fig.1.

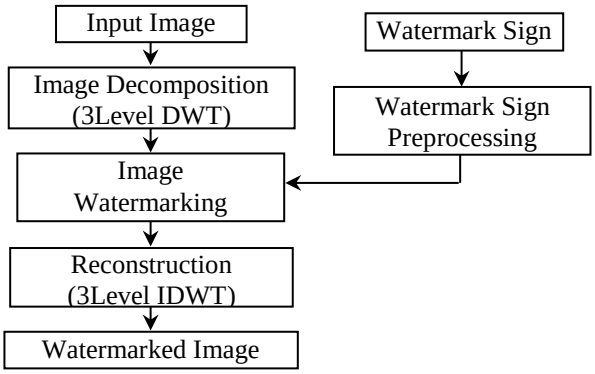


Fig.1. Common block scheme of watermarking.

The input image is decomposed into 3 levels via two dimensional Discrete Wavelet Transform (2D DWT). The transformed by the 2D discrete cosine transform (DCT) image of the digital watermark is included in one of the three 2D DCT transformed blocks from the 3th level of the 2D DWT – LH3, HL3 or HH3. The choice of the watermark insertion block with size PxQ is based on the maximum of entropy. Watermarking is performed by the following formula:

$$C_{k'} = C_k + a \times V_k, k = 1, 2, \dots, P \times Q \quad , \qquad (1)$$

where: C_k are the coefficients of the transformed insertion block, V_k are the coefficients of the transformed watermark, k is the consecutive number of the coefficients and a is parameter which determines the depth of watermarking. The value of the coefficient a must satisfy the markup threshold. The threshold is dynamically determined in relation to the input image, which is 10 percent lower than the entropy of the selected block. In reconstruction block the inverse wavelet transform (2D IDWT) is applied and the output watermarked image is obtained.

The developed decoder is informed. The transformed images - marked and original are including at the input. The purpose of the decoder is to determine what message (sign) is included in the watermarked image. The information about the block where the watermark is written, the markup threshold, and the markup factor is recorded in the header of the format used. The watermark is retrieved by the following formula:

$$V_{k'} = (C_{k'} - C_k)/a, k = 1, 2, \dots, P \times Q \quad , \qquad (2)$$

where: $C_{k'}$ are the coefficients of the transformed insertion block of the watermarked image, C_k are the coefficients of the transformed correspondent block of the original image, $V_{k'}$ are the coefficients of the transformed watermark. The watermark is received in the reconstruction unit. Over the data received by the decoder applies inverse discreet wavelet transform.

III. Experimental Results

For the analyses of efficiency of the developed algorithm for watermarking of medical images three test images, shown in Fig.1a, b, c, with size 512x512 and 256 gray levels are used.

Fig.1a. Input X-ray test image “Spine 1”.

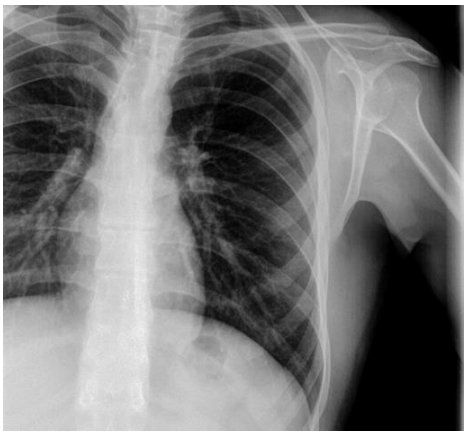


Fig.1b. Input X-ray test image “Spine 2”.

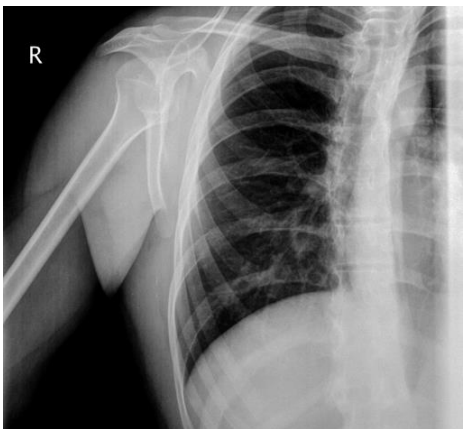


Fig.1c. Input X-ray test image “Spine 3”.

Based on the described sequence in Section II, algorithm for embedding of watermark and algorithm for extraction of the watermark have been developed and have been simulated by the developed with MATLAB programs.

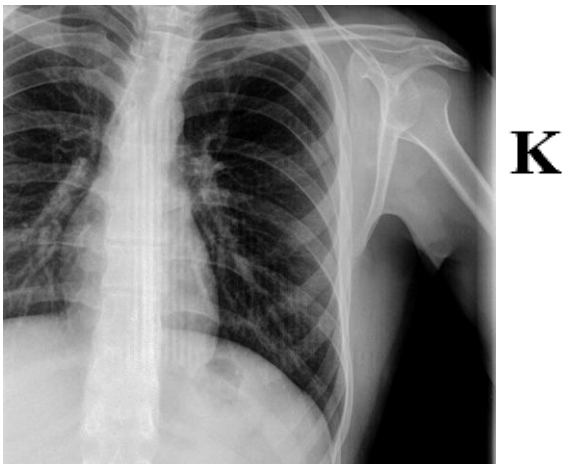
The test images shown in Fig.1 are transformed by the 3 levels 2D DWT and the input watermark (letter K) is embedded into the transformed block (LH3, HL3 or HH3) with maximum entropy of each image.

The robustness of the watermark against some popular attacks are simulated with the post processing of watermarked images by adding 100% of Gaussian noise with mean 0 and variance 0.01; adding 100% of Salt and Pepper noise; filtration with median filter with size 3x3; filtration of Gaussian noisy image with average filter; filtration of Salt and Pepper noisy image with median filter. The obtained results for the test images are summarized in Tabl.1.

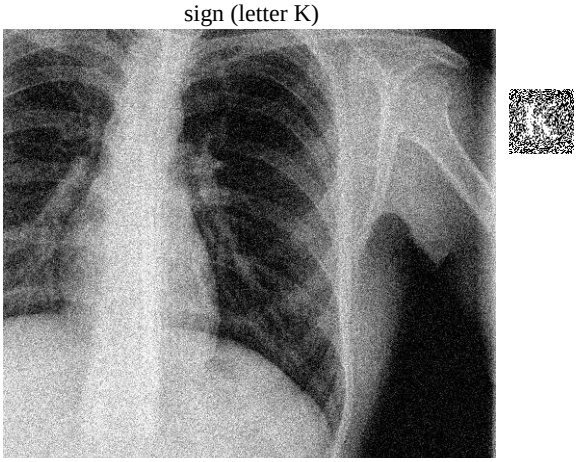
Tabl.1

Test Images	"Spine 1"	"Spine 2"	"Spine 3"
Reconstructed Watermarked image			
SNR, dB	19.86	18.16	18.71
PSNR, dB	45.8	42.72	43.49
MSE	1.71	3.48	2.91
NMSE	0.01	0.015	0.01
NMSE, %	1	1.5	1
NC	0.57	0.53	0.55
NC, %	57	53	55
Watermarked image with Gaussian noise			
SNR, dB	4.66	4.67	4.69
PSNR, dB	30.6	29.22	29.47
MSE	56.67	77.74	73.55
NMSE	0.34	0.34	0.34
NMSE, %	34	34	34
NC	0.34	0.46	0.42
NC, %	34	46	42
Watermarked image with Salt and Pepper noise			
SNR, dB	14.54	14.02	14.27
PSNR, dB	40.48	38.57	39.05
MSE	5.82	9.04	8.1
NMSE	0.035	0.04	0.038
NMSE, %	3.50	4	4
NC	0.36	0.41	0.39
NC, %	36	41	39
Watermarked image with median filtration			
SNR, dB	16.13	17.4	17.99
PSNR, dB	45.07	41.95	42.77
MSE	2.02	4.15	3.44
NMSE	0.012	0.018	0.016
NMSE, %	1.20%	1.80%	1.60%
NC	0.57	0.52	0.55
NC, %	57%	52%	55%
Watermarked image with Salt and Pepper noise and median filtration			
SNR, dB	18.55	16.99	17.6
PSNR, dB	44.49	41.54	42.38
MSE	2.31	4.57	3.76
NMSE	0.01	0.02	0.017
NMSE, %	1%	2%	1.70%
NC	0.56	0.52	0.55
NC, %	56%	52%	55%
Watermarked image with Gaussian noise and average filtration			
SNR, dB	8.86	8.6	8.6
PSNR, dB	34.8	33.15	33.38
MSE	21.51	31.49	29.88
NMSE	0.13	0.14	0.14
NMSE, %	13%	14%	14%
NC	0.34	0.43	0.41
NC, %	34%	43%	41%

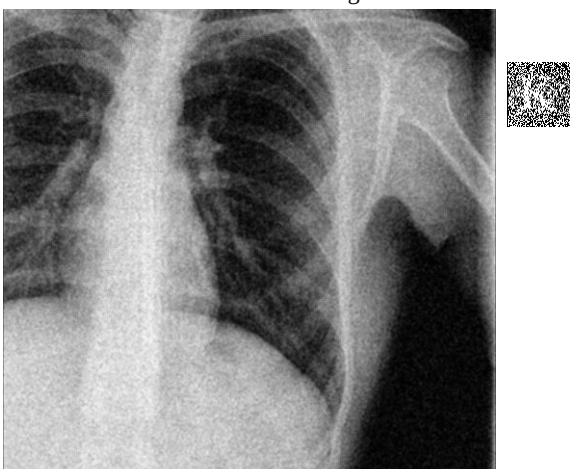
To estimating the efficiency of the developed algorithms for watermarking of medical images the following metrics are used: peek signal to noise ratio (PSNR) estimate how transparent is the watermark to the human eyes; normalize cross-correlation (NC) is used to determinate how close the extracted watermark is compared to the original. High value of NC means that there are little differences between them; mean square error (MSE) and normalized mean square error (NMSE) are used to determinate how much the watermark image has change compared to the original.



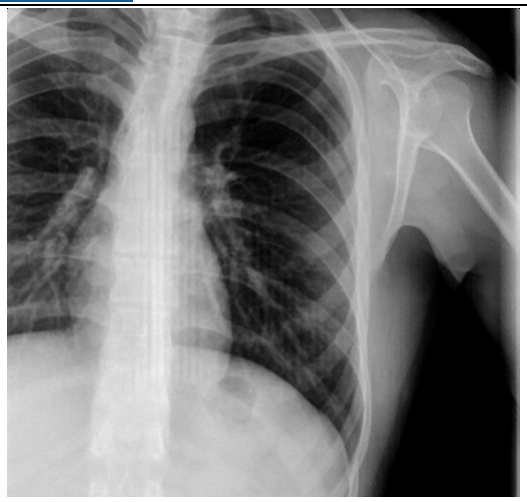
a. Input watermarked image and original watermark sign (letter K)



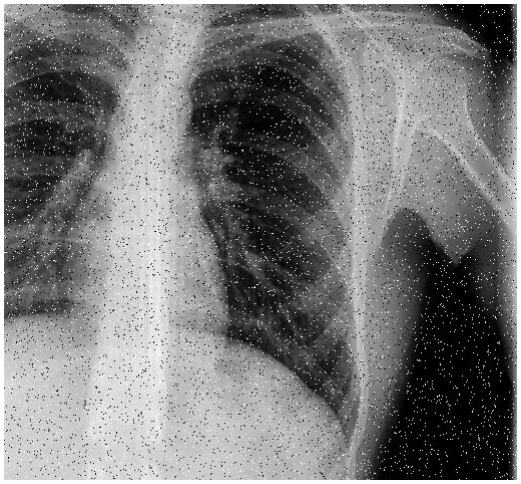
b. Input watermarked image with Gaussian noise and extracted watermark sign



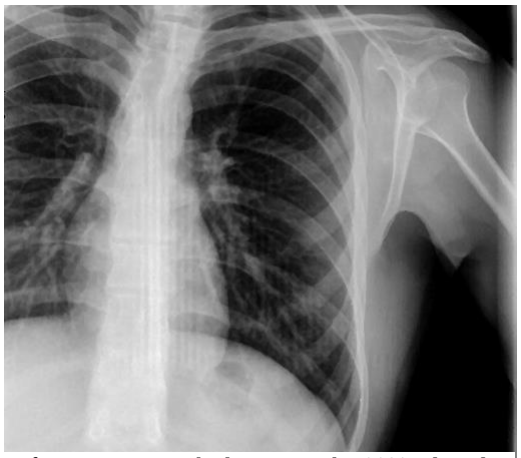
c. Input watermarked image with Gaussian noise and average filter and extracted watermark sign



d. Input watermarked image with median filter and extracted watermark sign



e. Input watermarked image with 100% salt and pepper noise and extracted watermark sign



f. Input watermarked image with 100% salt and pepper noise and median filter and extracted watermark sign

Fig.2. Results for watermarked image "Spine 2" with different post processing attacks

In Fig.2 a-f the visual results for watermarked image "Spine 2" with different post processing attacks are shown. On the right corner of each image is shown the extracted watermark.

IV. Conclusion

An algorithm for digital watermarking of medical images using Wavelet transform and DCT is presented. The obtained experimental results for some attacks over the three test medical images are made on the base of mean-squared error, signal to noise ratio and normalized cross-correlation of the reconstructed images. They shows that the developed algorithm for watermarking allows high robustness to possible attacks based on image processing operations as transforms, filtrations and etc. On the other hand the embedded watermark is practically invisible for the doctors and retains largely the information in the original images. This will allow to a great extent to verify the reliability of the medical data transmitted and recorded as images.

All this leads to the conclusion that the developed algorithm can be used successfully for watermarking not only of medical but on other type of data presented as images.

V. Acknowledgement

The author thank 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] P. Koushik, G. Ghosh, M. Bhattacharya. "Biomedical Image Watermarking in Wavelet Domain for Data Integrity Using Bit Majority Algorithm and Multiple Copies of Hidden Information". *American Journal of Biomedical Engineering*, 2012, vol. 2(2), pp. 29-37.

[2] H. Nyeem, W. Boles, C. Boyd. "A Review of Medical Image Watermarking Requirements for Teleradiology". *Journal Digital Imaging*, 2013, vol. 26, pp.326–343.

[3] L. Siau-Chuin, J.M. Zain. "Reversible medical image watermarking for tamper detection and recovery", *3rd IEEE International Conference on Computer Science and Information Technology*, 2010, vol. 5, pp. 417 – 420.

[4] I. Cox, M. Miller, J. Bloom, J. Fridrich, T. Kalker. *Digital watermarking and steganography*. 2nd Edition, Elsevier, Burlington, 2007.

[5] Coatrieux, H. Maitre, B. Sankur, Y. Rolland, R. Collorec. "Relevance of watermarking in medical imaging". *Proceeding of IEEE EMBS International Conference on Information Technology Applications in Biomedicine*, 2000, pp. 250–255.

[6] W. K. Pratt. *Digital Image Processing*, 4th Ed., John Wiley & Sons. Inc., Hoboken, New Jersey, 2007.

[7] R. C. Gonzalez, R. E. Woods. *Digital Image Processing*, Third Ed., Pearson Education Inc., 2008.

[8] A. Ustubioglu, G. Ulutas. "A New Medical Image Watermarking Technique with Finer Tamper Localization". *Journal of Digital Imaging*. Springer International Publishing, 2017, pp.1-17.

[9] Mahasweta, J. Joshi et al. "Watermarking in DCT-DWT Domain", *International Journal of Computer Science and Information Technologies* (IJCSIT), Vol. 2 (2), 2011, pp. 717-720.

[10] Neha Singh, Mamta Jain, Sunil Sharma, "A Survey of Digital Watermarking Techniques", *International Journal of Modern Communication Technologies & Research*, August 2013

Extension of Error Correction Capability of Two Dimensional Single Error Correction - Double Error Detection Encoding Technique

T. R. Nikolić[1], G. S. Nikolić[1], M. K. Stojčev[1], G. S. Jovanović[1], B. D. Petrović[1]

Abstract – **Wireless Sensor Networks (WSNs) are energy constraint networks that require reliable data transfer at a low cost of energy. There are two basic approaches to recover erroneous packets, ARQ and FEC. FEC is used with aim to increase link reliability and reduce number of packet retransmission. Given that most bit errors during data transmission are single-bit or double-bit errors and multiple-bit errors are present but rare, extension of the conventional error correction capability of 2D-SEC-DED code has been involved and performance evaluation are derived.**

Keywords – **Wireless Sensor Networks, Forward Error Correcting Code, 2D SEC-DED Coding Scheme**

I. Introduction

Wireless Sensor Networks (WSNs) are composed of large number of spatially distributed, autonomous and battery-powered small devices, called Sensor Nodes (SNs). SNs have the ability to sense the physical environment, process the obtained information and communicate using the radio interfaces [1]. An SN in a WSN is typically equipped with transducer, radio transceiver, microcontroller unit, and power source (usually battery). The lifetime of a WSN depends on how efficiently the battery life of each SN is in use. The radio transceiver is very expensive in terms of energy consumption, while the other components in the SN data processing consume significantly less.

Errors in data transmitted between SNs appear as a consequence of noise, interference, signal fading, etc. To provide reliable data communication two techniques are used, one is Automatic Repeat reQuest (ARQ) and the other one is Forward Error Correction (FEC). In ARQ the sender node adds error detecting code, usually Cyclic Redundancy Checks (CRC) or parity check codes to the data. In FEC the source node encodes data using some error correcting code which add redundancy to the packet and lets the receiver to correct errors in data packet if present, thus making retransmission outdated, since the communication activity is the most power hungry when compared to sensing and data processing.

This paper is focused on block codes. Among the widely used block codes in WSNs are Hamming, single error correction - double error detection (SEC-DED), CRC, erasure codes. Hamming codes have very simple structure and high code rate, and therefore they are used in FEC schemes when errors are random and the error rate is low.

In this paper we propose extension of error correction capability of two dimensional SEC-DED encoding technique. The extension deals with software manipulations of the received packet. In this manner, all single- and double-errors within the data field of the received packet are corrected.

[1]T. R. Nikolić, G. S. Nikolić, M. K. Stojčev, G. S. Jovanović, and B. D. Petrović, are with University of Nis, Faculty of Electronic Engineering, Niš, Serbia, E-mail: tatjana.nikolic@elfak.ni.ac.rs

II. SEC-DED Codes

SEC-DED codes are constructed by extended Hamming codes with an extra parity bit. The Hamming distance of SEC-DED is 4, which allows the decoder to distinguish between single-bit and two-bit errors. Thus the decoder can detect and correct a single-error and at the same time detect (not correct) a double error. Our observations, which are in accordance with [2], show that most bit errors are single- or double-bit errors, while burst and multiple bit errors rarely occur. Thus, it is likely that an encoding scheme that corrects single and double-bit errors can reduce a significant portion of the errors.

During one complete period of the adopted protocol implemented in our WSN, data gathered from several sensor elements are stored into the microcontroller memory as a matrix of order $k = k_r \times k_c$ bits (Information Bits Field, IBF), where k_r corresponds to the number of sensor elements in SN (sensed data) and k_c to the number of bits per element (ADC resolution of sensed data), see Fig. 1.

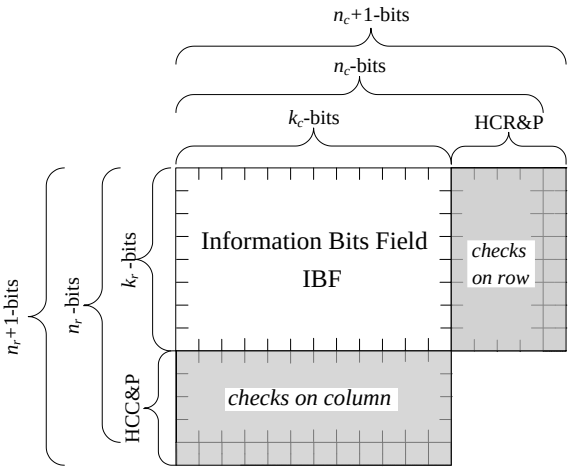


Fig. 1. 2D-SEC-DED code

2D SEC-DED code is two-dimensional code. We parameterize this code by the following expression

$$2D\ SEC - DED^{d_{min}}_{k_{ec}+h_{op},k_{er}+v_{op}}[k_r, k_c] \quad (1)$$

where d_{min} is the minimal Hamming distance, $k_{ec}(k_{er})$ is the number of parity check bits per row (column), $h_{op}(v_{op})$ is the number of overall parity check bits per horizontal (vertical) direction. For more details, related to creating 2D SEC-DED code, the reader can consult Reference [3].

In our solution we use synchronous rendezvous protocol called M-RPLL, which uses identical packet formatting and control mechanisms already described in [4-5], but implements different data encoding (2D-SEC-DED instead of CRC). The format of the control message (RTS, CTS, and ACK/NACK) for M-RPLL is presented in Fig. 2. If we put in expression (1) $k_r = 1$, $k_c = 11$, $k_{er} = 0$, $k_{ec} = 4$, $h_{op} = 1$, $v_{op} = 0$, and $d_{min} = 4$, the 2D SEC-DED code is transformed into one dimensional (16, 11) SEC-DED perfect code, see Fig. 1. Such encoding scheme employs single error correcting and double error detecting (SEC-DED) for the Data field. The format of the data message DATA is presented in Fig. 3. If we put in expression (1) $k_r = 11$, $k_c = 11$, $k_{er} = 4$, $k_{ec} = 4$, $h_{op} = 1$ and $v_{op} = 1$, we obtain (16, 11) SEC-DED perfect code per row and column, thus forming 2D-SEC-DED encoding scheme according to the procedure for generating a systematic code. In our design, 11 sensor elements are installed within the SN ($k_r = 11$), each sensed with 11-bit ADC resolution ($k_c = 11$).

III. Error Correction Capability of 2D SEC-DED

When conventional 2D SEC-DED error correction method is used [6], the wireless receiver, after accepting the control message, checks the SEC-DED code in two steps. In step 1, the single parity bit (overall parity) h_{op} is checked. If it is bad, the receiver assumes that 1-bit error occurred, and it uses the other parity bits (Hamming code bits, k_{ec}), in step 2, to correct the error, according to the Table I. It may happen, of course, that three or even five bits are bad, but the SEC-DED code cannot detect all such errors. If the h_{op} is good, then there are either no errors, or two bits are bad. The receiver switches to step 2, where it uses the other parity bits (Hamming code bits, k_{ec}), to distinguish between these two cases. Again, there could be four or six bad bits, but this code cannot handle this cases.

To correct single- and double-errors at low latency cost (by software), we extend detecting and correcting processes in the algorithm presented in [3] with additional two macro-steps (Macro-Step_4 and Macro-Step_5), see Fig. 4. Each macro-step performs already defined activities, but the number of iterations is increased.

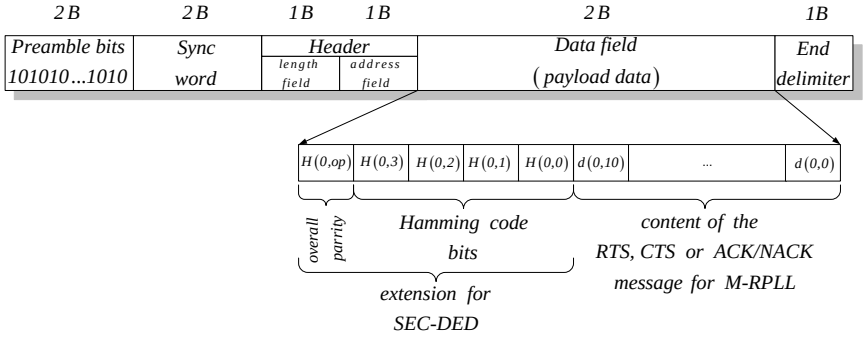


Fig. 2. Format of control messages for M-RPLL

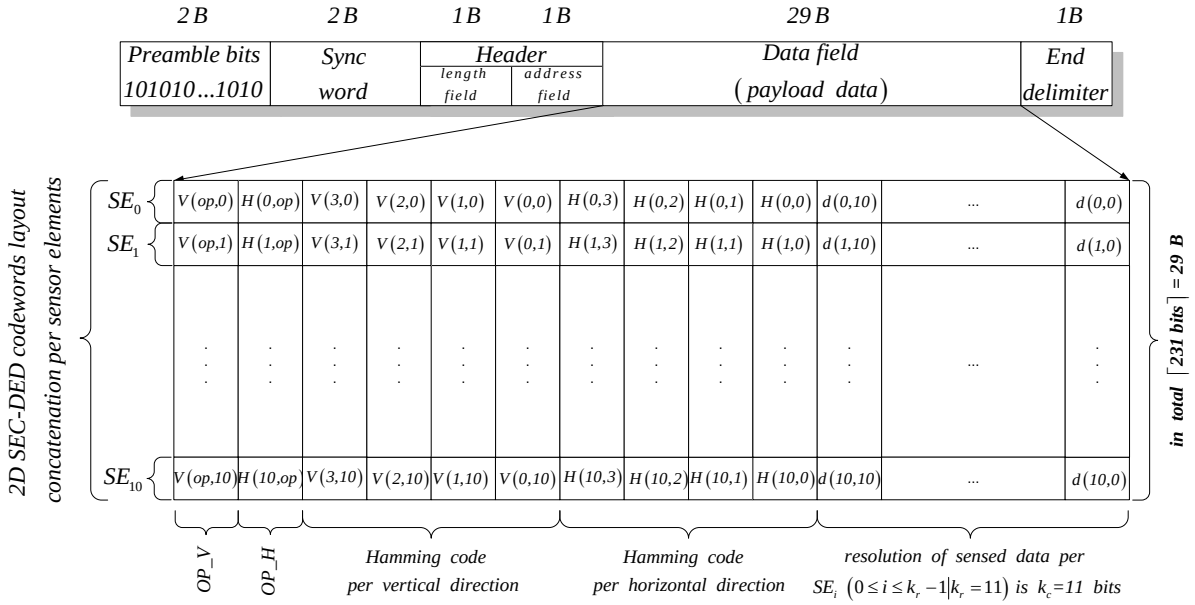


Fig. 3. Format of a data message for M-RPLL using systematic code

TABLE I
ERROR DETECTION FLAGS

Syndrome	Overall Parity	Error Type	Description
0	0	no error	No error
≠0	1	single error	Correctable (Syndrome point to incorrect bit position)
≠0	0	double error	Not correctable
0	1	parity error	Correctable (Parity error has occurred)

By using such procedure the obtained effects are the following: 1) Macro-Step_1 – error detection based on the syndrome and OP_H, and single bit error correction per horizontal direction; 2) Macro-Step_2 – error detection based on the syndrome and OP_V, and single bit error correction per vertical direction; 3) Macro-Step_3 – error detection based on the syndrome and OP_H, and single bit error correction per horizontal direction, again; 4) Macro-Step_4 – error detection based on the syndrome, OP_H and OP_V, and double bit error correction per elements at intersection points in matrix; and 5) Macro-Step_5 – error detection based on the syndrome and OP_H/OP_V, and single bit error correction per horizontal/vertical direction, again.

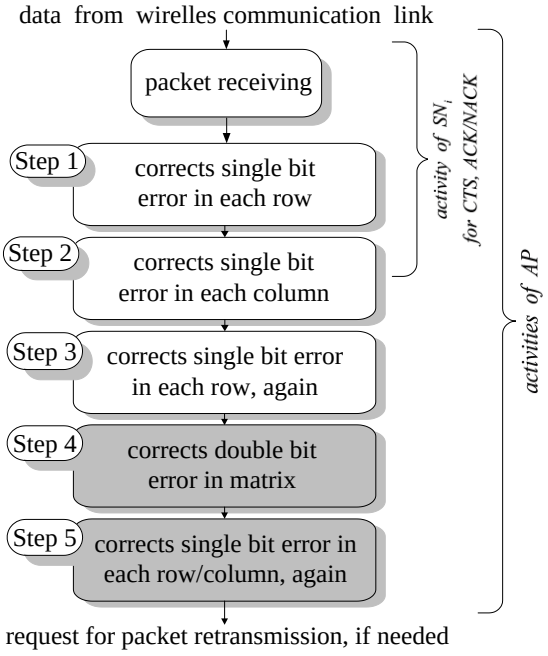


Fig. 4. Detecting and correcting processes by sensor node (SN_i) and access point (AP)

As can be seen from Fig. 4, the sender SN_i performs activity specified in Macro-Step_1, only (deals with error detection and correction of SEC-DED coded messages CTS and ACK/NACK). The AP executes Macro-Step_1 for RTS message, and activities from Macro-Step_1 up to Macro-Step_3 for 2D SEC-DED DATA coded message. Accordingly, after performing the three aforementioned steps, all single bit errors can be detected and corrected in any row or column. With the analysis carried out, all double errors are only detected but not corrected.

The patterns of possible double errors for a randomly selected data bit sequences, which cannot be corrected with the conventional algorithm (three Macro-Steps), are sketched in Fig. 5. Double bit errors in the IBF field can occur in any row or column, but for easier viewing, the pattern examples are reduced to matrix of order 4x4.

	pattern 1					pattern 2					pattern 3			
	1	2	3	4		1	2	3	4		1	2	3	4
1	x	x				x	x				x	x		
2	x	x				x	x	x			x	x		
3							x	x					x	x
4													x	x

Fig. 5. Possible positions of double errors

Our goal now deals with correction of all detected double bit errors. This is achieved by extending the previous algorithm, i.e. by introducing two additional steps. In Macro-Step_4, firstly the possible positions of all uncorrected multiple errors are determined. They are located in the intersection of those rows and columns of the matrix in which multiple errors are detected. In this way all the positions on which exists an error are detected, but besides them, some positions on which there are no real errors are marked. After that, all bits whose positions were detected as erroneous were inverted. With this bit inversion in any row or column, two real errors are corrected, but a new single error is involved. In accordance with the fact that double errors have a much higher likelihood than triple, we assume that in any row or column there can be two errors, while triple errors are very rare. So we can assume the error that was injected in the Macro-Step_4, of the proposed algorithm, is the third error in the order. Bearing in mind that some correct data bits have been inverted in this step, the values of these bits must be restored to old correct values. Therefore, to remedy this fault, we introduce Macro-Step_5, as additional. The conducted processing of this step is equivalent to Macro-Step_1 or Macro-Step_2 because the newly injected single bit errors can be corrected per horizontal or vertical direction.

IV. SIMULATION RESULTS

For data encoding, injecting, detecting and correcting errors during data transmission, a software simulation model, created in MATLAB, was used. Error-free, single- and double-/multiple- errors within a data field where injected by the MATLAB *randerr* function by using Binomial probability model for BER (bit error rate) as parameter. Let note that the most appropriate probability model of bit error from data/computer communication and networking point of view is the Binomial probability model [7]. In essence the Binomial function deals with integer valued discrete function and therefore it is appropriate model of bit error. Probability $P(i,m)$ that there are i errors in a code or message of size m ($i \leq m$) is given by a binomial frequency function, defined as:

$$P(i,m) = C_i^m \alpha^i \cdot (1-\alpha)^{m-1} \quad (2)$$

where $\alpha = BER$, $C_i^m = \binom{m}{i} = \frac{m!}{i!(m-i)!}$, $i \in \{0,1,2\}$, $i = 0$ corresponds to error-free transfer, $i = 1$ to single bit error transfer, and $i = 2$ to double-/multiple-bit error transfer. The *randerr* function [8] generates binary matrices with a specified number of zeros and ones and is a meant for testing error-control coding. The command $= randerr(N, M, [0,1,2; P_0, P_1, P_2])$, generates an $N \times M$ binary matrix having the property that each row (see Fig. 3) contains zero $'1's$ with probability P_0, one $'1'$ with probability P_1, etc. The error probability model, derived using Eq. (2), is presented in Table II. As can be seen from Table II most bit errors are single-bit or double bit errors and burst/multiple errors are present but rare. Thus, it is likely that an encoding scheme that corrects single and double-bit error can reduce a significant portion of the errors.

TABLE II
ERROR PROBABILITY MODEL

BER	Probability P_i per row [%]		
	P_0	P_1	P_2
$\alpha = 10^{-3}$	81.131	16.973	1.896
$\alpha = 10^{-4}$	97.931	2.046	0.023
$\alpha = 10^{-5}$	99.791	0.208	0.001
$\alpha = 10^{-6}$	99.979	0.020	~0.001

Notice: P_0 - error-free, P_1 - single-bit error, P_2 - double-/multiple-bits error

TABLE III
PACKET DELIVERY RATIO FOR M-RPLL PROTOCOLS

	BER				
	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
N	262800				
P_SE	262642	213288	43063	4745	434
P_D/M	259279	31081	345	4	0
PDR [%]	0.07	99.75	99.99	99.9999	~100

Notice: N - total number of transmitted packets, P_SE - # of packets with single error, P_D/M - # of packets with double-/multiple-errors

By conducting the error recovery procedure defined in Fig. 4, and the number of injected errors defined by random function (Table II), we obtain the results given in Table III. By analyzing the results, we can conclude the following:

1) For performance estimation we used a metric Packet Delivery Ratio (PDR) defined as coarse-grain indicator point to the ratio of the correctly received packets at the receiver to the total number of packets sent by the sender.

2) For BER=10^{-4}, 100% of a single-bit error and 99.99% of double-/multiple-bit errors are corrected. Let note that, from fault tolerant point of view, the involvement 2D SEC-DED/SEC-DED encoding technique is equivalent to implementation of Triple Modular Redundant (TMR) scheme at word-level voting, i.e. instead of using three packet replicas we use only one.

3) The proposed 2D SEC-DED implementation is effective when the BER (<10^{-2}) is not high and most errors are single-bit. When most of the error are bursty the proposed scheme is not as effective in reducing packet losses as is the case when erasure encoding is implemented, but at cost of higher energy consumption and computationally complexity.

V. CONCLUSION

During data transfer in WSN errors appear. In order to solve this problem in an efficient manner FEC technique is used. One of the most common FEC techniques is 2D-SEC-DED. Its main drawback is inability to correct double errors in information bits of the packet. In this paper, we propose an extension of the conventional 2D-SEC-DED procedure. The obtained simulation results show that all double errors in information field can be corrected. Having in mind that communication activity is the most power hungry one, by using 2D-SEC-DED code we drastically decrease the number of packet retransmissions and thus prolong the lifetime of the sensor node.

ACKNOWLEDGEMENT

This work was supported by the Serbian Ministry of Education and Science, Project No TR-32009 – "Low power reconfigurable fault-tolerant platforms".

REFERENCES

[1] I. F. Akyildiz, M. C. Vuran, Wireless Sensor Networks, Chester UK, John Wiley & Sons Ltd; 2010

[2] F. Koushanfar, et all, Fault Tolerance in Wireless Sensor Networks, in: Handbook of Sensor Networks, by I. Mahgoub and M. Ilyas (eds.), Section VIII, Ch. 36, pp. 36.1-13, CRC Press, Boca Raton, 2004

[3] Goran S. Nikolic, et all, "Reliable data transfer Rendezvous protocol in wireless sensor networks using 2D-SEC-DED encoding technique", Microelectronics Reliability, Volume 65, October 2016, pp 289-309

[4] M. R. Kosanović, M. K. Stojčev, "RPLL - Rendezvous Protocol for Long-Living Sensor Node", Facta Universitatis Series: Electronics and Energetics, Vol. 28, No. 1, (2015), pp. 85-102

[5] M. Brzozowski, K. Piotrowski, P. Langendoerfer, "A Cross-Layer Approach for Data Replication and Gathering in Decentralized Long-Living Wireless Sensor Networks", ISADS 2009, In Proc. of the 9th ISADS, 2009. pp. 49-54

[6] S. Lin, D. J. Costello, Error Control Coding, 2nd ed. Upper Saddle River, New Jersey: Pearson Prentice Hall; 2004.

[7] C. T. Bhunia, Information Technology Network and Internet, New Age International Publisher, New Delhi, (2005), pp. 173

[8] Mathworks, *randerr*, Generate bit error patterns, at: http://www.mathworks.com/help/comm/ref/randerr.html?requestedDomain=www.mathworks.com, (accessed 2018)

Background subtraction and detection of outliers in two-dimensional measured data

Szilvia Nagy[1], Brigita Sziová[2], Levente Solecki[3] and László T. Kóczy[2]

Abstract – **For colonoscopy images the main information is in the fine structure of the surface of the bowel or colorectal polyps, similarly to the case of combustion engine surface scans, where the grooving and wear can be detected from the fine pattern superponed to a cylinder curvature.**

In both cases appear outliers, colonoscopy images and to have many reflections, whereas the roughness scanners detect small dust particles and well as the vibrations from the environment.

The method presented in this paper takes care of both the problems using histogram stretching together with a special type of filtering. Also, masks are introduced in order to control the effect of the operators.

The effects of the processing steps on the structural entropy of the image is also studied, as structural entropies are used in characterization of the images.

Keywords – **Colonoscopy, Roughness, Background subtraction, Reflection, Outliers, Entropy.**

I. INTRODUCTION

In measured 2D data, such as images or surface scans there can be characteristic lengths that carry the information which is needed, while larger or smaller scale patterns are not necessary. Also, in measured data such points can arise, which do not fit into the trend of the other measured points. In the case of images or scanned surfaces such outliers can also form larger domains. In the following considerations we give a method for eliminating the outliers as well as removing larger scale background.

We apply our method for two kinds of data, the first one is microgeometrical scan of combustion engine cylinder surface, where the data consist of height values measured on a regular 2D grid, the second is colonoscopic images, where the data consist of 3 colour intensity values at points of regular grid, again. The background that should be neglected is the curvature of the cylinder surface and the bowel wall respectively, while the outliers are the measurement noise, and the impurities in the first case and the light reflections from the wet surface in the second one.

Colonoscopy is mainly used in diagnosis of colorectal cancer, which is one of the most frequently occurring cancer types. It develops from special types of colorectal polyps [1,2], thus diagnosing in the pre-cancerous, polyp phase helps survival greatly. Not all the polyp types have the tendency to develop into malign object, these polyps can remain in the bowel without causing any complications, moreover, it is better to leave them intact, as all the surgeries can potentially cause damage. The polyps can be classified as benign or malign according to the pattern on their surface [3-6]. Of course, the first task is to find the polyp, which is done by qualified medical experts nowadays, however there are several groups [7-9] studying the possibility of an automated detection. Previously, we developed a method that can classify an image segment to containing or not containing polyp classes, which works well for several types of images [10-11], however, for low quality pictures, the method fails. This was the reason we tried to preprocess the images.

In combustion engines the inner surface of the cylinder must be smooth enough to let the piston move, and at the same time it must store oil and the tiny particles arising as wear byproducts. The cylinder lining's pattern depends on the process it is made by. One of the most traditional and widespread method is honing, which results in straight grooves with a slight slope compared to the surface of the piston. The movement of the piston, however, causes fine scratches parallel to its movement. These patterns can be detected by surface scanners.

In order to build reliable classification scheme, we used Rényi entropy combinations and tested, whether there was difference between the new and the worn surfaces. However, in this classification, the large-scale pattern of the background dominates, and the fine pattern's own entropy is hard to distinguish. Outlier bring even more unwanted information to the system, thus we decided to remove them from the pictures.

In this article, after describing the image properties in Section 2, we give the histogram stretching and mask acquiring method together with the preparation of the outlier pattern to be subtracted. Section 3 describes the treatment of the large-scale background with polynomial fitting and large-scale averaging and give the results. The last section contains the conclusion.

II. IMAGE PROPERTIES

In the case of the colonoscopy mages, we used image databases from multiple sources [7-9]. These images are taken in non-ideal circumstances, the bowel is always dark inside, so the endoscope has a light source, which is point-like. At the same time the inner surface is wet, moving, pink (ideally), i.e. it is very far form ideal image source. They are almost

[1]Szilvia Nagy is with the Department of Telecommunications, Széchenyi István University, H-9026 Győr, Egyetem tér 1, Hungary, nagysz@sze.hu

[2]Brigita Sziová and László T. Kóczy are with the Department of Information Technology, Széchenyi István University szibr@sze.hu, koczy@sze.hu.

[3] Levente Solecki is with the Department of Vehicle Manufacturing, Széchenyi István University, solecki@sze.hu.

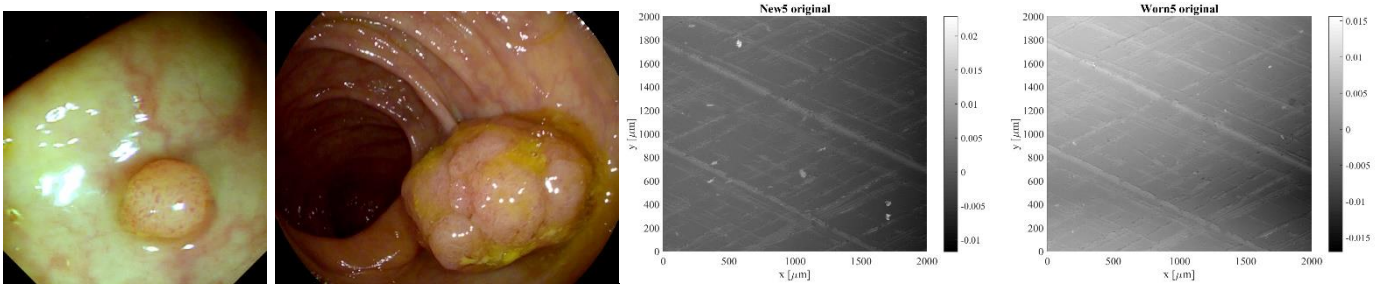


Fig. 1. First two photographs: endoscopy images (picture No. 219 from database CVC_Colon [7] and No. 116 from ETIS-Larib [8]) with visible reflections, blood vessels, polyp patterns, bubbles and impurities. 3rd and 4th subplot: new and worn surface scans at measuring point 5 of the combustion engine cylinder surface, with visible curving and outliers (the outliers are trimmed down to make the other details visible).

monotonic, with lots of reflections, multiple types of patterns from veins to polyp pits to the more-or-less distorted cylindrical shape of the bowel and the polyp protruding into the bowel space.

In the case of the combustion engines, we took silicone replicas from the new engine, and of the same part of the surface after 500 hours polycyclic endurance test. For preparing the samples Struers RepliSet-F5 replica kit was used. The replicas measured at 5 different points along their axes (each to different environments: the first point was between the first two piston rings, the 5th location a bit over the lowest turning point of the piston) with TalysurfCLI2000 surface scanner's optical sensor based on chromatic length aberration (CLA). These images are much easier to treat than the colonoscopy ones, as they have quite uniform pattern, a regular background and few, very small-sized outliers.

III. Contrast enhancement, mask preparation and reflection filtering

After separating the colour channels of the colonoscopy images, we had to treat 3 separate 2D data, similarly to the case of the surface scans. Studying the intensity histograms of the images, in both cases the outliers – the reflections and the impurities – form a large peak around the maximum intensity value. In the case of the colonoscopy images there is a large peak at the minimum intensity value as well, because all the images have black frame. In the case of the scanned surfaces there are usually some measurement error toward the low levels as well, this means however, only a small peak at the minimum intensity value.

As a first step, we stretched the histogram so, that the peaks around the highest and lowest values would be cut off. Instead of these points we used the maximum and the minimum values of the new histogram, respectively. During this process we also defined a mask both for those points where the intensity was below the lower threshold, and for those above the upper threshold. The low-masks consisted of 0s in those points where the intensity was below the lower limit, and 1 in the other points, while the high-masks were 1 everywhere, except for the points where the intensity was above the higher threshold.

As a next step we extended the masks to those neighbouring points, where the intensity value was a by a certain level higher than its environment excluding the already masked points.

With the new masks we calculated such a matrix-to-be-subtracted which was 0 everywhere, except for the masked points, where the values were calculated the following way. For each of the points we determined the average intensity of its surroundings of size 41 by 41 but did not count the points with 0 mask values into the surroundings. We put the difference between the average and the actual pixel intensity to the matrix-to-be-subtracted and used it for eliminating the outliers.

The masks were used differently in the case of the two types of data. For surface scans, both the high and the low mask covered valuable domains of the matrix, while in the case of the colonoscopy images the black mask denoted those parts of the image where the colonoscope could not provide data. As a result, for colonoscopy, these black points had to be excluded from the points of the matrix-to-be-subtracted, while for engine cylinder surfaces, not.

As a last step we applied a mean filtering around the edges of the masks, if it was necessary.

As an example, the matrix-to-be-subtracted and the resulting picture of the second subplot of Fig. 1 can be seen in Fig. 2 and Fig. 3. (in the case of the cylinder surfaces the matrices to be subtracted are hardly visible due to the large size of the scan (2000 by 2000 points) and the small sizes of the measurement errors, especially for the low outliers.

IV. Background subtraction

As in the case of the combustion engine cylinder surfaces the background is a cylinder of very large diameter compared to both the honing pattern and the size of the scanned surface element, as a first approach we fitted a second order polynomial to the matrix of the scanned surface. This approach was effective in some cases, sometimes, higher order polynomials were necessary. There were more surface scans, however, where this method resulted in even more complicated background pattern. Clearly, polynomial fitting could not be used in the case of the bowel pictures, as the background is way too complicated for a polynomial.

The simplest solutions for determining the background pattern seemed to be a large-scale averaging throughout the complete picture. Of course, in the case of the colonoscopy pictures the black frame should not be counted into the average, as it does not belong to the valuable part of the image, thus we applied the low-mask again. The resulting background is surprisingly near the ideal one, its effect can be seen in the first

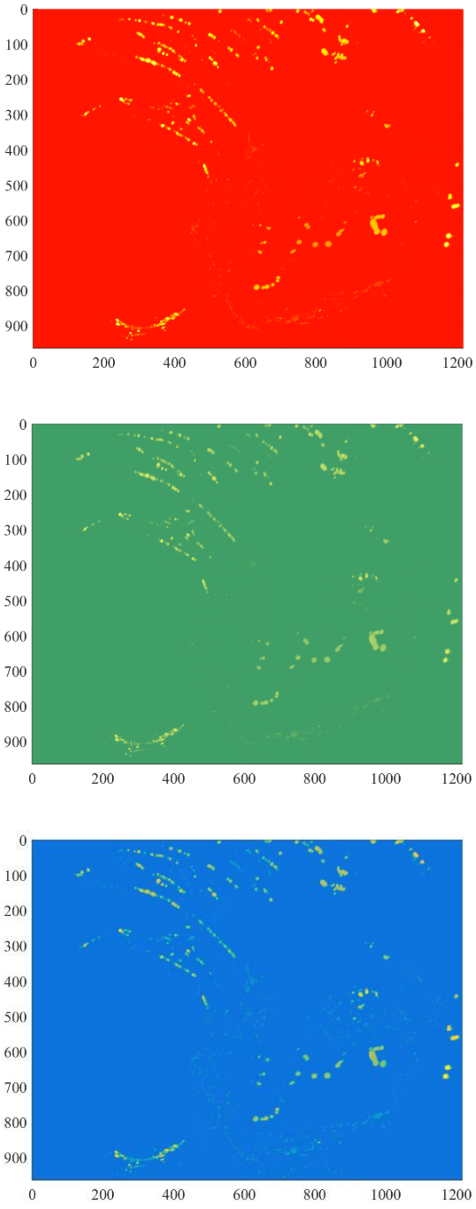


Fig. 2. Matrix-to-be-subtracted of the second picture of Fig. 1 (picture 116 from ETIS-Larib [8]) fo the three colour channels R, G, and B.

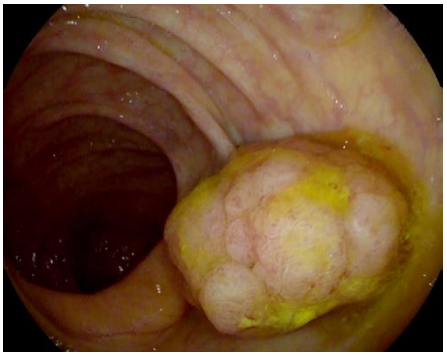


Fig. 3. The corrected picture after reflection subtraction for a colonoscopy image, for the 2nd picture of Fig. 1.

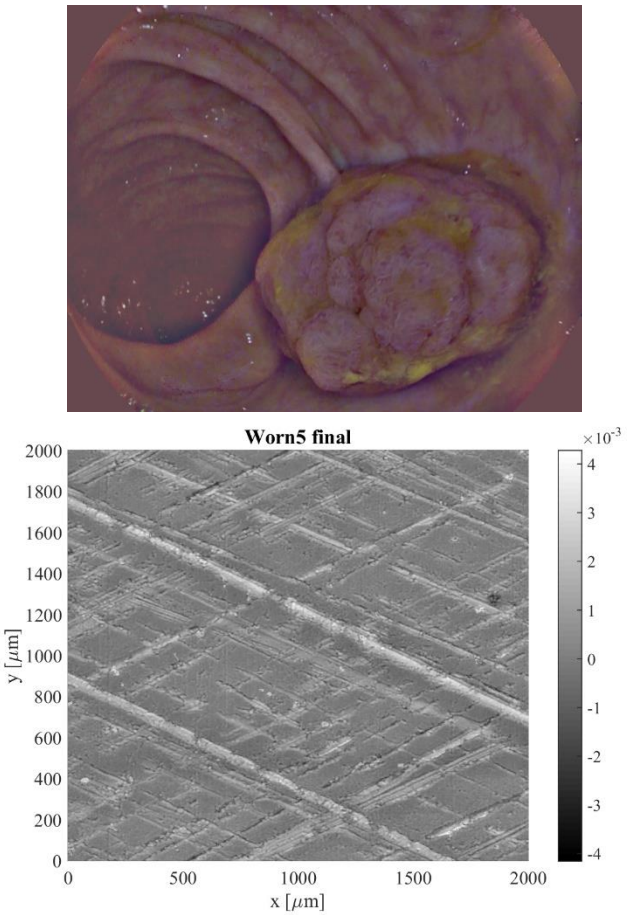


Fig. 4. The corrected pictures after reflection and background subtraction for the 2nd and 4th subplots of Fig. 1.

subplot of Fig. 4. In the case of the combustion engines, the changing of the order of the process made the detection of the outliers simpler. The resulting picture can be seen in the second subplot of Fig. 4.

V. Structural Entropy as a Basis of Classification

The purpose of this image processing was to determine, whether it can improve the classification efficiency of our scheme [10,11]. For the engine surfaces we started to develop a classification [12], where the wear of the surface could be determined from the scan, while for the colonoscopy images determining the polyp content was the goal. In the case of both data types we used a fuzzy classification scheme with structural entropy as one of the antecedents.

Structural entropy-based characterization was first used in quantum mechanics [13], [9-11], then it was adapted scanning microscopy images of semiconductor device parts [12]. Later it was found to characterize the surface roughness of gold catalyzers well [13].

These structural entropy quantities are defined using the generated entropies of Rényi. The *n*th entropy can be written as

$$S_n = \frac{1}{1-n} \ln \left(\sum_{i=1}^{N} {I_i}^n \right), \quad (1)$$

Here I_i means the normalized pixel intensity, and N the number of pixels. The pixel intensities must fulfill the requirements of being a probability distribution, i.e., all intensity should be non-negative and its sum should be 1.

Pipek and Varga studied two of Rényi entropy differences

$$S_{str} = S_1 - S_2, \text{ and } -\ln q = S_0 - S_2, \quad (2)$$

and found that these quantities together describe the shape of the distribution. The quantity q is known in quantum mechanics, it is the spatial filling factor.

If $S_{str}(q)$ is plotted for an image, it can be told from the plot, what type of localization does the intensity have, i.e., if the picture contains mainly exponentially decreasing components, its $S_{str}(q)$ point will be near the curve belonging to exponential distributions.

VI. The processing steps' effects on the structural entropy maps

The structural entropies moved to very low (10^{-4}) values in the case of the combustion engines but remained along the same path for both the new and the worn surfaces. The new-worn classification is not improved, moreover, in many locations it became less effective.

In the case of the colonoscopy images, the structural entropies also decrease in magnitude, however, the processed pictures usually perform slightly better than the original pictures.

In both cases the localization types remain very slow, Gaussian, or even slower.

VII. Conclusion

A method for subtracting the outliers and the background for two types of 2D data is presented. Using the histogram of the images, the lowest and highest value points, that are sparsely populated are cut off, and marked as outliers on masks. The masks are used for excluding these outliers from the averages when the new values corresponding to the masked points are calculated.

Background subtraction is based on large-scale averaging of the data proved to be more effective than polynomial fitting even in the case of regular surfaces such as combustion engine cylinder surfaces. The processing makes the image features more distinguishable to the naked eye, however, for computational intelligence-based classification methods, its effect is not always positive.

Acknowledgement

The authors would like to thank the financial support of the project EFOP-3.6.2-16-2017-00015, GINOP-2.3.4-15-2016-00003, and the ÚNKP-17-4 New National Excellence Programme of the Ministry of Human Capacities of Hungary. This work was supported by National Research, Development and Innovation Office (NKFIH) K108405 and K124055.

References

[1] K. Søreide, B.S. Nedrebø, A. Reite et al., „Endoscopy Morphology, Morphometry and Molecular Markers: Predicting Cancer Risk in Colorectal Adenoma", Expert Rev. Mol. Diagn, vol. 9, pp. 125-137, 2009.

[2] I. Rácz, M. Jánoki, and H. Saleh, "Colon Cancer Detection by 'Rendezvous Colonoscopy': Successful Removal of Stuck Colon Capsule by Conventional Colonoscopy", *Case Rep. Gastroenterol*., Volume 4, Karger, 2010, pp. 19–24.

[3] J. R. Jass, "Classification of colorectal cancer based on correlation of clinical, morphological and molecular features", *Histopathology*, Volume 50, Wiley, 2006, pp. 113–130.

[4] S. Kudo, S. Hirota, T. Nakajima, et al., "Colorectal tumours and pit pattern". J Clin Pathol, vol. 47, pp.880-885, 1994.

[5] S. Kudo, S. Tamura, T. Nakajima, et al. Diagnosis of colorectal tumorous lesions by magnifying endoscopy. Gastrointest Endosc, vol. 44, pp. 8-14, 1996.

[6] S. Kudo, C.A. Rubio, C.R., Teixeira, et al. Pit pattern in colorectal neoplasia: endoscopic magnifying view. Endoscopy, vol. 33, pp. 367-373, 2001.

[7] J. Bernal, F. J. Sanchez, F. Vilariño. "Towards Automatic Polyp Detection with a Polyp Appearance Model", Pattern Recognition, Vol. 45, pp. 3166-3182, 2012.

[8] J.S. Silva, A. Histace, O. Romain, X. Dray, B. Granado, "Towards embedded detection of polyps in WCE images for early diagnosis of colorectal cancer", Int J Comput Assisted Radiology and Surgery, Vol. 9, pp. 283-293, 2014.

[9] J. Bernal, F. J. Sanchez, G. Fernández-Esparrach, D. Gil, C. Rodrígez,F. Vilariño, "WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians", Computerized Medical Imaging and Graphics, Vol. 43, pp. 99-111, 2015

[10] Sz. Nagy, F. Lilik, L. T. Kóczy, "Entropy based fuzzy classification and detection aid for colorectal polyps", IEEE Africon 2017, Cape Town, South Africa, 15-17 September 2017.

[11] Sz. Nagy, B. Sziová, L. T. Kóczy, "The effect of image feature qualifiers on fuzzy colorectal polyp detection schemes using KH interpolation - towards hierarchical fuzzy classification of coloscopic still images", accepted for publication at FuzzIEEE2018, Rio de Janeiro

[12] L. Solecki, Sz. Nagy, "Wavelet Analysis and Structural Entropy Based Intelligent Classification Method for Combustion Engine Cylinder Surfaces", Proceedings of the 8th European Symposium on Computational Intelligence and Mathematics, ESCIM, 5-8th October 2016, Sofia pp. 115-120.

[13] J. Pipek and I. Varga, "Universal classification scheme for the spatial localization properties of one-particle states in finite *d*-dimensional systems", *Phys. Rev. A*, Volume 46, APS, Ridge NY-Washington DC, 1992, pp. 3148--3164.

[14] I. Varga and J. Pipek, "Rényi entropies characterizing the shape and the extension of the phase space representation of quantum wave functions in disordered systems", *Phys. Rev. E*, Volume 68, APS, Ridge NY-Washington DC, 2003, 026202.

[15] L. M. Molnár, Sz. Nagy, and I. Mojzes, „Structural entropy in detecting background patterns of AFM images", Vacuum, Volume 84, Elsevier, Amsterdam, 2010, pp. 179-183.

A. Bonyár, L. M. Molnár, G. Harsányi, "Localization factor: a new parameter for the quantitative characterization of surface structure with atomic force microscopy (AFM)", *MICRON*, Volume 43, Elsevier, Amsterdam, 2012, pp. 305-310.

C# Implementation of Split and Merge Algorithm for Image Segmentation

Velin Iordanov[1] and Ivo Draganov [2]

Abstract – **In this paper, we propose a C# implementation of the split and merge algorithm with a main applicability directed towards image segmentation. The algorithm process color and grayscale images by splitting them following quad-tree decomposition over intensity homogeneity criterion. Then adjacent blocks, including such from different tree levels, are merged together to form candidate regions for segmentation. Area filtration of the regions is the final processing step prior to generating the output map of the application. Testing is done with images containing printed text over complex background and promising results are registered.**

Keywords – **Image Segmentation, Split and Merge, C#, Area Filtration, Text Extraction.**

I. INTRODUCTION

Object detection in digital imagery and its recognition is a research area in which the main goal is to create a program for automatic segmentation of content visually embedded on a complex background.

Methods for object detection, in particular printed text, on a complex background form three groups: bottom-up [7, 12, 14], heuristic top-down [13, 2, 14] and top-down with machine learning [6, 8].

In "bottom-up" methods, the location of the text is not actually determined, it directly segments the image of areas, and those collapsed as symbolic are merged into words. Lienhart [7] segments partitioned images with a split-merge algorithm [5] and a growing region algorithm [9], while Bunke [12] groups the pixels of the text using a color clustering algorithm .These methods are sensitive to the size of the symbols, the available noise and the background type.

Through heuristic "top-down" methods, text blocks are first identified in the image processed by heuristic filters, then segmented into text and background areas. Difficulties arise here, both in detection and in segmentation. In [14] it is suggested that the text is contained in areas with a highly variable horizontal structure and additional spatial properties to be established by analyzing the related components. Smith et al. [11] localize the text by first locating the vertical edges with a preformed pattern, and then grouping them into text areas through a blurring process. The latter two approaches are quick, but many false searches occur, because often parts of the background also have a highly contrasting horizontal structure. Wu et al. [13] describe a text localization method based on texture segmentation. The method has a high computational complexity and is sensitive to background noise. In a more recent work Garcia [2] suggests a localization of text called a change in edge orientation based on the fact that text strings contain edges of different orientation. The disadvantage of the method is the non-use of multiple parallel edges, also characteristic of text strings. Apart from the peculiarities of the individual symbols, Sobettka et al. [12] offer detection based on a string when locating it.

Text detection problems - top-down heuristic methods are empirical and rely on predefined (operator) visual signs that often prove to be inappropriate on a complex background; multiple false searches (80-500%) are generated; learning methods somewhat solve this problem, but the following issues remain problematic: how to avoid the considerable time of classification applied to the entire image and how to reduce the variation in character size and change in brightness in terms of more stable features obtained in the pre-training phase? The process of normalization of the signs here occupies a central place as an object of intensive research.

In this study we propose a new direct approach of detecting isolated printed characters in images based on their size falling in preliminary defined limits over horizontals and verticals. The segmentation is based on contrast discrimination in small area employing the split-and-merge algorithm which attempts to overcome the major limitations of the aforementioned realizations. It is considered fast and reliable tool given a priori knowledge of the font sizes within processed images.

In Section II of the paper full description of the implemented algorithm is presented. It is done in C# following a flexible object-oriented module, also described as a structure in the same section. In Section III some experimental results are presented and then discussed in Section IV followed by a conclusion in Section V.

II. IMPLEMENTATION DESCRIPTION

A. Algorithm steps

Grayscale images are directly passed to the input of the system. If the image is in color then we covert it grayscale according to:

$$m[i][j] = 0.3 * r[i][j] + 0.59 * g[i][j] + 0.11 * b[i][j], \quad (1)$$

where $i=[0,M-1]$, $j=[0,N-1]$ and the resulting array should be of type double for preserving precision. After all calculations we go back to unsigned char representation within the range from 0 to 255.

[1]Velin Iordanov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: velin.iordanov@gmail.com

[2]Ivo 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

The obtained array *m[M][N]* is divided into quarters (Fig. 1) - if any of the dimensions *M* and/or *N* does not allow for an exact division (odd number), then the first halves (left and/or right) are taken as a half of the whole increased with 1 (it will always be even) and the rest with 1 less (e.g. if *M*=7, the quartile lengths to the left of the whole array are assumed to be 4 (half the nearest even number) and the lengths of the quarters to the right are taken 3).

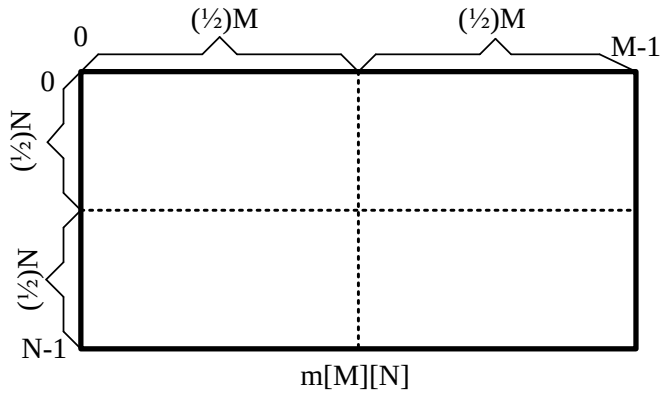


Fig. 1. Two-dimensional array split in quarters

For each quarter the element with maximum and minimum value is found and so is the difference in absolute value (positive number) and it is compared with predefined threshold - thresh1. If thresh1 is greater, all the values of the quarter elements are replaced with the average of their old values (again rounded to an integer, but the averaging itself is not an integer but a double). If thresh1 is less, the current quarter is divided again into quarters (if again one of its two or both sizes is an odd number - the approach from above is applied) and recheck for each new one if the absolute value of the difference between the maximum and minimum is greater than or less than thresh1 and the above steps apply. The splitting of quarters continues until a single element (1 value of the entire matrix) is reached if necessary.

Here we have rectangular blocks of different sizes (generally from *(M/2)x(N/2)* to 1x1 elements) each containing elements of the same values (the average of their original values). For each of these blocks, check whether their width or height (both sizes) exceeds thresh2 - if only one of these sizes exceeds it - all values of the given block are equalized to zero.

The array with dimensions *MxN* elements already has blocks containing zeros and positive values (the same within each block separately). For each block consisting of non-zero elements, all adjacent blocks are checked, which are also non-zero. The check is done by value - if the difference per module between the value of the current block (it is one for the whole) and the currently comparing neighboring (also one) is less than thresh3 - the two blocks are assumed to be "connected" and both are replaced with the weighted average of their two previous values - the number of pixels in one area of its brightness, collected by the number of pixels of the other in its brightness and divided by the total number of pixels for the two areas.

So far we have obtained a two-dimensional array *MxN* elements in which we have a number of areas (macro blocks) whose values are constant (equipotential domains). For each area obtained, the width and height (in number of elements) of the smallest rectangle surrounding it (Fig 2). If either the width or height of an area is less than thresh4, the area values are equal to zero.

Now we have a two-dimensional array MxN element containing equipotential domains, each with a value between 0 and 255 (including boundaries) - it is necessary one to be able to "return" it as a source parameter (e.g. by a pointer to the memory area) and we can also save it from this function in a binary file on the disk. All other intermediate areas (excluding the last received array) from the RAM for intermediate operations should be cleared.

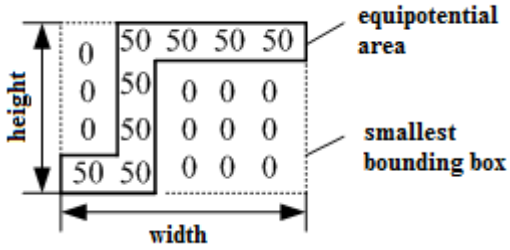


Fig. 2. Equipotential area dimensions estimation

B.Source design

The diagram of the classes SplitAndMergeSegmenter and ImageSegment are given in Fig. 3. The splitting is implemented recursively within the ImageSegment class where all obtained segments with its internal values and position within the matrix of the whole image. Division is applied until block size reaches 1x1 pixels.

The method Merge use a list of all registered segments. According to the algorithm description from Section II.A weighted average intensity is associated to the resulting bigger regions after merging.

The essential parts from the definitions of the classes are listed below:

```
public class SplitAndMergeSegmenter
  {
    private const double RedMultiplier = 0.3;
    private const double GreenMultiplier = 0.59;
    private const double BlueMultiplier = 0.11;
    private readonly ICollection<ImageSegment>
segments;
    private readonly Bitmap image;
    private readonly int thresh1;
    private readonly int thresh2;
    private readonly int thresh3;
    private readonly int thresh4;
    public SplitAndMergeSegmenter(Bitmap image, int
thresh1, int thresh2, int thresh3, int thresh4)
    {
      this.image = image;
      this.segments = new List<ImageSegment>();
      this.thresh1 = thresh1;
```

```
            this.thresh2 = thresh2;
            this.thresh3 = thresh3;
            this.thresh4 = thresh4;
        }
    private void Merge(IEnumerable<ImageSegment> imageSegments)
        {
            foreach (var segment in imageSegments)
            {
                var block = imageSegments.Where(x => CanMerge(x, segment));
                if (block == null)
                {
                    continue;
                }
                foreach (var item in block)
                {
                    CombineSegments(segment, item);
                }
            }
    …
    }

    public class ImageSegment
      {
        public ImageSegment(int[,] pixels, int topLeftX, int topLeftY, int bottomLeftX, int bottomLeftY, int topRightX, int topRightY, int bottomRightX, int bottomRightY)
        {
            this.Pixels = pixels;
            this.topLeftIndexX = topLeftX;
            this.topLeftIndexY = topLeftY;
            this.topRightIndexX = topRightX;
            this.topRightIndexY = topRightY;
            this.bottomLeftIndexX = bottomLeftX;
            this.bottomLeftIndexY = bottomLeftY;
            this.bottomRightIndexX = bottomRightX;
            this.bottomRightIndexY = bottomRightY;
            this.segments = new List<ImageSegment>();
        }
        public int[,] Pixels { get; set; }
        public int topLeftIndexX { get; set; }
        public int bottomLeftIndexX { get; set; }
        public int topRightIndexX { get; set; }
        public int bottomRightIndexX { get; set; }
        public int topLeftIndexY { get; set; }
        public int bottomLeftIndexY { get; set; }
        public int topRightIndexY { get; set; }
        public int bottomRightIndexY { get; set; }
        public bool isZeroed { get; set; }
        public bool isCombined { get; set; }
    public ICollection<ImageSegment> segments { get; set;}
        public void SetAvarageValue(int number)
        { for (int i = 0; i < this.Pixels.GetLength(0); i++)
            { for (int j = 0; j < this.Pixels.GetLength(1); j++)
                { this.Pixels[i, j] = number; } } } }
```

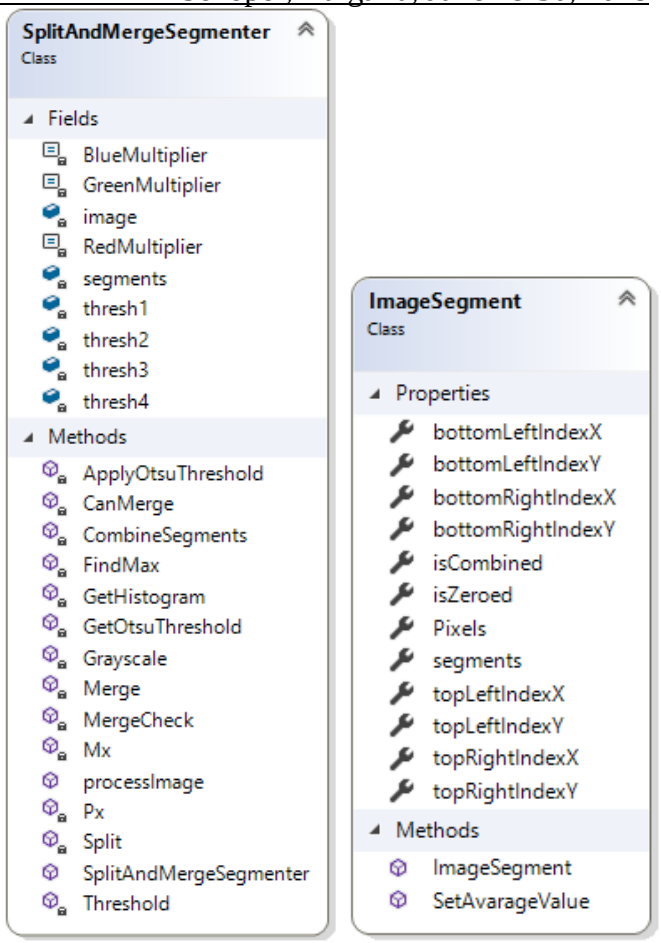


Fig. 3. Diagrams of the main application classes

III. Experimental results

The experimental test set includes 100 color images in RGB format at 24 bpp with 320x240 pixels in dimension. They are processes on PC compatible machine with Intel Core 2 Quad Q8300 CPU running at 2.5 GHz with 4 GB of RAM under Windows 10 OS. Empirically obtained thresholds based on font sizes present in the test images are set to be: thresh1 = 100, thresh2 = 5, thresh3 = 70, thresh4 = 1. Segmentation accuracy and execution time of the current implementation are presented in Table I compared to those of the implementation of Wu et al. [13].

Table I
Segmentation accuracy and time

Para-meters / Method	Correctly segmented symbols, %	Fragmented symbols, %	Missed symbols, %	Average execution time, sec
[13]	79.6	7.6	12.8	-
Proposed	81.3	10.5	18.2	0.5

A single image with its resulting segmented appearance can be observed in Fig. 4. Due to the low resolution with high compression ratio in MPEG-1 format (being an excerpt from SDTV footage with subsequent downsampling) considerable amount of false detections are attached to the symbols.

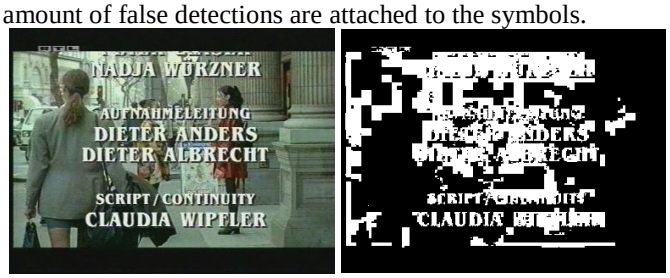


Fig. 4. Low resolution TV footage with lots of false detections

Some sample segmented images outside the main test set are given in Fig. 5 and Fig. 6 where different type of errors in segmentation appear. All the characters from the image in Fig. 5 are well detected but their edges are jagged because of the small scale to which splitting blocks emerge at the last segmentation step.

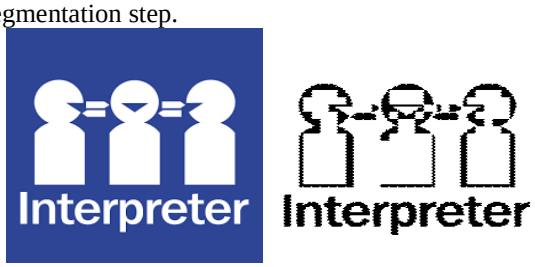


Fig. 5. Test results over artificially generated image

In the image from Fig. 6 there is large number of missing letters from the lower part of the text field. The reason for that is the small font size. Here, additional deteriorations exist as a sequence of the more complex forms of the characters.

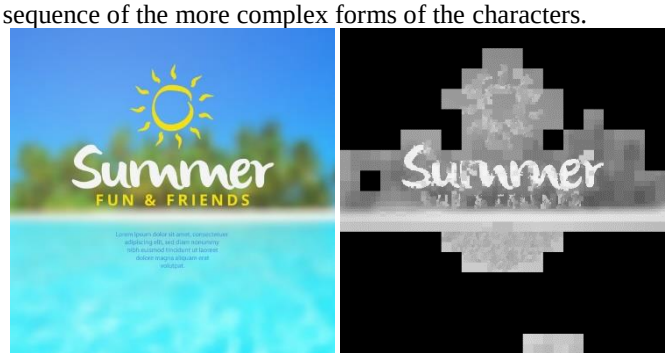


Fig. 6. Processing results for text with various font types and sizes

IV. DISCUSSION

The main influencing factors are textures from the complex backgrounds in the scenes. This hampers the process of accurate merging for the blocks that belong to the print symbols. Another factor is the thickness of the symbols, which, if below a certain threshold, impedes the removal of part of the blocks after segmentation. When area filtration is carried out, sometimes it leads to erosion of the symbols. Further, accurate segmentation is limited by overlapping translucent text on a complex background. There solid color fill of the individual symbols acquire color component variations, which also leads to disruptions in separate parts.

V. CONCLUSION

The proposed implementation of the split and merge algorithm in C# offers performance which satisfies potential users from practical point of view using contemporary Desktop computers. Its object-oriented class definitions may be easily extended. It is a practical ground for future modifications of the algorithm in order to cope with fonts more complicated in form and imposed in smaller sizes. Object textures from the background resembling present characters is the other direction to which further enhancements should be sought.

REFERENCES

[1] D. Chen, K. Shearer, H. Bourlard, “Text Enhancement with Asymmetric Filter for Video OCR”, In Proc. of the 11th Int. Conf. on Image Analysis and Processing, pp. 192–198, Sept. 2001.

[2] C. Garcia, X. Apostolidis, “Text Detection and Segmentation in Complex Color Images”, In Proc. of the Int. Conf. on Acoustics, Speech and Signal Processing, pp. 2326–2329, 2000.

[3] O. Hori, “A Video Text Extraction Method for Character Recogntion”, In Proc. of the Int. Conf. on Document Analysis and Recognition, pp. 25–28, Sept. 1999.

[4] H. Kamada, K. Fujimoto, “High-speed, High-accuracy Binrization Method for Recognizing Text in Images of Low Spatial Resolutions”, In Proc. of the Int. Conf. on Document Analysis and Recognition, pp. 139–142, Sept. 1999.

[5] R. Laprade, M. Doherty, “Split-and-merge Segmentation using an F Test Criterion”, SPIE image understanding and man-machine interface, pp. 74–79, 1987.

[6] H. Li, D. Doermann, “Text Enhancement in Digital Video using Multiple Frame Integration”, In Proc. of the ACM Multimedia, vol. 1, pp. 385–395, Orlando, Florida, USA, 1999.

[7] R. Lienhart, “Automatic Text Recognition in Digital Videos”, In Proc. SPIE, Image and Video Processing IV, pp. 2666–2675, Jan. 1996.

[8] R. Lienhart, A.Wernicke, “Localizing and Segmenting Text in Images and Videos”, IEEE Trans. on Circuits and Systems for Video Technology, vol. 12, no. 4, pp. 256–268, 2002.

[9] T. Pavlidis, Y.T. Liow, “Integrating Region Growing and Edge Detection”, IEEE Trans. on Pattern Analysis and Machine Intelligence, pp. 225–233, 1990.

[10] T. Sato, T. Kanade, E. K. Hughes, M. A. Smith, “Video OCR for Digital News Archives”, In Proc. of the IEEE Workshop on Content Based Access of Image and Video Databases, pp. 52–60, Bombay, Jan. 1998.

[11] M. A. Smith, T. Kanade, “Video Skimming for Quick Browsing based on Audio and Image Characterization”, Technical Report CMU-CS-95-186, Carnegie Mellon University, July 1995.

[12] K. Sobottka, H. Bunke, H. Kronenberg, “Identification of Text on Colored Book and Journal Covers”, In Proc. of the Int. Conf. on Document Analysis and Recognition, pp. 57–63, 1999.

[13] V.Wu, R. Manmatha, E. M. Riseman, “Finding text in images”, In Proc. of the ACM Int. Conf. Digital Libraries, pp. 23–26, 1997.

[14] Y. Zhong, K. Karu, A. K. Jain, “Locating Text in Complex Color Images”, Pattern Recognition, vol. 10, no. 28, pp. 1523–1536, 1995.

Web-Based Application for Digital Image Classification using TensorFlow

Ivan Videv[1] and Ivo Draganov[2]

Abstract – **In this paper, we investigate the capabilities of applying TensorFlow for digital image classification. A complete web-based application is created with easy portability to various operating systems, which cover all the functionalities incorporated into the library. It follows the client-server model and take advantage of all the options for tuning, such as random crop, random scaling, etc. Testing with a set of 772 images of 3 types of objects taken in natural environment prove the applicability of the framework for classification purposes. The system can be extended to solve general and strictly specialized tasks.**

Keywords – **Digital Image Classification, TensorFlow, Deep Learning, Convolutional Neural Network.**

I. Introduction

The algorithms for digital image classification employ two types of preliminary training schemes with labelled datasets:

- Unsupervised learning – the input information is not divided and described prior to its use – there are no categories for association;
- Supervised learning – there are means (teacher) for describing the initial data for training, so the categories exist before the process takes place. Classification and regression are the two main stages here.

K-means clustering [1] is one of the very popular unsupervised learning techniques. Main measure for pertaining of an object to certain cluster is the distance to its center:

$$V = \sum_{i=1}^{k} \sum_{x_j \in S_i} \left(x_j - \mu_i\right)^2. \quad (1)$$

The number of clusters is initially known but their centers can change their position during training.

Other kind of unsupervised learning is the associative type [2]. Its primary application is directed towards finding relations between variables in databases with large volume. If a certain sequence occurs more frequently than others do, then its sub-sequences would also occur more frequently that other shorter series.

Unsupervised learning is often realized by regression analysis [3]. It relies on estimating the dependence between input and output parameters the latter of which are not discrete.

Decision trees [4] are the foundation of another model where the main feature is hierarchy of rules represented by connected branches. Each node of the tree is a test for a particular property of the data.

Classification and regression could be embedded together and one such approach that takes benefit of this incorporation is the k-Nearest Neighbors (k-NN) [5]. The output data are k in number closest samples. They are property of the object to a certain class, which is expressed more often among all the other k classes.

Support Vector Machines (SVM) [6] are another example of unsupervised learning algorithm. The idea laying behind this method is to find hyper-plane, which separates most accurately two classes of data. If the data clusters are not separable by linear function in n-dimensional space, a transition to (n+1)-dimensional space is made and then a search for a linear function is attempted.

Artificial Neural Networks (ANN) [7] are another example of self-adaptive structures that make unsupervised learning possible. They consist of a network of artificial neurons modeling the real neurons from the human brain. Interaction with the outside world relies on adaptation of input weights depending on expected output obtained through properly selected activation function.

An example of specialized neural network modelling the perception ability of the human visual system is the neocognitron [8]. There are input layer of cells imitating the operation of the photosensitive region from the human cortex followed by cascade of modular structures built from 2 types of cells – simple and complex. The first type react to particular (fixed) stripes of light while the second are sensitive to wider variety of linearly described patterns of illumination. There is a third type of cells as well – the hyper-complex one, which are capable of making analysis of geometric forms.

Hopfield networks [9] introduce bidirectional connections among neurons allowing repetitions. Asynchronous and synchronous mode of updating are possible depending on whether a single neuron, randomly selected according to a rule, or all neurons are going to be updated at once.

Deep learning [10] is the next point of the natural evolution of the classification techniques without supervisor. Since SVM is faster in execution that ANN until recently it was the main approach for practical implementations. With the growth of computing power and the introduction of GPUs (Graphical Processing Unit) for parallel processing ANN become attractive for wider use too. Deep Neural Networks (DNN) have higher levels of abstraction – multiple hidden layers that allow detection of properties going closer to the output. Thus, modelling of more complex data with fewer computations becomes possible.

Neural networks could be constructed with memory [11] using Long Short-Term Memory (LSTM) cells. There is resemblance to Hopfield networks for them. These cells

[1]Ivan Videv is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: noxicul@gmail.com.

[2]Ivo 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

memorize data for a given time interval. They have body, input, output and delay gates. Each gate may be modelled as a traditional artificial neuron calculating its output by an activation function through input weights. These networks reveal very good recognition results for speech and handwriting.

Convolutional Neural Networks (CNN) [12] use Deep Learning and multiple hidden layers are contained within its structure. They do not form cycles in contrast to the networks with repetitions. Their operational principle relies on minimal amount of preprocessing. A series of filters are applied directly over the image and properties are derived further used for classification.

In this paper a newly developed web-based application is presented for digital image classification using TensorFlow. The purpose of the study is to test the applicability of the environment to the average non-professional user running it from a personal Desktop computer. In Section II the system architecture and the client design are given. Experimental results are included in Section III followed by discussion in Section IV. Then, a conclusion is made in Section V.

II. IMPLEMENTATION DESCRIPTION

A. System architecture

The general architecture of the testing system is given in Fig. 1. The server part is run over Ubuntu Server v. 16.04. The main processing functionalities are included inside php script modules executed by the PHP engine at the requests of the user through the client side. It is done from a web browser via a communication back and forth with the web server – in this instance Apache. Uploading images for classification is made available by FileZilla ftp client to the vsftpd ftp server. All temporary data and the personal information of the users is contained inside the MySQL database.

TensorFlow [13] is essential part of the proposed architecture. It is a library, developed by Google Inc., for mathematical computations using graphs and tensors. Constructing different configurations for neural network implementations is one of its main applications. Some of the qualities it possess is the strong parallelization, multiple central and graphical processing units support, and dedicated API for data exchange with the system.

Inside the proposed implementation, the Inception-v3 model [14] of convolutional neural network is used (Fig. 2). A successor of the GoogleNet architecture it is preliminary trained into 1000 classes [15] on a dedicated server by the development team. Transfer learning method is applied for knowledge transfer from solving a particular problem to propose a solution to another one.

The initial training is accomplished with the ImageNet database and then the outer layer of the network is displaced. There are 42 layers inside with included factorization – the traditional 7x7 convolution from Inception-v1 is factorized to 3x3 convolutions. In addition, there is normalization of the supporting classifiers, which redound for the better convergence of the deeper layers of the neural network. Re-training only the outer layer of the model saves execution time because the necessity of retransforming the inner layers to conform new shapes becomes obsolete.

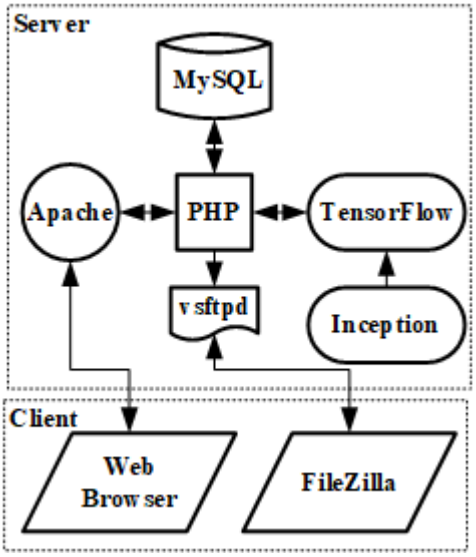
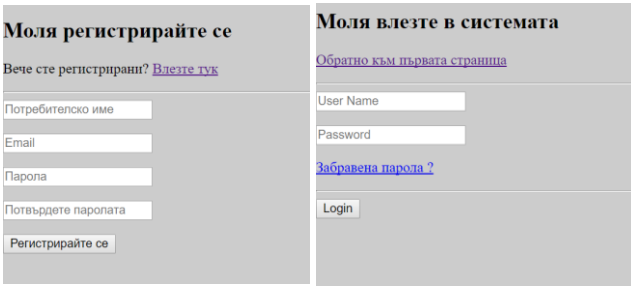


Fig. 1. Classification system architecture

Fig. 2. Inception-v3 structure, [14]

B.Client design

The graphical user interface for registering and admitting a client to the system is given in Fig. 3.



a) b)

Fig. 3. Registration panel a) and login panel of the client interface

Settings selection and initiating the training of the outer layer is placed over another panel of the client application (Fig. 4).

The following parameters could be defined:

- Number of training steps;
- Random crop size;
- Random intensity value;
- Random scale size;

- Enabling/Disabling a flip;
- Testing set size;
- Validation set size;
- Training rate;
- Training set size;
- Enabling/Disabling training over the full group;
- Enabling/Disabling displaying wrongly classified images.

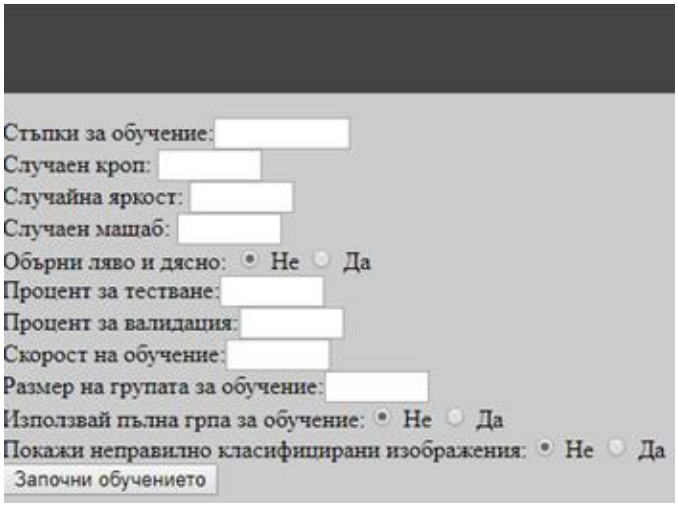


Fig. 4. Settings selection panel

III. Experimental results

Image test set consists of 3 groups of photos randomly selected from Google Images. They all contain fruits at different stages of ripening in various backgrounds:

- Grapes – 315 images;
- Cherry – 375 images;
- Pineapple – 82 images.

Grapes and cherry are very close in appearance and significantly different from pineapple so it is expected to have lower classification accuracy between the first two groups. Additional image of a car over ocean background is also included in the testing to evaluate the behavior of the system. Totally of ten test images (3 of grapes, 3 of pineapple, 3 of cherry and 1 of the car) are used not being a part of the training set. The hardware test platform is built with Intel Core i5-4440 3.1 GHz CPU with 4 cores, 2 GB RAM. In Fig. 5 are presented some of the training images from the three groups.

Fig. 5 Sample training images

Classification confidence for all 10 test images is presented in Fig. 6 at different training steps.

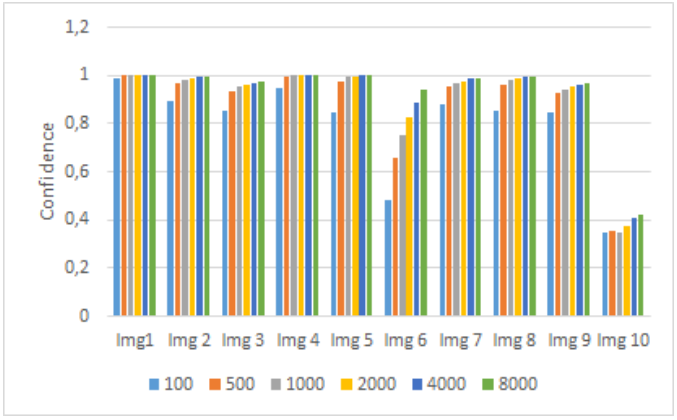


Fig. 6. Classification performance of the system

Initially 4000 steps are performed for the training phase. The execution time is 5 min 30 sec and the overall accuracy is 97.1 %. Comparison between the confidence levels with and without the "random crop", function is shown in Fig. 7. The random crop introduces random selection of the elements of the tensors describing the images, so a partial matching can be tried.

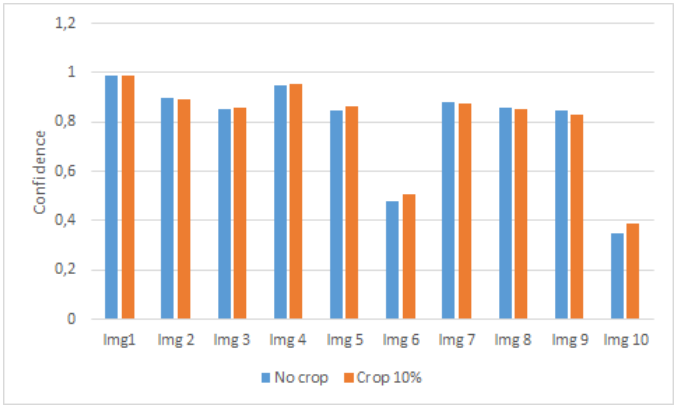


Fig. 7. Influence of the "random crop" function on the success rate of the system

There is one of the test images (Image 6, containing grapes) which returns slightly different results from the other images on average (Fig. 7 and 8).

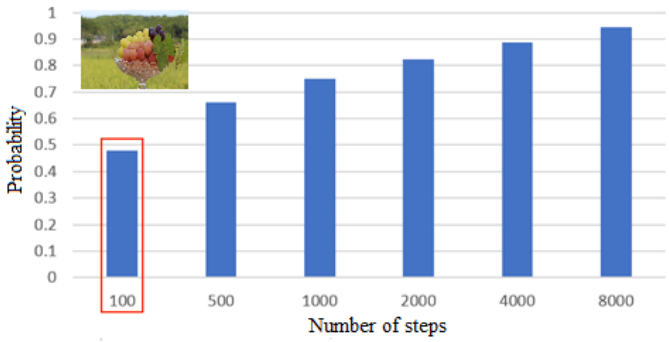


Fig. 8. Lower matching rate for Image 6

The probability for it is 47% at 100 steps, rising up to 66% at 500 steps. Probable cause for its lower matching rate is the mutual disposition with surrounding objects. These are the leaves and grass in the background with different texture and dominant size.

When objects of completely different type are passed to the system, such as the car from Fig. 9, the results are drastically changed. The probabilities for the three classes become - 40.5%, 31.1%, and 28.2%. Still the most probable match is with very low rate – below 50%.

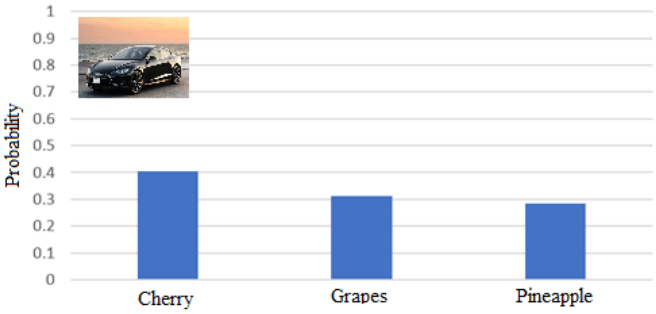


Fig. 9. Matching results for image with unfamiliar objects

IV. DISCUSSION

Efficient use of convolutional neural networks for visual classification purposes is a hard task because of the compromise that need to be done between accuracy of the model and the computational resource involved.

The experimental results show that the trained classifier copes well at the highest test level of 100 steps (iterations). Here is only one wrongly classified image – 6 of 9 passed to the input aside from the train set. The mismatch case possess very low confidence level of 0.47983. At 500 steps, the same image is correctly assigned with higher confidence of 0.66005. The execution time here is below 30 sec. Confidence comparable to that of the other test inputs – 0.94296 is achieved for 8000 steps for image 6. None of the other test images is categorized better in terms of confidence with the considerable rise of the number of steps – 0.95 is the typical value for 2000 steps.

Refining the model in 100 steps and 10 "random crop" takes almost an hour on the test platform. Tangible improvement has taken place with image 6 properly classified. Still, the results prove less accuracy from those obtained at 500 steps and without "random crop" where training takes 30 sec.

Adding noise to the images, as additional option, may be used to increase the accuracy even higher that is not possible only by incrementing the number of steps. TensorFlow supports Graphical Processing Units. In addition, it could be compiled with activated SSE4.1/SSE4.2/AVX/AVX2/FMA instructions sets.

In multiple cases, the accuracy of the model at 2000 steps achieved in 2 to 3 minutes of training is completely satisfactory. Common Desktop and mobile computers have the appropriate hardware for the purpose.

V. CONCLUSION

In this paper a web-based application for digital image classification is presented using TensorFlow. It proves to supply recognition rates enough for most practical cases. Need of proper training with account on the temporal accuracy at each iteration is a key phase of the whole tuning of the system. Execution times appear to be also satisfactory for the particular user working on a mid-class Desktop computer. Further development of the system suppose including a proper web-based interface for image category presentation, organizing topic based albums, etc.

REFERENCES

[1] J. A. Hartigan, M. A. Wong, "Algorithm AS 136: A k-Means Clustering Algorithm", Journal of the Royal Statistical Society, Series C (Applied Statistics), vol. 28, no. 1, pp. 100-108, 1979.

[2] J. J. Jenkins, W. A. Russell, "Associative Clustering during Recall", The Journal of Abnormal and Social Psychology, vol. 47, no. 4, pp. 818, 1952.

[3] S. Menard, *Applied Logistic Regression Analysis*, SAGE publications, 2018.

[4] S. R. Safavian, D. A. Landgrebe, "A Survey of Decision Tree Classifier Methodology", IEEE Transactions on Systems, Man, and Cybernetics, vol. 21, no. 3, pp. 660-674, 1991.

[5] S. A. Dudani, "The Distance-Weighted k-Nearest-Neighbor Rule", IEEE Transactions on Systems, Man, and Cybernetics, vol. 4, pp. 325-327, 1976.

[6] M. A. Hearst et al. "Support Vector Machines", IEEE Intelligent Systems and Their Applications, vol. 13, no. 4, pp. 18-28, 1998.

[7] R. J. Schalkoff, *Artificial neural networks*, New York, McGraw-Hill, 1997.

[8] K. Fukushima, S. Miyake, *Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Visual Pattern Recognition. In: Competition and Cooperation in Neural Nets*, Berlin, Heidelberg, Springer, pp. 267-285, 1982.

[9] J. J. Hopfield, "Neural Networks and Physical Systems with Emergent Collective Computational Abilities", Proceedings of the National Academy of Sciences, vol. 79, no. 8, pp. 2554-2558, 1982.

[10] Y. Lecun, Y. Bengio, G. Hinton, "Deep Learning", Nature, vol. 521, no. 7553, pp. 436, 2015.

[11] A. Graves, A.-R. Mohamed, G. Hinton, "Speech Recognition with Deep Recurrent Neural Networks", In Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2013), pp. 6645-6649, Vancouver, Canada, 2013.

[12] A. Krizhevsky, I. Sutskever, G. Hinton, "Imagenet Classification with Deep Convolutional Neural Networks", In: Advances in Neural Information Processing Systems, pp. 1097-1105, 2012.

[13] M. Abadi et al. "TensorFlow: A System for Large-Scale Machine Learning", In: OSDI, pp. 265-283, 2016.

[14] C. Szegedy et al. "Rethinking the Inception Architecture for Computer Vision", In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, p. 2818-2826, Seattle, WA, USA, 2016.

[15] https://github.com/WillKoehrsen/Machine-Learning-Projects/blob/master/pre%20trained%20models/imagenet_class_names.txt

COMPUTER SYSTEMS AND INTERNET TECHNOLOGIES – ORAL SESSION

Expert System for Prevention of Transborder Financial Violations

George Popov[1], Peter Yovchev[2] and Alexander Balevsky[3]

Abstract **–In this article we propose the establishment of automated expert system for the prevention of financial fraud , and offenses, acting on the territory of Bulgaria. Similar systems are used in many countries worldwide. In this presentation will be discussed some of the aspects of the current situation and will be proposed an approach for building a nationwide expert system which will take into account national specificities.**

Keywords – **contraband, expert system, tax system, smuggling, financial fraud.**

I. INTRODUCTION

A correctly functioning tax and customs administration and its related systems is crucial for the national security of any country. In this contemporary globalized world, there is a constant need for increasing the measures which would guarantee the effective work of the tax and customs administrations Upon entering the EU and expecting to join the Schengen group, Bulgaria has become an external border for the European community. An increase is expected in the interest of criminal organizations attempting illegal trans-border activities. The criminal organizations can afford the latest technology, and their illegal activities finance lucrative relationships with key decision-makers, experts, IT specialists [1,2,3,4].

International criminal cooperation makes detection and termination of such schemes even more difficult. The customs and tax administration is countering smuggling and contraband by:

- performing routine checks at border points;
- customs intelligence and investigations;
- special officers and informants in order to combat customs and tax fraud and violations;
- processing of customs declarations and company financial reports via the tax administration.

The main drawback of the above measures is that they are being done sporadically or during campaigns, following a signal or request, which means that:

- the offense is investigated after its completion, i.e. the revenue is received and the damages are done;
- collection of evidence is hindered, as attempts are made at cleaning company and personal records; it is difficult to establish accurate type and quantity of the goods etc.;

The processing is done by a person and the human factor is present, allowing for possibility of corruption, extortion, influence etc.

II. EXPERT SYSTEMS FOR DETECTING FINANCIAL AND CUSTOMS VIOLATIONS

Due to the electronic nature of the information relating to financial transactions and customs activities, expert systems for detection of trans border violations exist in many countries, These information systems use financial and customs data mining and generate possible customs violations and money fraud by searching for irregularities within certain quantities.

Such systems are difficult to port or even adapt in Bulgaria because of different:

- *legal base* – customs, financial etc.;
- *databases of different administrations* – varying formats of data gathering, data consistency. It is possible that a foreign expert system uses electronic data which is not available in Bulgaria;
- *quantitative data and its related indicators*. For example Bulgaria cannot be compared to the similarly sized Netherlands, nor with the dramatically different Brazil. Some indicators may be utilized with Romanian ones, albeit at a different ratio;

Systems that use data mining can make correlations, detect tendencies and anomalies within databases containing customs declarations, financial transfers and sales data from the tax administration. Various mathematical apparatus can be used for the knowledge gen-158

Decision support expert System for Prevention of financial Violations, including statistical methods, fuzzy logic, neural networks, image recognition, machine self-improvement, genetic algorithms, decision-trees, associative rules etc. A starter-class of similar systems can search for unusual values or anomalies in the data stream.

[1]George Popov is with the Faculty of Computer Science at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: popovg@tu-sofia.bg.

[2]Peter Yovchev is with the Faculty of Computer Science at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: fersina@mail.bg

[3]Alexandar Balevesky is from French faculty at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: alexandar.balevsky@gmail.com

III. PRECURSORY DATA ANALYSIS

In relation to all of the above, it is suggested that an integrated expert system for detection of financial, tax and customs violations is designed and implemented, which would process and analyze in real time the movement of documents, financial flows and goods. In order to accomplish that, a criterion system is needed which would include various indicators. Based on this system are the so called indicators for possible fraud. The following data is processed: goods, payments, supporting documents, dates, companies, and other information including banks, shipping companies info, warehouses, customs intelligence data and also wire-tapping. Movement of the goods . Goods and accompanying documents are verified. There are standard routes for international shipping. The following data is examined:

Trade chains:

- producer – wholesaler – seller – client;
- producer – wholesaler – producer;
- producer – wholesaler – exporter;
- importer wholesaler – seller – client; Searching for data anomalies within:
- *quantity* – sudden increase or decrease in operating balances, import/export levels or sales. Every deviation (anomaly) is an indicator for unusual situation and must be investigated further;
- *no economic logic* – an example could be excessive shipping charges or unusual routing, misleading value, re-exportation of the capitals and goods to the country of origin or to a country where similar goods are cheaper, cyclic movement of money and goods.

Document processing

A search for the following is done:

- *relations to compromised companies* – including companies owned or run by people/shareholders which have a history of customs fraud;
- *double accounting* – used by the companies in order to conceal tax and customs fraud;
- *forged or fake documents* – companies are forging documents related to goods' origin and quality of services.

Finance and cash-flow

The aim is to analyze cash-flows in order to detect anomaly (irregularity) within the flow of goods and documents. The following is processed:

- *nominal wire transfers* – related to companies' attempts at covering customs or tax fraud. Usually somebody makes unusual and large bank withdrawals in cash from a company's account, international transactions or unusually high-priced consulting services. Usually exact shipping or transport data is missing;
- *inconsistency between cash and goods flows* –such behavior usually indicates selling of contraband goods and/or goods for which state collectables are owed (VAT, excise, duty etc.). Companies and shareholders

In order to cover large transactions or deals, a set of companies is used or new companies are being registered. Analyzing companies and related persons can point out in the direction of possible risk company or behavior.

Additional information

Additional information may be obtained from the following sources:

- *intelligence information* – mobile data interception, GPRS, internet data, bank statements and transactions, wire-tapping. "Mole-searching"
- introducing various disturbances in order to observe the behavior;
- *movement of vehicles* – processing GPS data, border points data, traffic cameras.

IV. STRUCTURE OF THE SYSTEMS

A structure of the proposed system is given in [4]. Information is received via different administrative databases i.e. tax administration, customs, statistics and special sources (DB0,.., DBn). Different databases are integrated through Common Database API. The normalized data is transformed into metadata and organized into Multidimensional cube with its own API. The specialized On-line Analytical Processing (OLAP) engine is looking for correlations pre-set by the user via user interface.

CONCLUSION

The benefits of implementing such a system are obvious:

- Eliminating the human factor in investigating customs violations and disseminating valid and timely information to the competent authorities.
- Reducing the risk and preventing of trans border illegal financial transactions and fraud
- Monitoring and possibly eliminating or reducing the illegal movement of goods, money and give the information to the customs, tax authorities and the police.
- reduce of the risk of importing /and consuming/ potentially harmful goods and/or foodstuff;
- it is possible to design a mobile app along the user interface in order to notify regional/local

REFERENCES

[1] The ACFE's 2012 Report to the Nations on Occupational Fraud and Abuse, wwwacfe.com

[2] T. Bezlov, P. Gounev, G. Petrunov, E. Tzenkov and M. Tzvetkova (2007). Organised Crime in Bulgaria: Markets and Trends. Sofia: Center for the Study of Democracy.

[3] Open Society Institute Bulgaria: Schengen: Effects for Bulgaria, Sofia, 2008, ISBN 978-954-9828-63-4 (translated form Bulgarian language)

[4] Popov G., Ia. Dokov, Automated Expert System for Prevention of Custom Violations, Challenges in Higher Education & Research, vol. 12, Heron Press, Sofia, 2014

Research and Solutions for DDoS Detection and Mitigation with Software Defined Networks

Branislav Mladenov[1]

.

Abstract – **In this paper several Distributed Denial of Service (DDoS) detection and mitigation solutions are provided. The detection is handled by Software defined network (SDN) network and the SDN controller takes actions so the malicious traffic to be filtered or isolated.**

Keywords – **SDN, DDoS attack, DDoS mitigation, Openflow**

I. INTRODUCTION

Software defined network (SDN) is one of the most promising architecture that decouples control plane from data plane and centralize network management. Comparing with legacy networking it has many more advantages and provides flexibility and scalability of the network. The centralized SDN controller can easily detect and protect the network from many vulnerabilities and security problems that threats traditional networking. Based on its Application Programming Interface(API) and programmability it can take proper actions and defend the network in case of Distributed Denial of Service (DDoS) attacks. There are many solutions that can detect DDoS attacks and takes some actions but unfortunate even the attack is detected most common problem is that during the attack the internet service provider channel gets full so the access is blocked until the attack stops. Distributed Denial of Service (DDoS) attacks are one of the most challenging security threats today. Flooding corporate networks or web sites with huge amount of malicious traffic block their services. The main goal is to exhaust network or computer resources in order legitimate traffic not to reach its destination. SDN architecture provides many benefits with decoupling the control plane from data plane on one hand, but on the other hand we have a centralized infrastructure that can be attacked so whole network will be affected. In this article, solutions for fast detection on control plane or data plane level and taking counter measure for blocking the attack are proposed.

II. DDOS TYPE OF ATTACKS.

Denial of service attack method is most commonly used for service availability degradation of the target. There are several types of attacks and can be classified based on Open Systems Interconnection (OSI) model as: Layer 3 flood attacks, Layer 4 Transmission Control Protocol (TCP) state exhaustion attacks and Layer 7 application attacks [1].

Layer 7 application attacks are exploiting application level protocols like Hypertext Transfer Protocol (HTTP) or Simple Mail Transfer Protocol (SMTP) and they are on 24 % of DDoS attacks. These kinds of attacks are usually harder to detect and protect because the amount of traffic is not much more that the legitimate traffic. Layer 4 TCP attacks target to exhaust the resources on TCP level and takes round 20% of DDoS attacks. Both methods are not based on the volume of traffic rather on the combination of traffic that can affect the protocols. Few of the most dangerous DDoS attacks these days are: DNS torques water attack: flood attack targeting company DNS server with many lookup requests to consume DNS server resources; Secure Sockets Layer (SSL)-based DDoS attacks, they can be Hypertext Transfer Protocol Secure (HTTPS) flood or encrypted SYN flood attacks; permanent DDoS attack which damages the systems hard enough to break the hardware.
Layer 3 attacks are the most common attacks. The idea is to send a huge amount of User Datagram Protocol (UDP) or Internet Control Message Protocol (ICMP) traffic from different sources facing the targeted host [2]. There is a huge increase of these kinds of attacks due to the increase of Internet of things (IoT) devices in 2017. There were 91% more attacks than 2016 [3] Today it is much easier to hack smart household devices like smart television, fridge or Closed-circuit television (CCTV) because most of them have never been updated with the latest Operating System version so their vulnerabilities have not been fixed.

III. DDOS PROTECTION MECHANISM

There are many articles and researches for DDoS detection and mitigation. In general, there are two methods of detection [4]. First one is based on some well-known predefined protocol patterns. The second one is an anomaly based method which collects statistics of regular traffic and analyzes all the time. For both methods all incoming traffic is checked and if a deviation is detected, proper actions can be taken. Most commonly used devices for DDoS detections are web application firewalls (WAF) and Intrusion prevention systems (IPSs). WAFs are focused to detect layer 7 type of attacks while the IPS can detect all kind of protocols or patterns and try to block the malicious traffic while the legitimate traffic is passed. To learn which kind of traffic is legitimate and which is not the devices should be switched to a learning mode for several weeks before it is runt

[1] Branislav Mladenov is with the Faculty of Telecommunications at Technical University of Sofia 8 Kl. Ohridski Blvd, Sofia 1000 Bulgaria, E-mail: branislav.mladenov@gmail.com

into production. IPSs are smart devices and can block malicious traffic but unfortunately once the system detects the attack it is already too late because the IPS is deployed within customer's network and when the attack starts the whole internet bandwidth is allocated with malicious traffic and customer's services is totally blocked until the attack stops. Hosting companies are suffering from DDoS attacks because even only one of the customers has been attacked, all other customers are affected as well due to a reason that they share same internet service channel.

IV. Security solution with SDN

Software defined network has many benefits than traditional networking and some of them can be successfully used for DDoS protection and mitigation [5]. Separation the control plane from data plane makes the controller really easy to detect and react to DDoS attack very quick and efficient. With its global flow monitoring, the centralized controller can see and analyze the attack which comes from different sources and enters from different locations of the network. The controller can communicate with external IPS applications via its southbound interface and to provide summarized information about incoming traffic so prevention systems will analyze and take a decision which flows are part of the attack and which are legitimate traffic. SDN controller can rapidly and precisely creates policies based on the information provided from IPS that can be applied on the edge of the network. With its programmable interface the controller can communicate with other controllers within the Autonomous system (AS) or with other controllers that manage networks closer to attack sources or Tier 3 internet service provider can send information to its Tier 2 services provider when the attack is detected. The centralized controller is able dynamically and quickly to add or delete rules that can speed up DDoS mitigation. Comparing with the traditional network all features of SDN controller reduces the cost, the speed of reaction and the lack of human mistakes.

V. Security issues with SDN and DDOS

SDN network has many benefits that make network architecture more scalable, more secured and reduce cost it comes with its own disadvantages. SDN can be attacked on application layer, control layer or infrastructure layer [6]. All type of attacks targets resource exhaustion of SND infrastructure. Decoupling the control plane from data plane provides many features but in case of DDoS attack, the centralized SDN controller becomes a target because requests that will arrive on the edge switches will be new for them and not part of the switches Ternary Content Addressable Memory (TCAM) table so according to Openflow protocol version 1.3 the switch buffers the packet and ask the controller how to proceed with it [7]. During Infrastructure type of DDoS attack all malicious request will create a *packet_in* message that will be sent to the controller Fig.1. This can load the secured channel between the switch and the controller so no legitimate requests can be handled. On the other hand, the switch has a buffer with limited resources so in case of an attack it might be overloaded so again no new legitimate request can be processed [8]. SDN switch TCAM table has limited resources so during the attack it can be quickly filled up with malicious records. Control layer DDoS attacks are targeting the controller that has limited resources as well and if it has to process a huge number of requests its Central Processing Unit (CPU) and memory might be exhausted very fast. The communication between SDN controller and all switches is secured and encrypted which allocate much more resources for encryption and decryption of the traffic. Application DDoS attacks target the applications used by SDN controller and northbound API interface. Fig.1

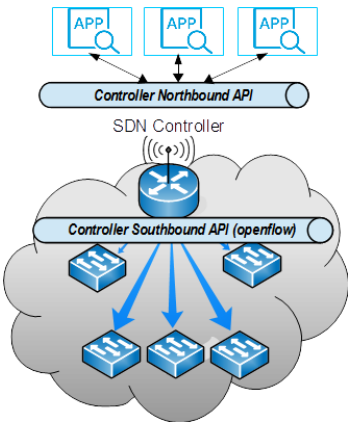


Fig1. SDN architecture.

There are many proposed solutions that can detect a DDoS attack and protect SDN network. One of the proposals is based on sFlow protocol [9]. sFlow agent can be installed on each switch it can collect flow statistics and send them to sFlow collector which can analyze them. The sFlow collector can summarize and create rules for malicious traffic flows that can be configured from SDN controller. In order to protect the channel between switches and controller rate limits can be configured on that link, so in case of attack the channel and the controller will not be overloaded with malicious requests. SDN controller can detect DDoS attack very quickly if thresholds are set on its resources or based on a number of *packet_in* requests received per seconds.

VI. Proposed solutions

In this section, several solutions for DDoS detection and mitigation are proposed.

A. *First solution.*

This solution proposes automatic DDoS attack detection in customer's network with automatic internet service provider notification. The notification can be provided through dedicated

for this purpose virtual local area network (VLAN) and Border Gateway Protocol (BGP) session. Once the ISP is notified it can forward the whole customer's traffic through its DDoS mitigation system to filter the malicious requests. Internet service provider's (ISP's) DDoS mitigation service is expensive so customers should use it only when there is an attack.

Based on the analysis in the previous section a DDoS attack can be detected fast either from the switches with sFlow protocol, or when the controller's thresholds are reached. The controller can automatically send a message to Internet service provider application which can activate its DDoS mitigation system. Fig.2

In traditional networking these functions are manual and the mitigation depends on the engineer on duty reactions. He or she has to be notified via monitoring system which may take 1 to 3 minutes and he or she should manually trigger the DDoS protections which may take 5 to 10 minutes. The proposed SDN solution has several benefits:

- Time for reaction. The time for reaction between the start of the attack and its mitigation is minimized because it is fully automatic. The detection can take up to 10 seconds and automatic notification less than a second.
- There are no manual actions which prevent human mistakes.
- Reduce cost. There is no need of engineer who monitors manual DDoS monitoring system. There is no need for additional DDoS monitoring system.
- Reduce cost of ISP's DDoS mitigation service because it is used during the attack only.

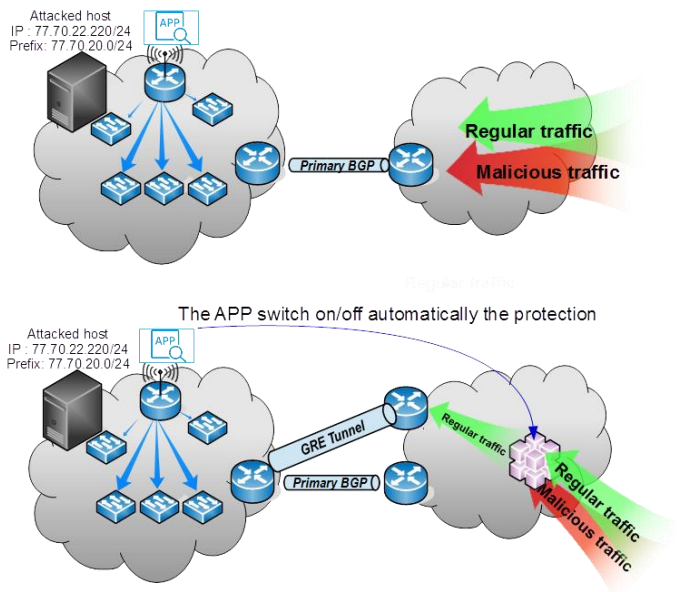


Fig.2 DDoS attack detection and notification.

B. Second solution.

This solution proposes automatic DDoS detection and changes traffic flow of attacked target. Hosting companies and cloud providers can be heavily impacted by the DDoS attack because all customers share one and the same internet service in most of the cases. The bandwidth is enough for legitimate traffic but in case of DDoS attack targeting only one customer, all other customers are affected when the channel gets full. If hosting company or cloud provider doesn't have DDoS protection provided from the ISP it is almost impossible to stop the attack before it affects the internet service channel. In most of the cases cloud providers have more than one channel for internet service so once the attack is detected and the target is identified, its prefix can be advertised via the second/backup internet service channel. This will not stop the attack but at least all other customers will not be affected anymore.

DDoS detection with SDN can be achieved very fast and based on the sFlow analysis the target can be identified in seconds. Once this is done the controller can automatically isolate the target via backup internet service provider link or can send commands to edge router that will advertise the targeted prefix via the secondary service provider link. Fig.3

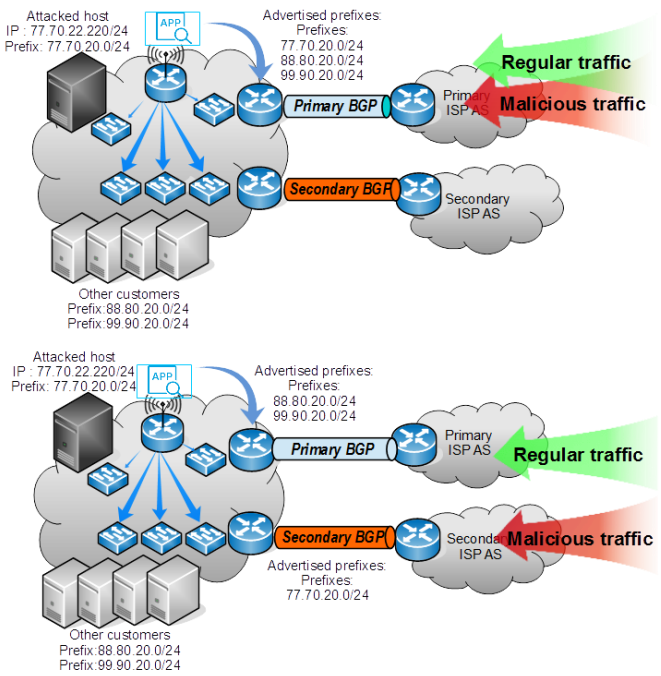


Fig.3. Advertising attacked prefix via secondary ISP.

In traditional networking, these activities can take more than 15 to 30 minutes and depends of manual actions of engineer on duty and its knowledge.

The proposed SDN solution has several benefits:

- Time for reaction. The time for reaction between the start of the attack and targeted destination traffic flow change is minimized because it is automatic. The detection can take up to 10 seconds and flow reroute can take less than a second.
- There are no manual actions which prevent human mistakes.
- Reduce cost. There is no need of engineer who manually has to change the flow of the targeted prefix.

C. Third solution

It proposes DDoS mitigation that blocks the traffic as close as possible to the source of the attack. As we explained in the previous section the switches can identify the target and the sources via sFlow protocol. This information can be sent to a sFlow collector that can create a database of the affected sources which can be used for following purposes Fig.4:

- Flow blocking rules can be created based on the database. The rules can be used by other SDN controllers that are closer to the sources. This can stop the malicious traffic before it hits the destination network and will not load all service providers' networks between the source and target.
- It might be the case that malicious sources are not part of any SDN network but just a traditional network with next generation firewalls that have API interface. The application can track all new records of the database and create blocking rules that which be applied automatically to the firewall in an outbound direction so the malicious traffic will not exit its network.
- The database can be used anytime during an attack or it can be analyzed and if some of the sources are used several times for different DDoS attacks it can be marked as malicious source. Once such malicious sources are identified in the database they can be blocked permanently by internet service providers.

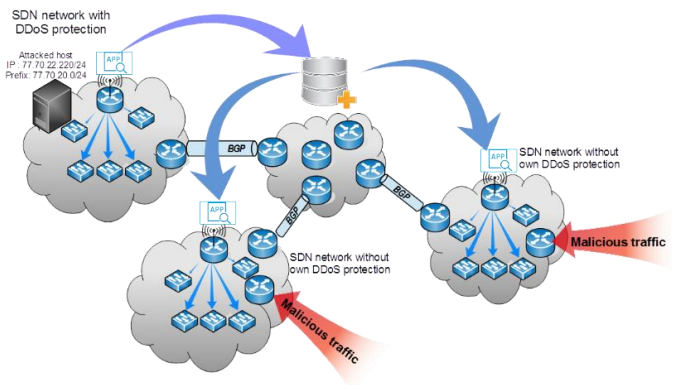


Fig.4 DDoS mitigation with malicious sources data base.

VII. CONCLUSION

Denial of service attack is still one of the biggest internet security challenges these days. In this paper, we discussed what type of DDoS attacks are most commonly used and types of protection mechanisms.

With this paper, several solutions that can help to detect DDoS attacks faster than traditional networks are proposed. With SDN controller programmable interface, DDoS mitigation can be achieved in a second and can be blocked even closer to the source which will stop malicious traffic. All proposed solutions can reduce cost and network resources. On another hand SDN networks are still vulnerable to DDoS attacks but counter measures can be taken to protect the controller. The essential part of this is fast and accurate attack detection so with proper monitoring and applications companies can protect their network and services without affecting legitimate traffic. This paper provides just few SDN solutions that can be part of much bigger and efficient DDoS protection architecture.

REFERENCES

[1] M.Jeyaraman, R.Saravanaprabu "A Survey and New Perspectives on Classifying the DDOS Attack with Their Characteristics" International Journal on Recent and Innovation Trends in Computing and Communication , IJRITCC | August 2016, Available

[2] N. Dayal, P. Maity, S.Srivastava, R. Khondoker "Research trends in Security and DdoS" . SecurityComm.Networks2016; 9:6386–6411 . Published online 9 February 2017 in Wiley Online Library (wileyonlinelibrary.com).

[3] P.Lee, D. Stewart, "Global TMT Predictions 2017 Analysis" , August 2016

[4] S. Luo, Jun Wu, B. Pei. "A Defense Mechanism for Distributed Denial of Service Attack in Software-Defined Networks". Frontier of Computer Science and Technology (FCST), 2015 Ninth International Conference

[5] W. Xia, Y. Wen, C. Foh, D. Niyato, and H. Xie, "A survey on softwaredefined networking," IEEE Commun. Surveys & Tutorials, vol. 17, no. 1, pp. 27–51, First Quarter 2015.

[6] P. Porras, S. Shin, V. Yegneswaran, M. Fong, M. Tyson, and G. Gu, "A security enforcement kernel for openflow networks," in Proc. First Workshop on Hot Topics in Software Defined Networks, 2012, pp. 121– 126.

[7] M. D. Yosr Jarraya, Taous Madi, "A survey and a layered taxonomy of software-defined networking," IEEE Commun. Surveys & Tutorials, vol. 16, no. 4, pp. 1955–1980, Fourth Quarter 2014.

[8] F. Hu, Q. Hao, and K. Bao, "A survey on software-defined network (SDN) and openflow: From concept to implementation," IEEE Commun. Surveys & Tutorials, vol. 16, no. 4, Fourth Quarter 2014.

[9] C. YuHunag, T. MinChi, C. YaoTing, C. YuChieh, and C. YanRen, "A novel design for future on-demand service and security," in Proc. 12th IEEE Int'l Conf. Commun. Technology, 2010, pp. 385–388.

Two-finger touch interface activation method for wearable devices

Yuri Dimitrov[1]

Abstract – **Wearable device interfaces are challenging due to the devices' small factors and their rich features. Research in this paper proves that it is possible for a wearable device computer-human interface to be activated based on recognition of a two-finger touch on the device surface which is around the display (bezel, rim) while the user's fingers don't hinder the user's sight on the device display.**

Keywords – **Wearable, Smartwatch, Touch, Interface**

I. INTRODUCTION

The wearable devices and in particular smartwatches and fitness trackers are becoming more and more popular during the last few years [1]. Despite this fact the methods used for interaction between the user and the device remain the same as those used typically in the classic electronic watches - push buttons as well as in the smartphones - touch screen based interfaces. Touch based interfaces, being efficient and intuitive used in smartphones, when performed over a small surface like a smartwatch display, often face the common "fat finger" problem [2] - due to the small target icons and other interface controls the user can't aim and push them correctly due to the line of sight obstacle caused by their own finger. There is also a common issue with the confirmation of the taken action.

Trying to solve these issues and to make their devices more usable some of the leading vendors in the smartwatch area have implemented in their recent models, different non-touch display and non-push button interfaces. These are ones like rotating digital crown [3], rotating bezel [4], touch sensitive rim around the display [5]. All of them are addressed to solve the above mentioned fat finger and user experience problems but the common disadvantage is that an interface pre-usage action is required (push a button in all examples above) in order for these interfaces to be activated. Due to the high power consumption and limited battery size/capacity wearable devices typically work in idle or power save mode and an intended action should be performed such as pushing a button or specific device movement (acceleration) in order for the device interface (access to the device menu) to be activated. That is why we consider that the methods of alternative interfacing with wearable devices and specifically those which allow instant interface activation have to be further researched.

II. RELATED WORK

The touch based interaction with wearable devices has been researched by many authors. Some of them have made research in areas of evaluating touch or pressure technologies on the device components other than on the display such as the devices' sides [6], band/strap [7], back [8] or changing the devices' position according to the band [9]. Others have explored the area and surfaces around the device like touches on user's skin close to the wearable device [10] or using touch sensitive fabrics close to the device [11]. There is also research about replacing standard touch interfaces with those based on taps [12] and [13].

The most closely related research [14] evaluates the interaction at the edge/outer side of round wearable devices. The major differences between the mentioned research and ours are: the sensors' position place - while on the mentioned model they are on the edge, in our model they are on the bezel which is a much more accessible part of smartwatch devices which are worn on the wrist; the experimental model size - our one is about 30% smaller (42 mm in diameter against 60 mm) and the same as most of the current smartwatches; two-finger multi-touch positions had never been researched nor evaluated during the above mentioned research.

In order for a better approach to be found for wearable device interface activation, this paper presents two Experiments where

Fig. 1. a). 3D Model for Experiment 1; Fig. 1. b). Compare the 3D Model and other smartwatches; Fig. 1. c). Experiment 1 process

[1]Yuri Dimitrov is PhD student in Computer Science Department at Technical University of Varna, 1, Studentska Str., Varna 9010, Bulgaria, E-mail: yndimitrov@gmail.com

the users interactions with wearable device models are researched and analyzed.

III. EXPERIMENT 1

The goal of the Experiment 1 is to prove that two finger touches in two separate and distinguished areas on the device bezel is possible to be registered in order for the device interface to be activated and further interface actions to be performed. In order the goal of the Experiment 1 to be achieved it is need to be proven that, during the two finger touch over the devise bezel, there are enough bigger non-touched spaces (the second and the forth in the sequence "Thumb finger (touched)", "Thumb finger to Other finger (non-touched)", "Other Finger (touched)", "Other Finger to Thumb finger (non-touched)") remained between the two touches. As enough bigger space will be counted a space with 2 or more non-touched hour marks.

Equipment:

In order for the touch areas on the bezel to be observed, a dedicated 3D model was designed and printed (Fig. 1. a).). The aim was the 3D model to be as close as possible to a real watch/smartwatch (Fig 1. b).). The dimensions of the 3D model are - diameter: 44 mm, height: 15 mm, bezel angle: 45^0, bezel width: 4 mm; color: white; material: PLA. As the aim of the model is bezel touch activities to be recorded no other standard watch/smartwatch interfaces like buttons or crown were added to the model. For more precise observation and the fingers touch areas recording 12 engraved lines were made - one per each 12-houred analogue watch hour mark. The model had lugs for standard 18 mm watch strap to be used.

Group:

The test group for the Experiment 1 consisted of 12 people (male 10, female 2), aged 25-47 (Average 36.5), all right handed, without disabilities, all volunteers.

The process:

The following instructions were given to each participant: To wear the model like an ordinary watch on his left hand and to stand in a relaxed position. After that to raise his/her left hand in a position to see the model's virtual display. He/she was instructed that the only way to interact with the device is to touch the bezel with two fingers. Finally he/she was asked to put his/her leading hand (the right one) thumb and one of the other fingers over the bezel in a way to be able to also see the model's virtual display (Fig. 1. c).). The touched by user fingers engraved hour mark lines were manually recorded in MS Excel spreadsheet. Each participant made three consequential attempts and the data from all three attempts was recorded (there was no try attempt). After each attempt the participant was instructed to go back to a relaxed standing position.

Results:

During the experiment process, after half of the attempts were made (6 from 12), it was decided that the visual measurement process is not precise enough and the data wouldn't be reliable for the purposes of the experiment. Due to this decision the experiment was cancelled and it was decided to proceed directly to Experiment 2.

IV. EXPERIMENT 2

The goal of the Experiment 2 is the same like as Experiment 1 but using a much more precise way of measurement - recording touch areas via electronic touch sensors mounted into the device. For the purposes of the goal of the Experiment 2 as enough bigger non-touched space will be counted a space with 2 or more non-touched sensors (i.e. two areas with 2 or more non-touched sensors from 12 sensors in total).

Equipment:

For the Experiment 2 a new 3D model was designed and printed (Fig 2. a.).) - the same as the one used in Experiment 1 with the following differences - engraved hour mark lines were removed and the model was hollow with an open back in order to access the sensors and their wirings provided. On the bezel of the model 12 capacitive sensors (3x10 mm conductive metal metric screws with 6 mm in diameter heads) were mounted (Fig 2. b). The distance between each two sensors was also 6 mm. The experimental model was driven by an Arduino Mega 2560 computer with a 12 capacitive ports MPR121 extension board connected to it. The 12 capacitive sensors were connected to the MPR121 extension board sensors' ports. To each sensor was assigned an ID from 1 to 12 corresponding to the 12-hour standard watch hour mark place of each sensor. Specially developed for the research purposes software in the Arduino language was used for the experiment. The software detected the starts (touching contact) and the ends (releasing) of the touches made by the participant's fingers independently for

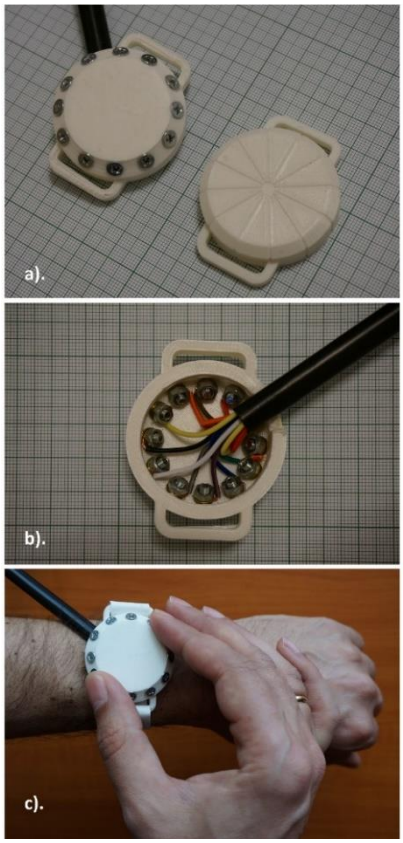


Fig. 2. a). 3D Model for Experiment 2 (left) compared to the 3D Model for the Experiment 1; Fig. 2. b). Experiment 2 3D Model sensors wiring; Fig. 2. c). Experiment 2 process

each sensor. The Arduino Mega 2560 computer was connected to a notebook via USB interface and sent the data to the Arduino IDE Serial Monitor. The raw data was then transferred to MS Excel for further processing.

Group:

The test group consisted of 20 people (male 16, female 4), aged 25-49 (average 36), all right handed, without disabilities, with clear and dry hands, without any other worn watches or hand jewelries, all volunteers.

The process:

Each participant was given the same instructions as during the Experiment 1 process. The experiment was held in office conditions. The IDs of the touched sensors were recorded in order the two finger positions to be precisely determined. Each participant made three consequential attempts and the data from all three attempts was recorded (Fig. 2. c).).

Results:

All 60 measurements which were made using the specially designed and printed 3D model for the Experiment 2, were successful and the raw data is presented in Table I.

TABLE I
EXPERIMENT 2 RAW DATA

Att. No.	Sensor ID (X - touched; Blank - non-touched											
	1	2	3	4	5	6	7	8	9	10	11	12
1	X	X					X					
2	X	X					X					
3	X	X					X					
4	X	X	X				X	X	X			
5	X	X					X	X				
6	X	X					X	X	X			
7	X	X					X	X				
8	X	X					X	X				
9	X	X					X	X				
10	X						X					
11							X	X				X
12	X						X	X				
13				X	X						X	
14					X						X	X
15				X	X					X	X	
16		X					X					
17	X	X					X					
18	X	X					X	X				X
19	X	X				X	X	X				
20	X	X				X	X					
21	X	X					X					
22	X	X					X	X				
23		X	X				X	X				
24		X	X				X	X				
25		X					X	X				
26	X	X					X					
27	X	X	X				X	X	X			
28	X	X				X	X	X				
29	X	X				X	X	X				
30	X	X				X	X	X				
31	X	X				X	X					
32	X	X				X	X	X				
33	X	X				X	X	X				
34	X	X	X				X	X				
35		X	X				X	X				
36	X	X					X	X				
37	X					X						X
38						X						X
39	X					X	X					X
40	X					X						X
41	X	X				X	X					
42						X						X
43	X					X	X					
44					X	X						X
45	X					X	X					
46	X	X					X	X				
47	X	X	X				X	X				
48	X	X					X	X				
49	X	X					X	X				
50	X	X					X	X				
51	X						X					X
52		X					X					
53	X					X	X					X
54		X	X				X	X				
55	X	X					X	X				
56	X	X					X	X				
57	X	X					X	X				
58	X	X					X	X				
59	X	X				X	X					
60	X	X				X	X					

The summarized data is presented in Table II where are shown the Average number of touched/non-touched sensors, Minimum and maximum number of touched and non-touched sensors, and Median of touched/non-touched sensors, per each of the two touched (by thumb and other finger) and the two non-touched (between the fingers touches) areas. The direction determining non-touched areas in columns 2 and 4 is clockwise.

TABLE II
EXPERIMENT 2 SUMMARIZED DATA

Sensors numbers	Thumb finger (touched)	Thumb finger to Other finger (non-touched)	Other Finger (touched)	Other Finger to Thumb finger (non-touched)
	1	2	3	4
Average number	1.90	4.42	1.88	3.77
Minimum number	1	3	1	3
Maximum number	3	6	3	5
Median value	2	4	2	4

In all of the 60 measurements the participants activated two touch areas (averages 1.90 for thumb activated sensors and 1.88 for the other finger activated sensors and medians of 2). No case with less than 3 non-touched sensors between both touch areas was recorded (Table II, Row “Minimum number”, columns 2 and 4 values). The symmetrical results for both touched areas and both non-touched areas says that very probably the similar results would be obtained if a research for people with left leading hand is performed.

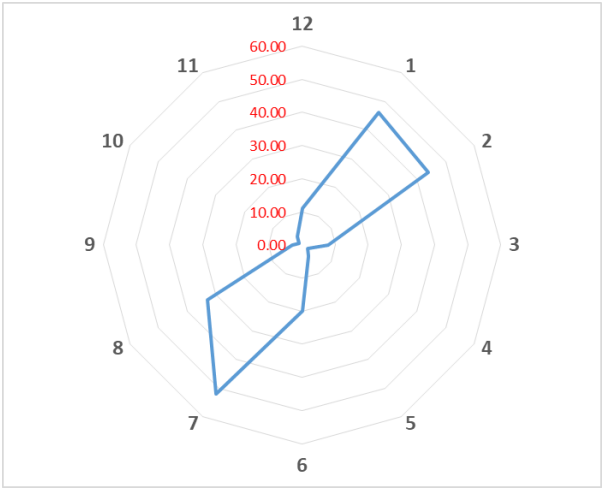


Fig. 3. Radar chart of the number of touches per sensor in all 60 measurements

The radar chart showing the numbers of touches per each touch sensitive sensor of the 3D wearable device/smartwatch model used in Experiment 2 during all 60 measurements is shown on Fig. 3.

V. Future work

As the goal of this paper is only to prove if a device interface could be activated intentionally by the user, following the results published in the paper, a research evaluating which kind of computer-human interfaces could be activated, executed and how they could be implemented in the process of wearable device input would be made in the best way and further results would be published.

The experiment model could also be improved in hardware aspect - capacitive board to be embedded into the device in order for wider range of experiments to be done. Measurement and recording Arduino software could be additionally developed for two-finger touch commands recognition.

A further investigation could be performed to evaluate if it is possible for the user pose (standing, sitting, laying) to be determined based on the activation of the different multi-touch areas over the bezel of smartwatch when the user is in a different pose.

The two-finger touch interface activation could be evaluated also for other wearable devices as well as for touch interaction on other touch sensitive surfaces.

VI. Conclusion

During the research described in this paper two experiments were made. The goal of the experiments was to be proven that it is possible to determine an intended two-finger touch on touch sensitive surface around the device display - the wearable device (smartwatch) bezel.

In order for the research goal to be achieved the users' interaction process with models of wearable devices was studied. For these purposes two special 3D models of smartwatches were designed and printed - a mechanical model for the Experiment 1 and electronic model with touch sensitive bezel for the Experiment 2.

For the Experiment 2 a dedicated experimental set of Arduino Mega 2560 computer and MPR121 capacitive sensors board was built and specialized Arduino software was developed.

The Experiment 2 results clearly prove that two-finger touch on the bezel of rounded wearable device with standard watch size could be unambiguously determine in 100% of the users' attempts. It is possible based on two-finger touch (two point multi-touch) over the wearable device (smartwatch) touch sensitive bezel the human-computer device interface to be activated (the device to be waked up from its' idle mode) and order action or set of actions to be performed such as menu navigation, menu items selection, confirmation, cancellation, information input etc. In a respective way it is possible for the interface to be deactivated (the device to go back to idle mode) based on the recognition of the full release of the touched areas or a release of a specific area.

Acknowledgement

Thanks to Zhivko Kulev for the 3D Models designing and printing.

References

[1] IDC, "International Data Corporation (IDC) Worldwide Quarterly Wearable Device Tracker - Q3 2017", https://www.idc.com/getdoc.jsp?containerId=prUS432, last visited on March 6th 2018.

[2] K. A. Siek, Y. Rogers, and K. H. Connelly, "Fat Finger Worries: How Older and Younger Users Physically Interact with PDAs" pp. 267–280, 2005.

[3] Apple Inc., "Apple Watch", https://www.apple.com/watch/, last visited on March 6th 2018.

[4] Samsung Electronics, "Gear S3", http://www.samsung.com/uk/wearables/gear-s3/, last visited on March 6th 2018.

[5] Misfit, "Misfit Vapor smartwatch", https://misfit.com/misfit-vapor, last visited on March 6th 2018.

[6] R. Darbar, P. K. Sen, and D. Samanta, "PressTact: Side Pressure-Based Input for Smartwatch Interaction" *Proc. 2016 CHI Conf. Ext. Abstr. Hum. Factors Comput. Syst. - CHI EA '16*, pp. 2431–2438, 2016.

[7] Y. Ahn, S. Hwang, H. Yoon, J. Gim, and J. Ryu, "BandSense: Pressure-sensitive Multi-touch Interaction on a Wristband" *Proc. 33rd Annu. ACM Conf. Ext. Abstr. Hum. Factors Comput. Syst. - CHI EA '15*, pp. 251–254, 2015.

[8] P. Baudisch and G. Chu, "Back-of-device interaction allows creating very small touch devices" *Proc. 27th Int. Conf. Hum. factors Comput. Syst. - CHI 09*, no. c, p. 1923, 2009.

[9] R. Xiao, G. Laput, and C. Harrison, "Expanding the input expressivity of smartwatches with mechanical pan, twist, tilt and click" *Proc. 32nd Annu. ACM Conf. Hum. factors Comput. Syst. - CHI '14*, pp. 193–196, 2014.

[10] C. Zhang *et al.*, "TapSkin: Recognizing On-Skin Input for Smartwatches" *Proc. 2016 ACM Interact. Surfaces Spaces - ISS '16*, pp. 13–22, 2016.

[11] S. Schneegass and A. Voit, "GestureSleeve: using touch sensitive fabrics for gestural input on the forearm for controlling smartwatches" *Proc. 2016 ACM Int. Symp. Wearable Comput. - ISWC '16*, pp. 108–115, 2016.

[12] I. Oakley, D. Lee, M. R. Islam, and A. Esteves, "Beats: Tapping Gestures for Smart Watches" *Proc. 33rd Annu. ACM Conf. Hum. Factors Comput. Syst. - CHI '15*, pp. 1237–1246, 2015.

[13] H.-S. Yeo, J. Lee, A. Bianchi, and A. Quigley, "Sidetap & Slingshot Gestures on Unmodified Smartwatches" *Proc. 29th Annu. Symp. User Interface Softw. Technol. - UIST '16 Adjun.*, pp. 189–190, 2016.

[14] I. Oakley and D. Lee, "Interaction on the Edge: Offset Sensing for Small Devices" *CHI '14 Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, pp. 169–178, 2014.

Deployment of Mobile Edge Location Service

Denitsa Kireva[1], Evelina Pencheva[1], Ivaylo Atanasov[1], Ventsislav Trifonov[1]

Abstract – **Multi-access Edge Computing (MEC) provides virtualized "cloud native" functionality into the radio access network. In this paper, we study the deployment of location service in MEC environment. The location service procedures are mapped onto positioning protocol functionality. Location service models are proposed, formally described and verified.**

Keywords – **Multi-access Edge Computing, Application Programming Interfaces, Finite state machines.**

I. Introduction

Multi-access Edge Computing (MEC) provides cloud and IT capabilities in a close vicinity to end users at the edge of the network. It offers customized mobile broadband experience by providing contextual and location information. MEC addresses challenges such as high latency, security vulnerability, low coverage, and lagged data transmission [1]. A number of security, safety and data analytics mobile edge applications may be deployed based on location information [2]. Location-based services are a promising way of exploiting the special possibilities created by ubiquitous mobile devices and wireless communication. Advanced location-based applications will require highly accurate information about the geographic location of mobile objects and functionality that goes beyond simply querying the user's position, for example determining all mobile objects inside a certain geographic area. Different aspects of location-based services related to positioning algorithms are studied in [3], [4]. Location-based services are beneficial for retailers, healthcare providers, airports, government organizations and many other enterprises around the world for multimedia broadcasting [5].

MEC is a novel paradigm and it is currently under standardization. In [6], the authors study how the capabilities of well established Parlay X Terminal Location Web Service may be used to provide location service in MEC environment. In this paper, we explore the way Mobile Edge Location Application Programming Interfaces (API) defined by ETSI may be implemented in the radio access network.

In the next sections, we describe the architecture for Mobile Edge Location Service, provide functional mapping of service API onto positioning protocol and model the location service state as seen by mobile edge application and by the network.

II. Architecture of Mobile Edge Location Service

The Mobile Edge Location service provides authorized applications with location-related information [7]. It exposes information to applications, such as:

- the location of specific UEs currently served by the radio node(s) associated with the mobile edge host;
- information about the location of all UEs currently served by the radio node(s) associated with the mobile edge host;
- optionally, information about the location of a certain category of UEs currently served by the radio node(s) associated with the mobile edge host;
- a list of UEs in a particular location;
- information about the location of all radio nodes currently associated with the mobile edge host;
- location in form of geolocation, Cell ID, etc.

The architecture for deployment of MEC Location service is shown in Fig.1. MEC platform exposes essential functionality required to run MEC applications on a particular virtualization infrastructure enabling them to provide and consume mobile edge services. The MEC platform is in the close vicinity of the eNodeB. The MEC platform may be integrated with Location server in order to provide Location service. The protocol between the target User equipment (UE) and the Location server is LTE Positioning Protocol (LPP) [8]. LPP is used to position the target UE using positioning-related measurements obtained by one or more reference sources or to transfer assistance data.

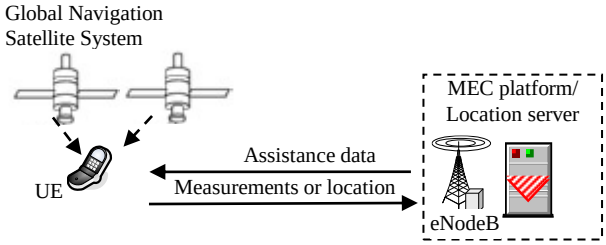


Fig.1 Architecture for Mobile Edge Location Service

The LPP procedures include the following:

- Transfer of positioning and protocol capabilities related to LPP and the positioning methods supported by LPP;
- Transfer of assistance data related to positioning;
- Transfer of location information where location information applies both to an actual position estimate and to values used in computing position (e.g., radio measurements or positioning measurements);
- Abort procedure used by one endpoint to notify the other endpoint to abort an ongoing procedure between the two endpoints.

III. Functional Mapping of Location Service API onto LLP

The UE Location Lookup is the procedure for applications acquiring the current location information of a specific UE or a group of UEs. In this procedure, the Location Service will

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

report the lookup result once on each request. The Mobile edge application looks up an UE location by sending a request to the resource representing the UE location, which includes the UE(s) identifier, e.g. UE IP address. The MEC platform needs to determine which of the positioning methods has to be used in order to provide the requested accuracy. To do this, it starts an LPP dialogue with the target, as shown in Fig.2.

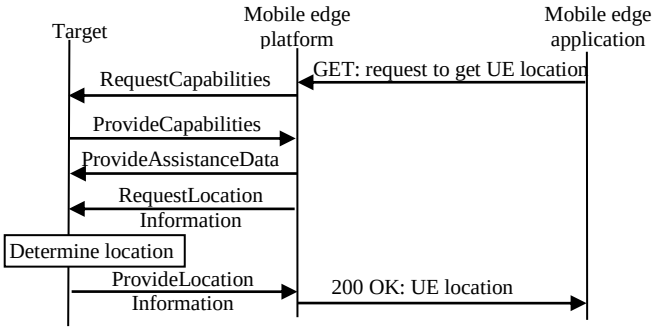


Fig.2 Flow of UE location lookup

The UE Information Lookup is the procedure for applications acquiring information of a list of UEs in a particular location. In this procedure, the Location Service reports the lookup result once on each request. First, the mobile edge application looks up UE information in a particular area by sending a request to the resource representing the UE Information, which includes location area information. The mobile edge platform identifies the base stations serving the cells in the location area. Next, each of the base stations identifies UE for which there is an established UE context. UE context is a block of information in a base station associated to an active UE. The block of information contains the necessary information required to maintain the radio access network services towards the active UE. At least UE state information, security information, UE capability information and the identities of the UE-associated logical S1-connection are included in the UE context. Having information about UE served by the base stations associated with the requested location area, the mobile edge platform queries the UEs about their capabilities and requests location information. The mobile edge platform may transfer additional assistance data to the target in one or more additional LPP messages. When the mobile edge platform receives information about UE(s) location it determines whether the respective UE is in the requested location area, and if so the UE is included in the UE list returned as a response to the mobile edge application. The flow for UE information lookup is shown in Fig.3.

The UE Location Subscribe is the procedure for applications acquiring up-to-date location information of a specific UE or a group of UEs in a period, which help the applications to track the UE(s). In this procedure, the Location Service will continue to report the subscribed information until the subscription is cancelled. Fig.4 illustrates a subscription to periodic UE location information reporting and periodic notifications.

The UE Information Subscribe is the procedure for applications to receive notifications of UE Information updates for the list of UEs in a particular location. The UE information can be used to update the status changes or periodic notifications of UE information. This procedure initiates similar action in the network as for UE Location Subscribe procedure.

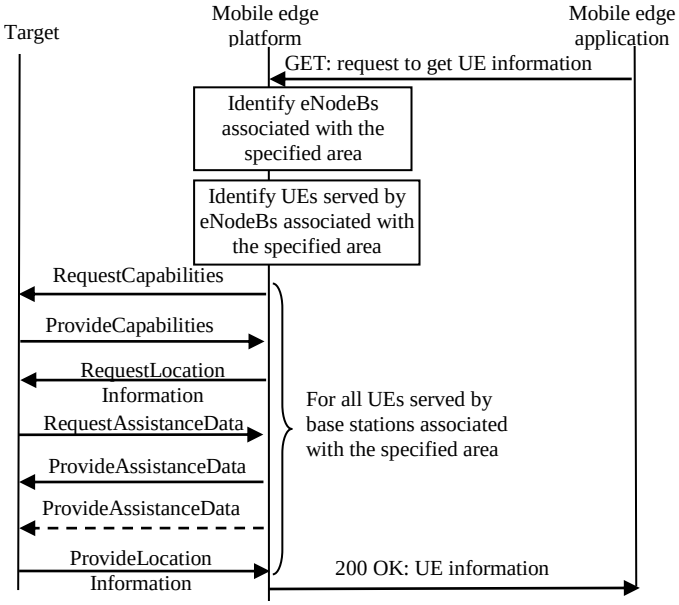


Fig.3 Flow for UE information lookup

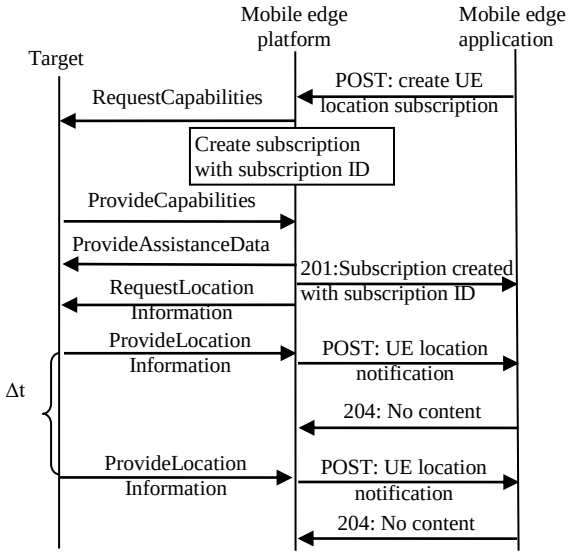


Fig.4 Flow of UE location information subscribe

The mobile edge application may at any time modify the subscription (e.g. the reporting period, or specified area) or to terminate the subscription. In order to modify existing subscription, mobile edge application updates the subscription resource by sending a PUT request to the resource containing all the subscriptions of the specific subscription type with the modified data structure specific to that subscription. Location service returns "200 OK" with the message body containing the accepted data structure specific to that subscription.

The Subscribe Cancellation is the procedure for applications to cancel the subscription, with which the Location Service must stop reporting the subscribed information to the application. Fig.5 shows the flow for subscription termination.

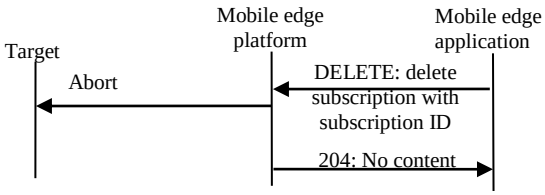


Fig.5 Flow of Location Subscribe Cancellation

IV. Location Service Models

The application and network views on location service status need to be synchronized during the process of UE(s) positioning. In this section, we propose models representing location service status as seen by a mobile edge application and the network view, and provide a method for its formal verification.

The application view on the location state is shown in Fig.6.

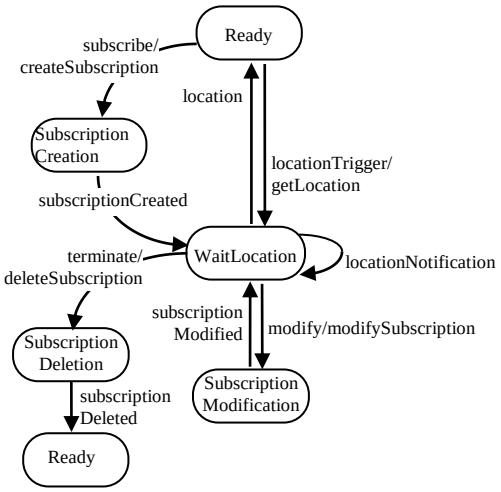


Fig. 6 Application view on the location service state

In Ready state, there is no particular event triggering location request or the location request has already been completed. In Ready state, when the application logic requests location information, it sends a query (getLocation) to the network. In Ready state, the mobile edge application may subscribe for location information notifications. Being in SubscriptionCreation state, the application waits for subscription acknowledgement. In WaitLocation state, the application waits for location information requests on demand or in the subscription frame. In WaitLocation state, when location information requested on demand is received, the location service moves to Ready state. In WaitLocation state, when location information notification is received, the location service remains in the same state. In WaitLocation state, the subscription for location information may be modified or terminated. In SubscriptionModification state, the application waits for subscription modification acknowledgement. In SubscriptionDeletion state, the application waits for subscription termination acknowledgement.

We use the mathematical formalism of Labeled Transition Systems (LTSs) to describe the location service state models. An LTS is defined as a quadruple of set of states, set of inputs, set of transitions, and an initial state.

By L_{App}= (S_{App}, Inp_{App}, $\rightarrow_{App}$, s_0^{App}) it is denoted a Labeled Transition System (LTS) representing the Application's view on location service state where:

S_{App} = { Ready [s_1^A], SubscriptionCreation [s_2^A], WaitLocation [s_3^A], SubscriptionModification [s_4^A], SubscriptionDeletion [s_5^A]};

Inp_{App} = {locationTrigger[t_1^A], subscribe[t_2^A], subscriptionCreated[t_3^A], location [t_4^A], locationNotification [t_5^A], modify [t_6^A], subscriptionModified[t_7^A], delete[t_8^A], subscriptionDeleted[t_9^A]};

$\rightarrow_{App}$ = { (s_1^A t_1^A s_3^A), (s_1^A t_2^A s_2^A), (s_2^A t_3^A s_3^A), (s_3^A t_5^A s_1^A), (s_3^A t_6^A s_4^A), (s_4^A t_7^A s_3^A), (s_3^A t_8^A s_5^A), (s_5^A t_9^A s_1^A)}.

s_0^{App}= { s_1^A }.

Short notations are given in brackets.

Fig.7 shows the network view on the location service state.

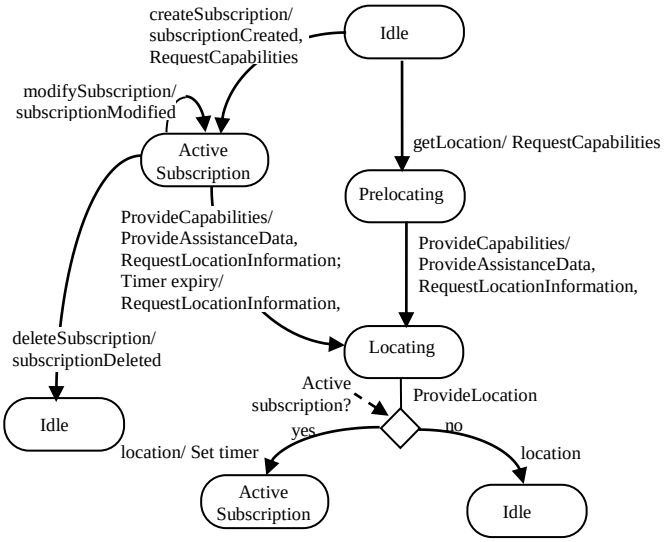


Fig. 7 Network view on the location service state

In Idle state, there is no location request or the location request is completed. Being in this state, the location service moves to ActiveSubscription state, if a request for subscription to location information is received. The location service moves to PreLocating state when a request for location information is received. In Prelocating state, positioning capabilities and assistance data are transferred. The location service queries the UE(s) about device capabilities and provides assistance data. In Locating state, the target UE(s) is being located. The transition from Locating to Idle state occurs when location information is provided. In case of active subscription, the timer for periodic location notifications is set and the location services moves to ActiveSubscription state. In ActiveSubscription state, the subscription may be modified or terminated.

By L_{MEC}= (S_{MEC}, Inp_{MEC}, $\rightarrow_{MEC}$, s_0^{MEC}) it is denoted an LTS representing the network view on location service state where:

S_{MEC}={ Idle[s_1^M], ActiveSubscription[s_2^M], Prelocating[s_3^M], Locating[s_4^M], };

Inp_{MEC} = { createSubscription[t_1^M], getLocation[t_2^M], ProvideCapabilities[t_3^M], TimerExpiry [t_4^M], ProvideLocation[t_5^M], modifySubscription[t_6^M], deleteSubscription[t_7^M]};

$\rightarrow_{MEC}$ = { (s_1^M t_1^M s_2^M), (s_1^M t_2^M s_3^M), (s_2^M t_3^M s_4^M), (s_3^M t_3^M s_4^M), (s_2^M t_4^M s_4^M), (s_4^M t_5^M s_1^M), (s_4^M t_5^M s_2^M), (s_2^M t_6^M s_2^M), (s_2^M t_7^M s_1^M)}

s_0^{MEC}={ s_1^M }.

Having formal description of the models representing the location service status as seen by mobile edge application and network, we can prove that these models are synchronized i.e. they expose equivalent behavior.

Intuitively, in terms of observed behavior, two LTSs are equivalent if one LTS displays a final result and the other LTS displays the same result. The idea of equivalence is formalized by the concept of bisimilarity [9]. In practice, strong bisimilarity puts strong conditions for equivalence which are not always necessary. Weak bisimilarity allows internal transitions to be ignored.

Proposition: The labeled transition systems L_{App} and L_{MEC} are weakly bisimilar.

Proof: As to definition of weak bisimulation, it is necessary to identify a bisimilar relation between the states of both LTSs and to identify respective matching between transitions. Let U_{AppMEC} be a relation between the states of L_{App} and L_{MEC} and U_{AppMEC}= {(s_1^A, s_1^M), (s_3^A, s_4^M)}. Then for the following events the respective transitions are identified:

1. The application requests for location information: for (s_1^A t_1^A s_3^A) $\exists$ (s_1^M t_2^M s_3^M) $\sqcap$ (s_3^M t_3^M s_4^M).
2. The application creates subscription for location information: for (s_1^A t_2^A s_2^A) $\sqcap$ (s_2^A t_3^A s_3^A) $\exists$ (s_1^M t_1^M s_2^M) $\sqcap$ (s_2^M t_3^M s_4^M).
3. The location information requested on demand is provided: for (s_3^A t_4^A s_1^A) $\exists$ (s_4^M t_5^M s_1^M).
4. The application is notified about location information: for (s_3^A t_5^A s_3^A) $\exists$ (s_4^M t_5^M s_2^M) $\sqcap$ (s_2^M t_4^M s_4^M).
5. The application modifies the subscription for location information: for (s_3^A t_5^A s_3^A) $\sqcap$ (s_3^A t_6^A s_4^A) $\sqcap$ (s_4^A t_7^A s_3^A) $\exists$ (s_4^M t_5^M s_2^M) $\sqcap$ (s_2^M t_6^M s_2^M) $\sqcap$ (s_2^M t_4^M s_4^M).
6. The application terminates the subscription for location information: for (s_3^A t_8^A s_5^A) $\sqcap$ (s_5^A t_9^A s_1^A) $\exists$ (s_4^M t_5^M s_2^M) $\sqcap$ (s_2^M t_7^M s_1^M).

Therefore L_{App} and L_{MEC} are weakly bisimilar.■

V. Conclusion

Location service retrieves real-time geographical data for mobile devices or smartphones to provide information, entertainment or security. Some applications allow consumers to "check in" at restaurants, coffee shops, stores, concerts, and other places or events. Deployment of location service in MEC environment enables customized applications for safety, security and big data analytics. In this paper, we study the way Mobile Edge Location Service may be implemented in the radio access network. We examine the mapping of Location API onto LPP procedures, propose models representing the application and MEC platform views on the location service state and a formal method for their verification.

Acknowledgement

The research is conducted under the grant of project DH07/10-2016, funded by National Science Fund, Ministry of Education and Science, Bulgaria.

References

[1] N. Abbas, Y. Zhang, A. Taherkordi and T. Skeie, "Mobile Edge Computing: A Survey," IEEE Internet of Things Journal, vol. 5, no. 1, pp. 450-465, Feb. 2018.

[2] M. Satyanarayanan, "The Emergence of Edge Computing, " IEEE Computer, January 2017, pp. 30-39.

[3] M. Aal-Nouman, O. H. Salman, H. Takruri-Rizk and M. Hope, "A new architecture for location-based services core network to preserve user privacy," 2017 Annual Conference on New Trends in Information & Communications Technology Applications (NTICT), 2017, pp. 286-291.

[4] J. Haifeng and W. Bin, "Design and Implementation of Location Service Based on Signal Strength Difference," 2017 International Conference on Intelligent Environments (IE), Seoul, 2017, pp. 1-5.

[5] Q. Wei and L. Duan, "Research and implementation of mobile advertising system based on location service," 2017 3rd International Conference on Information Management (ICIM), Chengdu, 2017, pp. 407-411.

[6] E. Pencheva, I. Atanasov, K. Kassev and V. Trifonov, "Location service in mobile edge computing," 2017 Ninth International Conference on Ubiquitous and Future Networks (ICUFN), Milan, 2017, pp. 617-622.

[7] ETSI GS MEC 013 Mobile Edge Computing (MEC); Location API, v1.1.1, 2017.

[8] 3GPP TS 36.355 Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access Network (E-UTRAN); LTE Positioning Protocol (LPP) (Release 13), v13.3.0, 2016.

[9] D. Escrig, J. Keiren, T. Willemse, "Games for Bisimulations and Abstraction," Cornel University Library, 2017, arXiv: 611.00401 [cs.LO].

Deployment of Mobile Edge Bandwidth Management Service

Denitsa Kireva[1], Evelina Pencheva[1], Ivaylo Atanasov[1], Ventsislav Trifonov[1]

Abstract – **Multi-access Edge Computing (MEC) provides processing and storage capabilities of the cloud into the radio access network. In this paper, we study the deployment of bandwidth management service in MEC environment. The bandwidth management service procedures are mapped onto functionality of the control protocol between radio access network and core network. An extension of the bandwidth management service is proposed that enables applications to monitor the status of configured bandwidth allocation. Models representing the status of bandwidth allocation as seen by mobile edge application and network are proposed, formally described and verified.**

Keywords – **Multi-access Edge Computing, Application Programming Interfaces, Finite state machines.**

I. Introduction

Multi-access Edge Computing (MEC) is the response of challenges to cloud computing related to requirements of massive Internet of Things communications for short delays and high bandwidth [1], [2]. It distributes the processing logic and storage capabilities within the Radio Access Network in close proximity to mobile devices [3]. The main purpose of MEC may be summarized as follows: real-time distributed computing, reduction of processor load in end devices (offloading), and localization of data [4].

The mobile edge platform provides infrastructure services, radio network information services, bandwidth management services and location services. The communication between mobile edge platform and mobile edge applications is based on Representational State Transfer (REST) style [5].

MEC standardization is ongoing with many challenges yet to be addressed [6]. In this paper, we explore the way Mobile Edge Bandwidth Management Service (BWMS) Application Programming Interfaces (API) defined by ETSI may be implemented in the radio access network.

In the next sections, we describe the functional mapping of BWMS API onto S1 Application Protocol (S1AP), propose an extension of BWMS functionality and model the resource state as seen by mobile edge application and by the network.

II. Functional Mapping of Bandwidth Management Service API onto S1AP

When the mobile edge system supports the feature Bandwidth Manager, the mobile edge platform or a dedicated mobile edge application enables an authorized mobile edge application to register statically and/or dynamically its bandwidth requirements and/or priority. Different applications, whether managing a single instance or several sessions (for example content delivery network), may request specific bandwidth requirements (bandwidth size, bandwidth priority, or both) for the whole application instance or different bandwidth requirements per session. The BWMS may aggregate all the requests and act in a manner that will help optimize the bandwidth usage [7].

Bandwidth management functionality is mapped onto E-RAB (E-UTRAN Radio Access Bearer) management functions of S1AP. S1AP provides the signaling between E-UTRAN and the evolved packet core required to perform S1AP functions [8]. This overall S1AP functionality for E-RAB management is responsible for setting up, modifying and releasing E-RABs, which are triggered by the MME. The release and modification of E-RABs may be triggered by the eNodeB as well.

Fig.1 shows a scenario where a mobile edge application registers to BWMS. This scenario is related to E-RAB Setup procedure. The purpose of the E-RAB Setup procedure is to assign resources on Uu and S1 for one or several E-RABs and to setup corresponding Data Radio Bearers for given UE.

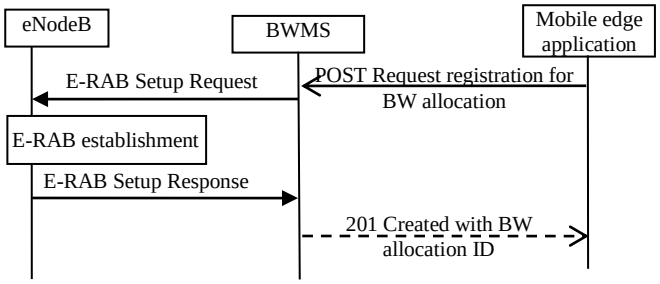


Fig.1 Flow of mobile edge application registration to BWMS

Fig.2 shows a scenario where a mobile edge application instance unregisters from BWMS. This scenario is related to E-RAB Release procedure. The purpose of the E-RAB Release procedure is to enable the release of already established E-RABs for given UE.

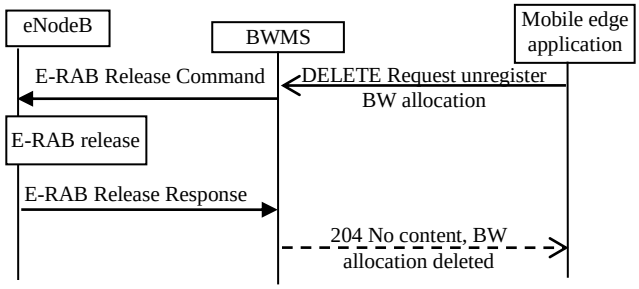


Fig.2 Flow of mobile edge application unregistering bandwidth allocation from BWMS

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

Fig.3 shows a scenario where a mobile edge application instance updates its requested bandwidth requirements on the BWMS. This scenario is related to E-RAB Modify procedure. The purpose of the E-RAB Modify procedure is to enable modifications of already established E-RABs for given UE.

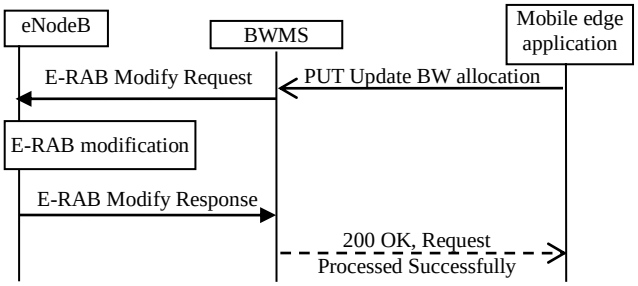


Fig.3 Flow of mobile edge application updating its requested bandwidth requirement on BWMS

Fig.4 shows a scenario where a mobile edge application instance gets its configured bandwidth allocation from the BWMS. The scenario does not include interaction with the eNodeB.

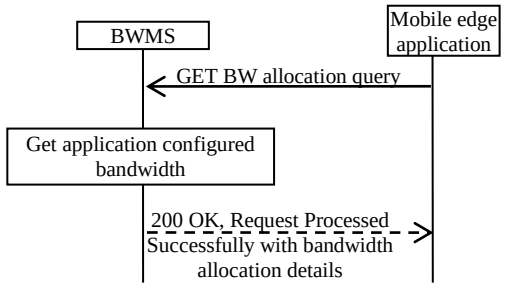


Fig.4 Flow of mobile edge application getting its configured bandwidth allocation from BWMS [7]

III. An Extension of BWMS Functionality

The eNodeB may have information about E-RAB release in case of unexpected service data flow deactivation (e.g. the user is out of coverage) or information about E-RAB modification. This information may be transferred to mobile edge application which has registered its bandwidth requirements to BWMS. The current BWMS specification does not describe capabilities for subscription and notifications about E-RAB related events.

We propose an extension of BWMS with functionality enabling mobile edge applications to subscribe for events related to E-RAB release and modification as far as these functions are supported by S1AP.

To be able to monitor selected E-RAB events, the application creates a subscription to certain specific E-RAB event that is available at BWMS. Fig.5 shows a scenario where the application uses REST based procedures to create a subscription for E-RAB event notifications.

Fig.6 shows a scenario where the application needs to update an existing monitoring of RAB event management. The monitoring update is triggered e.g. by the need to change the existing monitoring, or due to the expiry of the monitoring.

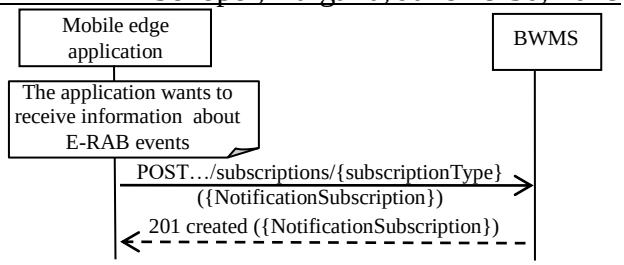


Fig.5 Flow pattern for subscribing to E-RAB event notifications

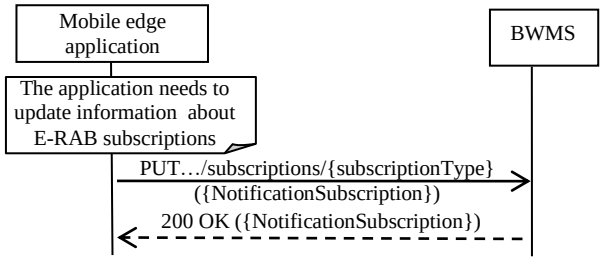


Fig.6 Flow of application updating RAB management monitoring

When the mobile edge application does not need to receive notifications about E-RAB related events, the application terminates the subscription by sending DELETE message.

Fig.7 presents the scenario where the BWMS sends a notification about E-RAB modification to the mobile edge application.

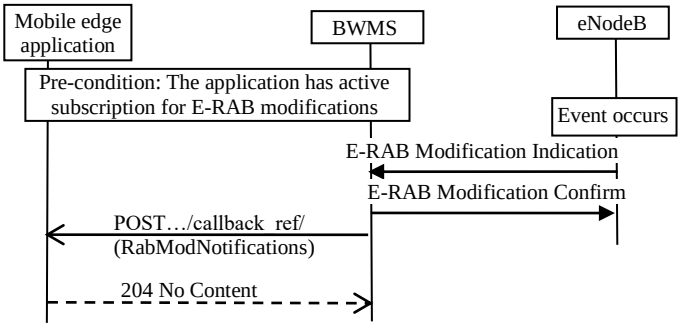


Fig.7 Flow of application subscription for notifications about E-RAB modifications

Fig.8 presents the scenario where the BWMS sends a notification about E-RAB release to mobile edge application.

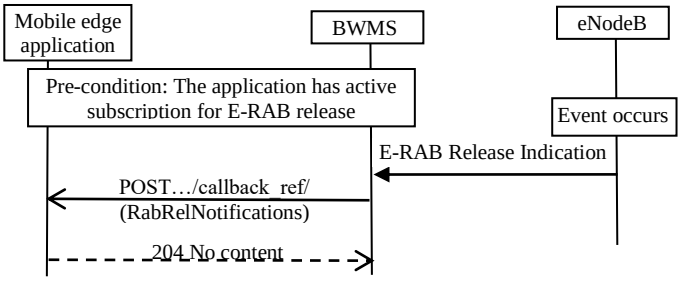


Fig.8 Flow of application subscription for notifications about E-RAB release

IV. RESOURCE STATE MODELS

The application and network views on the state of allocated bandwidth need to be synchronized during the process of RAB management for a particular UE. In this section we propose models representing bandwidth allocation state as seen by a mobile edge application and RAB state as maintained by the network, and provide a method for model formal verification.

The application view on the bandwidth allocation state is shown in Fig.9. In Null state, there is no registration for specific bandwidth allocation. In RegisteringBWM state, a mobile edge application performs registration for bandwidth allocation. In BWMRegistred state, the bandwidth allocation is successful. In UpdatingBWM state, the mobile edge application updates the configured bandwidth allocation. In UnregisteringBWM state, the mobile edge application de-registers its bandwidth allocation from BWMS.

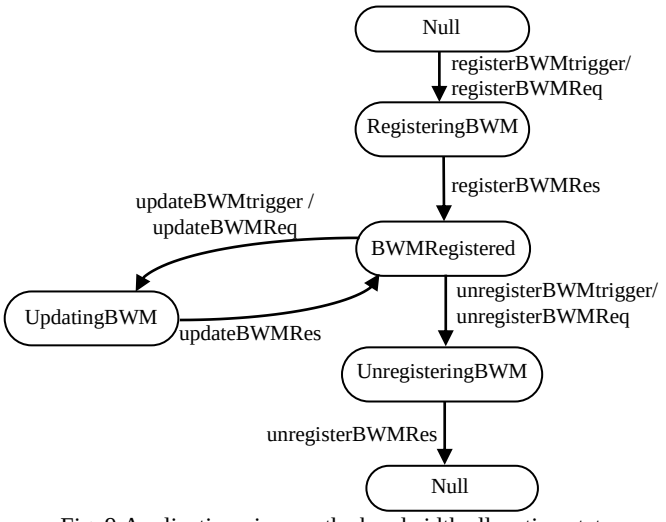


Fig. 9 Application view on the bandwidth allocation state

We use the mathematical formalism of Labeled Transition Systems (LTSs) to describe the location service state models. An LTS is defined as a quadruple of set of states, set of inputs, set of transitions, and an initial state.

By BWM_{App}= (S_{App}, Act_{App}, $\rightarrow_{App}$, s_0^{App}) it is denoted a Labeled Transition System (LTS) representing the Application's view on bandwidth allocation state where:

- S_{App} = {Null [s_1^A], RegisteringBWM [s_2^A], BWMRegistered [s_3^A], UpdatingBWM [s_4^A], UnregisteringBWM [s_5^A]};
- Act_{App} = {registerBWMTrigger[t_1^A], registerBWMRes[t_2^A], updateBWMTrigger[t_3^A], updateBWMRes [t_4^A], unregisterBWMTrigger [t_5^A], unregisterBWMRes [t_6^A]};
- $\rightarrow_{App}$ = {(s_1^A t_1^A s_2^A), (s_2^A t_2^A s_3^A), (s_3^A t_3^A s_4^A), (s_4^A t_4^A s_3^A), (s_3^A t_5^A s_5^A), (s_5^A t_6^A s_1^A)}.
- s_0^{App}= { s_1^A }.

Short notations are given in brackets.

Fig.10 shows the proposed simplified state model of E-RAB that has to be maintained by the eNodeB. When the user connects to the network an RRC connection and a default E-RAB are established. A user connection with specific QoS requires establishment of a dedicated E-RAB with appropriate QoS, which results in RRC connection re-configuration. The transitions in the state model correspond to the respective RRC and S1AP messages.

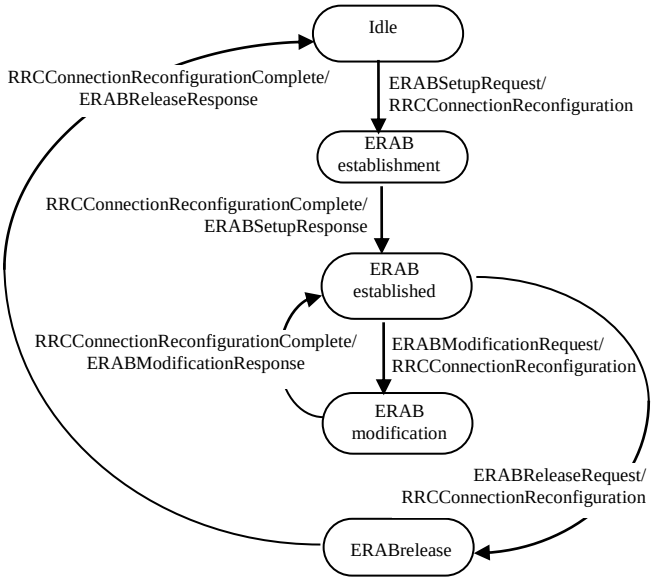


Fig.10 E-RAB state model maintained by the eNodeB

By BWM_{eNB} = (S_{eNB}, Act_{eNB} $\rightarrow_{eNB}$, s_0^{eNB}) it is denoted an LTS representing the eNodeB's view on E-RAB state, where:

- S_{eNB} ={ Idle [s_1^N], ERABestablishment [s_2^N], ERABestablished [s_3^N], ERABmodification [s_4^N], ERABrelease [s_5^N] }
- Act_{eNB} = {ERABSetupRequest [t_1^N], RRCConnectionReconfigurationComplete$_{Est}$ [t_2^N], ERABModificationRequest [t_3^N], RRCConnectionReconfigurationComplete$_{Mod}$ [t_4^N], ERABReleaseRequest[t_5^N], RRCConnectionReconfigurationComplete$_{Rel}$ [t_6^N]};
- $\rightarrow_{eNB}$ = { (s_1^N t_1^N s_2^N), (s_2^N t_2^N s_3^N), (s_3^N t_3^N s_4^N), (s_4^N t_4^N s_3^N), (s_3^N t_5^N s_5^N), (s_5^N t_6^N s_1^N)};
- s_0^{eNB} = { s_1^N }.

Having formal description of the models representing the location service status as seen by mobile edge application and network, we can prove that these models are synchronized i.e. they expose equivalent behavior.

Intuitively, in terms of observed behavior, two LTSs are equivalent if one LTS displays a final result and the other LTS displays the same result. The idea of equivalence is formalized by the concept of bisimilarity [9].

Proposition: The labeled transition systems BWM_{App} and BWM_{MEC} are weakly bisimilar.

Proof: First we identify a homomorphism $H_{BWM}(t_x^A, t_y^N)$ between actions of both state machines, shown in Table 1.

TABLE I. HOMOMORPHISM BETWEEN ACTIONS OF BWM_{APP} AND BWM_{MEC}

$H_{BWM}(t_x^A, t_y^N)$	Event description
$H_{BWM}(t_1^A, t_1^N)$	The mobile edge application sends a request to register to the BWMS.
$H_{BWM}(t_2^A, t_2^N)$	BWMS responds with a registration and initialization approval
$H_{BWM}(t_3^A, t_3^N)$	The mobile edge application sends a request to update a specific bandwidth allocation on the BWMS.
$H_{BWM}(t_4^A, t_4^N)$	BWMS responds with an update approval.
$H_{BWM}(t_5^A, t_5^N)$	The mobile edge application sends an unregister request to BWMS
$H_{BWM}(t_6^A, t_6^N)$	BWMS responds with a deregistration approval.

As to definition of strong bisimulation, it is necessary to identify a bisimilar relation between the states of both LTSs and to identify respective matching between transitions. Let U_{AppMEC} be a relation between the states of BWM_{App} and BWM_{MEC} where U_{AppMEC}= {(s_1^A, s_1^M), (s_2^A, s_2^N), (s_3^A, s_3^N), (s_4^A, s_4^N), (s_5^A, s_5^N)}. Then for the following events the respective transitions are identified:

1. In case of $H_{BWM}(t_1^A, t_1^N)$: for $(s_1^A\ t_1^A\ s_2^A) \exists (s_1^N\ t_1^N\ s_2^N)$.
2. In case of $H_{BWM}(t_2^A, t_2^N)$: for $(s_2^A\ t_2^A\ s_3^A) \exists (s_2^N\ t_2^N\ s_3^N)$.
3. In case of $H_{BWM}(t_3^A, t_3^N)$: for $(s_3^A\ t_3^A\ s_4^A) \exists (s_3^N\ t_3^N\ s_4^N)$.
4. In case of $H_{BWM}(t_4^A, t_4^N)$: for $(s_4^A\ t_4^A\ s_3^A) \exists (s_4^N\ t_4^N\ s_3^N)$.
5. In case of $H_{BWM}(t_5^A, t_5^N)$: for $(s_4^A\ t_5^A\ s_5^A) \exists (s_4^N\ t_5^N\ s_5^N)$.
6. In case of $H_{BWM}(t_6^A, t_6^N)$: for $(s_5^A\ t_6^A\ s_1^A) \exists (s_5^N\ t_6^N\ s_1^N)$.

Therefore BWM_{App} and BWM_{MEC} are weakly bisimilar.■

The synchronized behaviour of the models allows to prove in a mathematically formalized manner that the approach is consistently implementable. Mathematical formalism for equivalence of behaviour is used to generate model-based test situations in order to demonstrate compliance of a system's implementation with its specification

V. CONCLUSION

Different mobile edge applications running in parallel on the same mobile edge host may require specific static/dynamic up/down bandwidth resources, including bandwidth size and bandwidth priority. As all mobile edge applications and application sessions are competing over the same shared bandwidth resources, an optional central bandwidth resource allocator service may exist on the mobile edge platform. The BWMS API enables all registered applications to statically and/or dynamically register for specific bandwidth allocation per session/application.

In this paper, we study the way mobile edge BWMS may be implemented in the radio access network. We examine the mapping of BWMS API onto S1AP procedures related to E-RAB management, propose an extension of BWMS that enables applications to monitor the status of configured bandwidth allocation. We model the state of bandwidth allocation as seen by a mobile edge application and the E-RAB state as seen by the network and prove that both models expose equivalent behavior.

The multi access edge applications for bandwidth management are generally aimed at improving performance of the network and user experience. Examples of such applications are orchestration of video streams and video optimization, local content caching at the edge, backhaul optimization, etc.

ACKNOWLEDGEMENT

The research is conducted under the grant of project ДН07/10-2016, funded by National Science Fund, Ministry of Education and Science, Bulgaria.

REFERENCES

[1] H. Tanaka, M. Yoshida, K. Mori, N. Takahashi, Multi-access Edge Computing: A Survey, Journal of Information Processing, vol.26, February 2018, pp.87-97.

[2] N. Abbas, Y. Zhang, A. Taherkordi and T. Skeie, Mobile Edge Computing: A Survey, IEEE Internet of Things Journal, vol. 5, no. 1, pp. 450-465, Feb. 2018.

[3] Y. Ai, M. Peng, K. Zhang, Edge cloud computing technologies for internet of things: A primer, Digital Communication Networks, ScienceDirect, https://doi.org/10.1016/j.dcan.2017.07.001

[4] T. Taleb, K. Samdanis, B. Mada, H. Flinck, S. Dutta and D. Sabella, On Multi-Access Edge Computing: A Survey of the Emerging 5G Network Edge Cloud Architecture and Orchestration, IEEE Communications Surveys & Tutorials, vol. 19, no. 3, pp. 1657-1681, 2017.

[5] ETSI GS MEC 002, Mobile Edge Computing (MEC); Technical Requirements v1.1.1, 2016.

[6] S. Wang, X. Zhang, Y. Zhang, L. Wang, J. Yang and W. Wang, A Survey on Mobile Edge Networks: Convergence of Computing, Caching and Communications, IEEE Access, 2017, vol. 5, pp. 6757-6779.

[7] ETSI GS MEC 015 Mobile Edge Computing (MEC); Bandwidth Management API, v1.1.1, 2017

[8] 3GPP TS 36.331 Evolved Universal Terrestrial Radio Access (EUTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); S1 Application Protocol (S1AP); Release 14, v14.1.0, 2017.

[9] Escrig, D., Keiren, J., Willemse, T. Games for Bisimulations and Abstraction, Cornel University Library, arXiv:1611.00401 [cs.LO], 2016

Modified Simulation Framework for Realization of Horizontal Handover in LTE Cellular Networks

Aydan Haka[1]

Abstract - **The LTE technology provides simultaneously voice, data and video with different priority on networks. LTE cellular network provides uninterrupted delivery of these services while on the move, and this is possible through the Handover procedure. In this paper is proposed a simulation framework for realization of handover procedure for LTE technology, which realizes the UE mobility, prioritizes traffic and reorders packets.**

Key words - **4G, LTE, Horizontal Handover, QoS**

I. INTRODUCTION

In [1] the European Commission presents coordinated designation and authorization of the 700 MHz band for wireless broadband by 2020 and coordinated designation of the sub-700 MHz band. According to this decision this frequency bands will be used for terrestrial systems capable of providing wireless broadband communications services and for deployment of 5G technologies.

LTE is a widely used 4G technology defined by 3GPP, capable of realizing Broadband Wireless Access services. LTE supports high spectrum efficiency low latency scalable bandwidth from 1.4 MHz to 20 MHz, using MIMO technology, OFDM technique for downlink and SC-FDMA for uplink, and allows the user to access the service in a state of moving both in one cell or between cells (Handover) without any termination of communication. According to [2], there are two basic Handover technologies: Hard Handover – also called break-before-make; Soft Handover – also called make-before-break.

Furthermore Handover procedure is divided into two categories:

- Horizontal handover – automatic switching between access points in one technology;
- Vertical handover - automatic switching from one technology to another at the point where the service is delivered.

The realization of Handover procedure depends on eNodeB's Reference Signal Receive Power (RSRP) values measured from UE [3, 4]. As shown in [5] the Horizontal handover at LTE is realized through the X2 interface of the eNodeB. The X2 is a point-to-point interface, and it can be established between the serving eNodeB and its neighbors. In case the X2 interface is not configured or the connection is blocked, the Handover procedure can be implemented via MME using S1 interface.

[1]Aydan Haka is with the Faculty of Computing and Automation at Technical University of Varna, st. Studentska 1, Varna 3000, Bulgaria, E-mail: aydin.mehmed@tu-varna.bg

To keep mobile users satisfy, carrying out a Handover requires providing good QoS. This can be achieved by studying the delay value during Handover and prioritizing the different types of network streams.

II. RELATED WORKS

There are many developments using Handover mechanisms on LTE cellular networks that offer different solutions to improve QoS parameters of the serving network.

In [6], authors propose an approach that enhances the capability of LTE-femtocell networks when dealing with downlink variable bit rate video transmission and supporting efficient mobility management through an optimized handover policy. This is realized as making a pre-allocation of radio resources, based on the knowledge of future required video traffic of connected users. The results show that Handover period is increased. However, the QoS for variable bit rate video traffic increases.

In [7] is proposed a pre-scheduling mechanism for real-time video delivery over LTE. Proposed mechanism dynamically adjusts the data rates of transmission for providing a high QoS for real-time video before new connection establishment. The results show higher throughputs compared to the EXP/PF scheme.

Research [8] focuses on the mobility management and Handover issues between different cells - macro, micro and femtocells. There is proposed algorithm which improves the performance of femtocell in terms of decreased number of Handover, which improves the QoS.

Research [9] focuses on the analysis of specific type of LTE traffic, the video streaming in frequency division duplex (FDD) mode in Handover process on LTE network. The results show that the QoS for the high speed UEs, which generates video stream, is not increased significantly.

Unlike the solutions considered the proposed simulation framework do not focus on only one type of traffic. It operates with all the types of traffic which LTE supports. It can be used to investigate QoS based on LTE standard and proposed traffic prioritization mechanism, and how it affects to different UEs. Moreover the presented simulator generates the transmission matrix of allocated resources for every eNodeB for one frame. Based on this transmission matrix it can be research the delay, throughput and packet delivery ratio values.

III. LTE SIMULATION MODEL

The main concept of cellular networks is the division of services into small areas called cells. Each cell has its own coverage area and operates with different parameters. Each of

them contains an eNodeB, which serves all users in the range, and ensures UE mobility between cells.

This study will focus on analyzing QoS for mobile users performing the Handover between neighboring eNodeB's within LTE cellular network.

According to [10] the Fig. 1 shows distribution of built eNodeB's from different PLMN in suburban area of Varna, Bulgaria.

Fig. 1 Distribution of eNodeB's on suburban area of Varna, Bulgaria

In this figure eNodeB's are built up along the highway in one line, and UE's move in two directions – from West to East, and from East to West. Represented placement of eNodeB's was used for simulation topology as shown in Fig. 2.

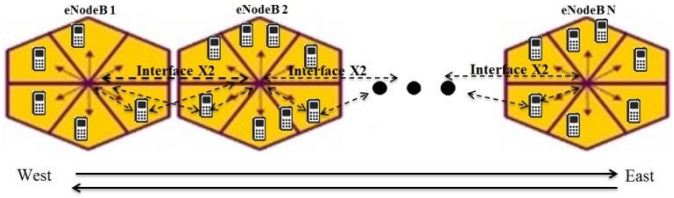


Fig. 2 Handover topology used for simulation

In the beginning every eNodeB has different number of connected UE's. They can be static or mobile, but this research focuses on mobile users. According to topology UE's can move in one of two directions (i.e. West to East or East to West). Every UE moves with different speed in one of the directions, and when reaches the end of serving cell initiates Handover. After that they connect to the next eNodeB in moving direction. Stages when UE performs the Handover process on the LTE network are as in Fig. 3.

The scheduler of every eNodeB performs proposed prioritization mechanism. According to this mechanism faster mobile users are with greater priority from slower users, because faster ones will reach end of cell first. So, faster UEs receive more resource unlike the others.

After the Handover is completed the number of UEs is changed dynamically for every eNodeB. Unchanged is just the number of static users, if available. The scheduler of eNodeB then prioritizes users and redistributes resources according to priority.

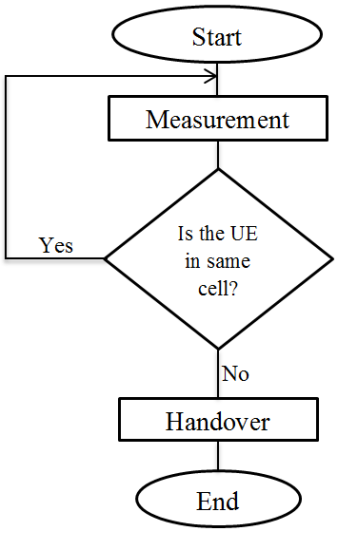


Fig. 3 Handover process diagram

IV. Proposed Algorithm for Prioritization of UEs in LTE

This simulation framework uses the same prioritization, presented from authors [11]. According to the priority the scheduler of eNodeB arranges UEs, where the order of priority is as: first, the users are ordered by payed priority – value from 0 to 7 where greater value indicates greater priority, second criteria is distance to eNodeB – users which are closer to eNodeB has greater priority, next criteria is speed of UE, if the user is mobile, where the high speed users has greater priority, in the end the users are prioritized according different type of service from required traffic flow.

V. Simulation Framework for Realization of Horizontal Handover in LTE

The simulation framework presented in this paper is improved version of simulator shown in [11]. Improvements consist in the fact that for mobile UEs a direction of movement can be selected in the range of the eNodeB, which allows implementation of mechanism for realization of Handover. Before the Handover is realized, it is checked at what speed and what direction the subscriber moves. After that, is calculated the distance that the UE will travel within five minutes in meters. The calculated value is added to distance to eNodeB given to the UE. After that, if the distance to eNodeB is greater than the radius of the serving eNodeB, the Handover to the target eNodeB is realized in to the movement direction. For the mobile UEs for which the value of distance to eNodeB is not greater than radius of cell, the Handover is not occurs, and they stay at the range of the serving eNodeB. The Handover is realized by the standard, and the context of the UE is transmitted to the next eNodeB. After the Handover is completed, the scheduler of eNodeB then prioritizes users by proposed mechanism, and redistributes resources according to priority. The Fig. 4 shows the new fields for selecting moving direction of UE.

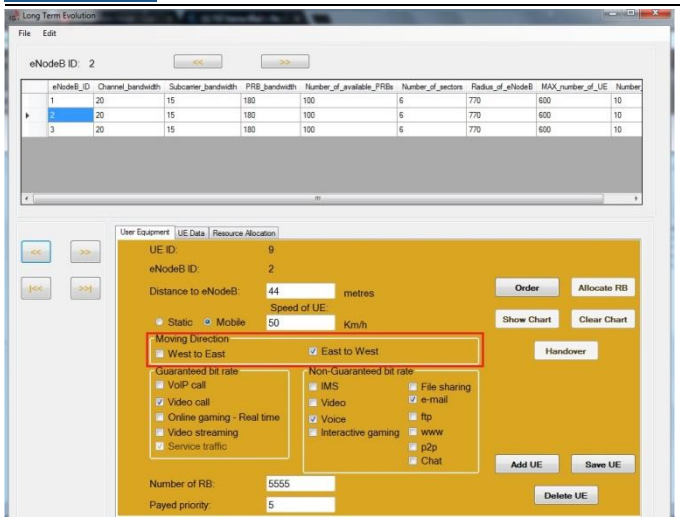


Fig. 4 Data of UE connected to eNodeB

Fig. 5 shows the base station information database and the related UEs. In the database for UE was added the fields "West_to_East" and "East_to_West", and they shows the selected moving direction of UE, if in the field is written "Checked". This figure shows connected UEs to eNodeB 2 before realization of Handover.

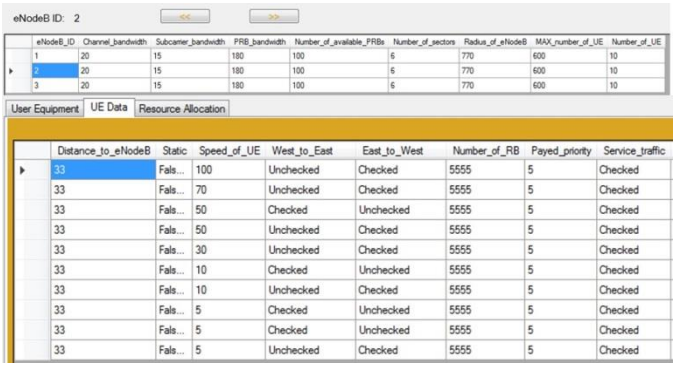


Fig. 5 Database with information of eNodeB 2 and related UEs

According to the speed of UE and the selected movement direction the high speed UEs, after the realization of Handover will connect to eNodeB 1 or eNodeB 3 in the first iteration. The low speed users are still connected to serving cell (i. e. eNodeB 2) as shown in Fig. 6, after realization of first Handover iteration. For eNodeB 1 and eNodeB 2 in the beginning there no connected UEs.

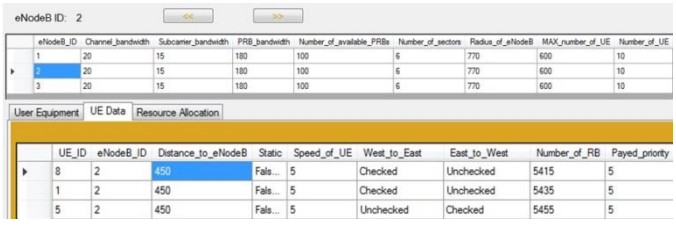


Fig. 6 UE data after realizing of the Handover on eNodeB 2

After the first iteration of Handover is complete, to eNodeB 1 are connected UEs with ID 4, 6, 9, 7 and 3 as shown in Fig. 7, and to eNodeB 3 are connected UEs with ID 2 and 10 as shown in Fig. 8.

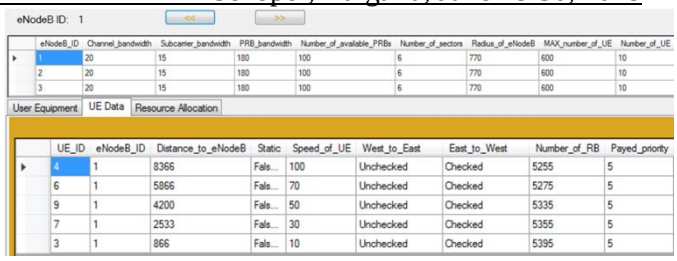


Fig. 7 Realized handover to eNodeB 1 in first iteration

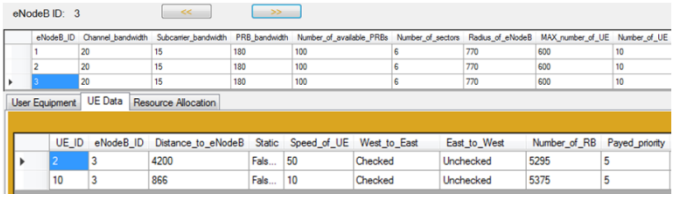


Fig. 8 Realized handover to eNodeB 3 in first iteration

After the second Handover iteration the low speed UEs will connect to eNodeB 3 – UEs with ID 8 and 1, and to eNodeB 1 - UE with ID 5. Fig. 9 and Fig. 10 shows realized handover for users in second iteration respectively for eNodeB 1 and eNodeB 3. After second iteration there no UEs connected to eNodeB 2.

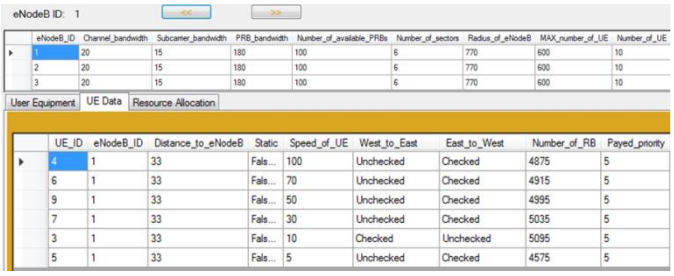


Fig. 9 Realized handover to eNodeB 1 in second iteration

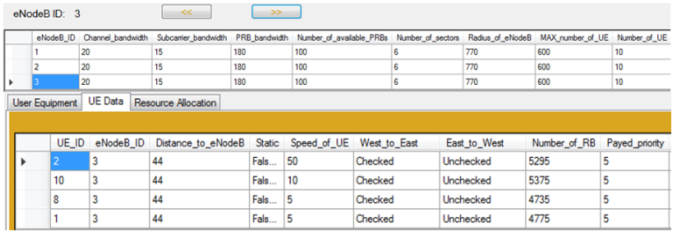


Fig. 10 Realized handover to eNodeB 3 in second iteration

VI. TESTS AND DISCUSSION

In this study, four tests with different number of UEs were carried out, which moves at different speeds. During the tests, users are moving from serving to the next eNodeB by performing a Handover procedure. After the Handover is completed, it is determined the number of realized and unrealized Handover procedures and the delay values. The delay values are calculated with the equation (1):

$$Delay = \frac{Total\ time\ for\ deliver\ a\ frame}{Total\ RB\ Sent} \qquad (1)$$

Fig. 11 shows the number of realized and unrealized Handover procedures. The Handover occurs when UE goes to

the edge of serving cell and the received signal is low. Then UE search for stronger signal and connect to next eNodeB.

According to the simulator the realization of Handover depends on UE location (i.e. distance to eNodeB) and movement speed. Because of this, mostly high speed UEs realizes more Handovers unlike the low speed UEs. As shown in Fig. 11, the number of realized handovers increases with increasing the number of high speed UEs unlike the number of unrealized, which depends on number of low speed UEs.

As the number of UEs increases, the number of Handover procedures increases, which gain the requirements of QoS. When a Handover is performed, improvement of the QoS may be achieved with decreasing the delay value for high speed UEs. This is performed from the scheduler of eNodeB and the prioritization mechanism on it. According to the prioritization mechanism the high speed UEs gain more resource blocks, and their requests are executed first. These UEs have greater priority, because they moves with high speed, and they may reach first the end of the cell, and will perform the Handover.

According to Fig. 12 it can be seen that the applied prioritization improves the QoS for high speed users, because unlike the other UEs, the high speed UEs has lower delay value. Although the delay values increase with increasing the number of UEs, they stay lowest for high speed UEs. The values for delay are calculated as average value from realized tests for resource allocation in one frame.

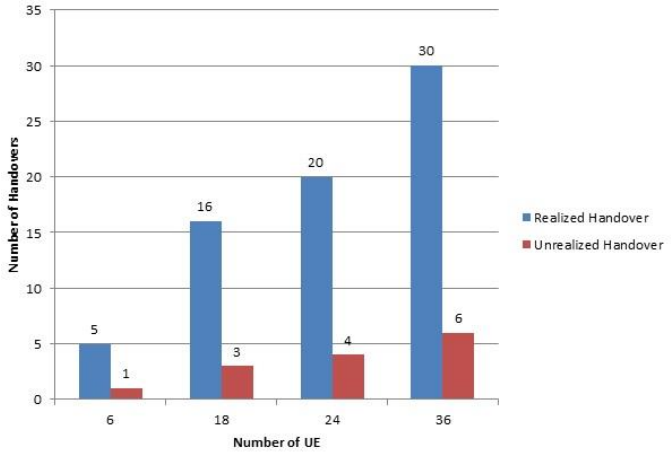


Fig. 11 Realized and Unrealized Handover procedures

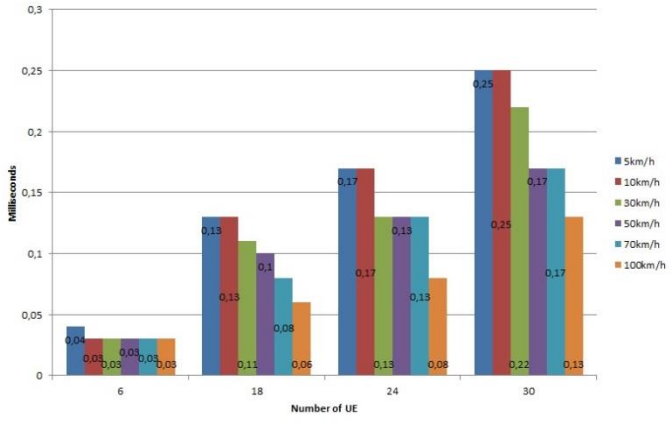


Fig. 12 Delay values with different number of UEs

VII. Conclusion

In this paper is proposed simulation framework for realization of horizontal Handover in LTE network. The framework performs an algorithm for realization of UE mobility between neighboring cells, according to the prioritization mechanism. Simulation's results show that the proposed prioritization mechanism improves QoS for high speed UEs. There are presented number of realized and unrealized Handovers and delay value for resource allocation by users. It was always assured a minimum delay value for allocated resources for high speed users realized Handover.

References

[1] Use of the 470-790 MHz frequency band in the Union. 2017. https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32017D0899 Last visit on 09.04.2018

[2] 802.21-2017-IEEE Standard for Local and metropolitan area networks--Part 21: Media Independent Services Framework. 2017. http://ieeexplore.ieee.org/document/7919341/, last visit on 02.03.2018

[3] Wang, Y., Chang, J., Huang, G. A Handover Prediction Mechanism Based on LTE-A UE History Information. //18th International Conference on Network-Based Information System (NBiS), Taipei, Taiwan, 2015

[4] Palla, S., Soumya, M. Self-Organizing Network Based Handover Mechanism for LTE Networks. //International Journal of Engineering Science and Computing, June 2017, Vol. 7, No. 6, pp: 13664 – 13668

[5] Alexandris K., Nikaein N., Knopp R. Analyzing X2 Handover in LTE/LTE-A. //IEEE: 2016 14th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt) – Tempe, AZ, USA, 2016

[6] Salhi, M., Trabelsi, S., Boudriga, N. Mobility-Assisted and QoS-Aware Resource Allocation for Video Streaming over LTE Femtocell Networks. //ECTI TRANSACTION ON ELECTRICAL ENG., ELECTRONICS, AND COMMUNICATIONS, 2015, VOL. 13, NO. 1

[7] Lai, W., Tai, C., Su, W. A Pre-Scheduling Mechanism in LTE Handover for Streaming Video. //applied science, 2016, Vol. 6, Issue 3

[8] Khan, M., Ashraf, M., Zafar, H., Ahmad, H. Enhanced Handover Mechanism in Long Term Evolution (LTE) Networks. //Internationa Journal of Communication Networks and Information Security (IJCNIS), 2017, Vol. 9, No. 1

[9] Latupapua, C. FJ., Priyambodo, T. K. Streaming Video Performance FDD Mode in Handover Process on LTE Network. //IJCCS (Indonesian Journal of Computing and Cybernetics Systems), 2018, Vol. 12, No. 1, pp: 43 – 52

[10] WIGLE.NET. https://wigle.net/ . Last visit on 31.03.2018.

[11] Aleksieva V., Haka A. Modified Scheduler for Traffic Prioritization in LTE Network. // Proceedings of the Second International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'17), 2017, Volume 2, pp. 228-238

Second order statistics of wireless signals over non-linear LOS fading channel with NLOS interference

Vesad Doljak[1], Dejan Milić [1], Suad Suljović[1], Emina Rizabegović[2] and Mihajlo Stefanović[1]

Abstract–In an urban environment, one can frequently encounter a situation where the wireless communication between a user and a wireless stations is impaired by the interfering signals coming from other stations and users. Often the station is in the clear view of the user, but the interfering transmitters are not. Therefore, this kind of wireless channel is interesting to be analysed in more detail. Here we consider a wireless fading channel where useful signal is subject to α-k-μ multipath fading, and co-channel interference is subjected to η-μ multipath fading. We denote the model as (α-k-μ)/(η-μ). This communication channel can be studied as ratio of α-k-μ random process and a η-μ random process. Since the interference is significant in today's wireless networks, the channel model ignores influence of receiver noise on the system performance. Probability density function and cumulative distribution function of (α-k-μ)/(η-μ) distribution are evaluated and level crossing rate of the (α-k-μ)/(η-μ) random process is calculated. By using these results, bit error rate probability, outage probability, and average fade duration of a radio transmission over (α-k-μ)/(η-μ) fading channel are determined. Influence of α-k-μ multipath fading severity and nonlinearity parameters on the level crossing rate is analysed.

Keywords –Level crossing rate, Cumulative distribution function, Average fade duration.

I. INTRODUCTION

In this paper, the model of the wireless channel (α-k-μ)/(η-μ) is formulated and estimated. This type of channel contains the desired α-k-μ signal and interference η-μ, and the ratio of this fading and interference itself contains several parameters. These parameters are α-k-μ short term fading nonlinearity parameter, Rician factor of α-k-μ short term fading, α-k-μ short term fading severity parameter, η-μ short term fading nonlinearity parameter and η-μ short term fading severity parameter. The (α-k-μ)/(η-μ) is a general fading channel and multiple known channel models can be derived from (α-k-μ)/(η-μ) fading channel.

There are several papers that discuss the distribution of fading and interference relationships, whose parameters affect the performance of the wireless telecommunication system. In the paper [1], there are first-order statistics for the distribution of η-μ and α-k-μ, eigenvalue functions and cumulative distribution. Using first-order statistics, the moments, the rate of the transient level of random processes α-k-η and η-μ can be calculated. In paper [2], the outage probability of selection combining diversity receiver in the presence of η-μ short term fading and Gama long term fading is evaluated. These results can be used in performance analysis of wireless relay comunication system with two sections in Nakagami-*m* multipath fading channel. Statistics of ratio of a random variable and product of two random variables is analysed in paper [3].

In this paper, two random variables and their ratio (α-k-μ)/(η-μ) are analyzed, and parameters that affect the performance of a wireless radio system functioning on the principle of a fading channel are evaluated. In this paper, we also consider a wireless communication system operating over the proposed fading channel. For this system, level crossing rate of the resulting signal to interference ratio random process is calculated. Probability density function can be used for evaluation of average symbol error probability for the proposed system, and outage probability can be evaluated by using cumulative distribution function of an (α-k-μ)/(η-μ) random variable. Obtained results can be used in performance analysis of wireless communication system operating over (α-k-μ)/(η-μ) multipath fading channel.

II. RATIO OF $\alpha-k-\mu$ AND $\eta-\mu$ RANDOM VARIABLES

The α-k-μ distribution can be used to describe small scale signal envelope variation in nonlinear, line of sight multipath fading environment. The η-μ distribution finds application for description of small scale signal envelope variation in nonlinear, non-line-of-sight multipath fading environments. The ratio *z* of a α-k-μ random variable *x* and a η-μ random variable *y* is:

$$z = x / y,\ \ x = zy,\ z = x^{2/\alpha} / y^{2/\alpha},\ z^{\alpha/2} = x / y \qquad (1)$$

Probability density function (PDF) function of variable *x* is [1, eq. (2.14)]:

$$p_x(x) = \frac{2\mu(k+1)^{\frac{\mu+1}{2}}}{k^{\frac{\mu-1}{2}} e^{\mu k} \Omega_1^{\frac{\mu+1}{2}}} x^{\mu} e^{-\frac{\mu(1+k)}{\Omega_1}x^2} I_{\mu-1}\left(2\mu\sqrt{\frac{k(k+1)}{\Omega_1}}x\right) =$$

$$= \frac{2}{e^{\mu k}} \sum_{i_1=0}^{+\infty} \frac{\mu^{2i_1+\mu}(k+1)^{i_1+\mu} k^{i_1} x^{2i_1+2\mu-1}}{\Omega_1^{i_1+\mu}\Gamma(i_1+\mu) i_1!} e^{-\frac{\mu(1+k)}{\Omega_1}x^2} \qquad (2)$$

with Ω=E [R^2], denoting average signal power, $I_n(\cdot)$ is the *n*-the order modified Bessel function of the first kind of order *c* by using well-known transformation [6, eq. (17.7.1.1)]:

[1]Vesad Doljak, Dejan Milić, Suad Suljović and Mihajlo Stefanović are with the Faculty of Electronics Engineering at University of Niš, Aleksandra Medvedeva 14, Niš 18000, Serbia, E-mail: vesko95@yahoo.com, dejan.milic@el.fak.ni.ac.rs, suadsara@gmail.com, misa.profesor@gmail.com

[2]Emina Rizabegović is with the Faculty of Technical Sciences University of Novi Pazar, Vuka Karadžića bb, Novi Pazar 36300, Serbia, E-mail: rizabegovicemina@gmail.com

$$I_v(x)=\sum_{k=0}^{+\infty}\frac{x^{v+2k}}{2^{v+2k}\,k!\,\Gamma(v+k+1)} \tag{3}$$

PDF of η-μ random variable y is [7]:

$$p_y(y)=\frac{4\sqrt{\pi}\,\mu^{\mu+\frac{1}{2}}h^{\mu}y^{2\mu}}{\Gamma(\mu)H^{\mu-\frac{1}{2}}\Omega^{\mu+\frac{1}{2}}}e^{-\frac{2\mu h}{\Omega}y^2}I_{\mu-\frac{1}{2}}\left(\frac{2H\mu}{\Omega}y^2\right)=$$

$$=\frac{4\sqrt{\pi}\,h^{\mu}}{\Gamma(\mu)}\sum_{i_2=0}^{+\infty}\frac{\mu^{2i_2+2\mu}y^{4i_2+4\mu-1}H^{2i_2}}{i_2!\,\Gamma\left(i_2+\mu+\frac{1}{2}\right)\Omega_2{}^{2i_2+2\mu}}e^{-\frac{2\mu h}{\Omega_2}y^2} \tag{4}$$

where $\Omega=\mathrm{E}[R^2]$, stands for the average power, while $\Gamma(a)$ denotes Gamma function, H and h are signal parameters, written in the function of parameter η_1 as [1]:

$$H=\frac{\eta_1-\eta_1^{-1}}{4},\ h=\frac{2+\eta_1^{-1}+\eta_1}{4} \tag{5}$$

First derivative of the ratio of α-k-μ and α-μ random variables is:

$$\dot{z}=\frac{2}{\alpha z^{\frac{\alpha}{2}-1}}\left(\frac{\dot{x}}{y}-\frac{x\dot{y}}{y^2}\right) \tag{6}$$

The first derivative of the ratio of an α-k-μ and η-μ random variables can be viewed as a conditional Gaussian distribution. The variance of z is [8]:

$$\sigma_{\dot{z}}{}^2=\frac{4}{\alpha^2 z^{\alpha-2}y^2}\left(f_1{}^2+z^{\alpha}f_2{}^2\right) \tag{7}$$

where variances involved are:

$$f_1{}^2=\pi^2 f_m{}^2\Omega_1,\ f_2{}^2=\pi^2 f_m{}^2\Omega_2 \tag{8}$$

and f_m is maximal Doppler frequency. Conditional probability density function of $\dot{z}$ is [9]:

$$p_{\dot{z}}(\dot{z}/zy)=\frac{1}{\sqrt{2\pi}\sigma_{\dot{z}}}e^{-\frac{\dot{z}^2}{2\sigma_{\dot{z}}{}^2}} \tag{9}$$

Joint probability density function of $z,\dot{z}$ channel y is:

$$p_{z\dot{z}y}(z\dot{z}y)=p_{\dot{z}}(\dot{z}/zy)\,p_{zy}(zy)=$$

$$=p_{\dot{z}}(\dot{z}/zy)\,p_y(y)\,p_z(z/y) \tag{10}$$

Conditional probability density function of z is:

$$p_z(z/y)=\left|\frac{dx}{dz}\right|p_x\left(yz^{\frac{\alpha}{2}}\right)=\frac{\alpha}{2}yz^{\frac{\alpha}{2}-1}p_x\left(yz^{\frac{\alpha}{2}}\right) \tag{11}$$

Joint probability density function of $z,\dot{z}$ and y is therefore:

$$p_{z\dot{z}y}(z\dot{z}y)=\frac{\alpha}{2}yz^{\frac{\alpha}{2}-1}p_x\left(yz^{\frac{\alpha}{2}}\right)p_y(y)\,p_{\dot{z}}(\dot{z}/zy) \tag{12}$$

Joint probability density function of $z,\dot{z}$ is [10]:

$$p_{z\dot{z}}(z\dot{z})=\int_0^{+\infty}p_{z\dot{z}y}(z\dot{z}y)\,dy=\frac{\sqrt{2}\,\alpha^2 h^{\mu}}{e^{\mu k}\Gamma(\mu)\left(f_1{}^2+z^{\alpha}f_2{}^2\right)^{\frac{1}{2}}}\cdot$$

$$\cdot\sum_{i_1=0}^{+\infty}\sum_{i_2=0}^{+\infty}\frac{\mu^{2i_1+2i_2+3\mu}(k+1)^{i_1+\mu}k^{i_1}H^{2i_2}z^{\alpha i_1+\alpha\mu+\frac{\alpha}{2}-2}}{\Omega_1{}^{i_1+\mu}\Omega_2{}^{2i_2+2\mu}\Gamma(i_1+\mu)\Gamma\left(i_2+\mu+\frac{1}{2}\right)i_1!i_2!}\cdot$$

$$\cdot\int_0^{+\infty}y^{2i_1+4i_2+6\mu}e^{-\frac{\mu(1+k)z^{\alpha}\Omega_2+2\mu h\Omega_1}{\Omega_1\Omega_2}y^2}e^{-\frac{\alpha^2z^{\alpha-2}y^2}{8\left(f_1{}^2+z^{\alpha}f_2{}^2\right)}\dot{z}^2}dy \tag{13}$$

LCR is defined as the rate at which a random process crosses level z in the positive or the negative direction. Using relacions (13), the average level crossing rate (LCR) of the ratio α-k-μ and η-μ random variables is [11]:

$$N_z(z)=\int_0^{+\infty}\dot{z}p_{z\dot{z}}(z\dot{z})\,d\dot{z}=\frac{\alpha^2\sqrt{2}}{e^{\mu k}\,\Gamma(\mu)\sqrt{f_1{}^2+z^{\alpha}f_2{}^2}}\cdot$$

$$\cdot\sum_{i_1=0}^{+\infty}\sum_{i_2=0}^{+\infty}\frac{\mu^{2i_1+2i_2+3\mu}k^{i_2}(k+1)^{i_2+\mu}z^{\alpha i_2+\alpha\mu+\frac{\alpha}{2}-2}h^{\mu}H^{2i_1}}{\Omega_1{}^{i_2+\mu}\Omega_2{}^{2i_1+2\mu}\Gamma(i_2+\mu)\Gamma\left(i_1+\mu+\frac{1}{2}\right)i_1!i_2!}\cdot$$

$$\cdot\int_0^{+\infty}y^{4i_1+2i_2+6\mu}dy\,e^{-\frac{\mu(1+k)z^{\alpha}\Omega_2+2\mu h\Omega_1}{\Omega_1\Omega_2}y^2}\int_0^{+\infty}\dot{z}d\dot{z}e^{-\frac{\alpha^2z^{\alpha-2}y^2\dot{z}^2}{8\left(f_1{}^2+z^{\alpha}f_2{}^2\right)}} \tag{14}$$

Resolving second integral in relacion (14), and using very well term [12]:

$$\Gamma(z)=\int_0^{+\infty}t^{z-1}e^{-t}dt \tag{15}$$

we can writing LCR of z:

$$N_z(z)=\frac{2\sqrt{2}\,h^{\mu}\pi f_m\left(\Omega_1+z^{\alpha}\Omega_2\right)^{\frac{1}{2}}}{e^{\mu k}\Gamma(\mu)}\sum_{i_1=0}^{+\infty}\sum_{i_2=0}^{+\infty}\frac{\mu^{i_1+\frac{1}{2}}(k+1)^{i_1+\mu}}{\Gamma(i_1+\mu)}\cdot$$

$$\cdot\frac{k^{i_1}H^{2i_2}\Omega_1{}^{2i_2+2\mu-\frac{1}{2}}\Omega_2{}^{i_1+\mu-\frac{1}{2}}z^{\alpha i_1+\alpha\mu-\frac{\alpha}{2}}\Gamma\left(i_1+2i_2+3\mu-\frac{1}{2}\right)}{\Gamma\left(i_2+\mu+\frac{1}{2}\right)\left((1+k)z^{\alpha}\Omega_2+2h\Omega_1\right)^{i_1+2i_2+3\mu-\frac{1}{2}}i_1!i_2!} \tag{16}$$

Probability density function of the ratio of a α-k-μ random variable and a η-μ random variable is [8]:

$$p_z(z)=\int_0^{+\infty} dy p_z(z/y)\,p_y(y)=\frac{2\alpha\sqrt{\pi}\,h^{\mu}}{\Gamma(\mu)e^{\mu k}}\sum_{i_1=0}^{+\infty}\sum_{i_2=0}^{+\infty}\frac{(\mu k)^{i_1}}{i_1!i_2!}\cdot$$

$$\cdot\frac{\Omega_1^{2i_2+2\mu}\left(\Omega_2(k+1)\right)^{i_1+\mu}H^{2i_2}\Gamma(i_1+2i_2+3\mu)z^{\alpha i_1+\alpha\mu-1}}{\Gamma(i_1+\mu)\Gamma\left(i_2+\mu+\frac{1}{2}\right)\left(z^{\alpha}(1+k)\Omega_2+2h\Omega_1\right)^{i_1+2i_2+3\mu}} \quad (17)$$

Cumulative distribution function of the ratio of a α-k-μ random variable and η-μ random variable is [13]:

$$F_z(z)=\int_0^{z}p_z(t)dt=\frac{2\alpha\sqrt{\pi}\,h^{\mu}}{\Gamma(\mu)e^{\mu k}}\sum_{i_1=0}^{+\infty}\sum_{i_2=0}^{+\infty}\frac{(\mu k)^{i_1}\left(\Omega_2(k+1)\right)^{i_1+\mu}}{i_1!i_2!\Gamma(i_1+\mu)}\cdot$$

$$\cdot\frac{\Omega_1^{2i_2+2\mu}H^{2i_2}\Gamma(i_1+2i_2+3\mu)}{\Gamma\left(i_2+\mu+\frac{1}{2}\right)}\int_0^{z}\frac{t^{\alpha i_1+\alpha\mu-1}dt}{\left(2h\Omega_1+\Omega_2(1+k)t^{\alpha}\right)^{i_1+2i_2+3\mu}} \quad (18)$$

Integral in expression (18) resolve by the form [14]:

$$\int_0^{\lambda}\frac{x^m}{\left(a+bx^n\right)^p}dx=\frac{a^{-p}}{n}\left(\frac{a}{b}\right)^{\frac{m+1}{n}}B_z\left(\frac{m+1}{n},p-\frac{m+1}{n}\right),$$

$$z=\frac{b\lambda^n}{a+b\lambda^n},a>0,\ b>0,n>0,0<\frac{m+1}{n}<p \quad (19)$$

where $B_z(a, b)$ is the incomplete Beta function, [4, Eq. 8.38]. Using term (19), we can write the expression for $F_z(z)$:

$$F_z(z)=\frac{\sqrt{\pi}}{\Gamma(\mu)e^{\mu k}}\sum_{i_1=0}^{+\infty}\sum_{i_2=0}^{+\infty}\frac{(\mu k)^{i_1}\Gamma(i_1+2i_2+3\mu)}{2^{2i_2+2\mu-1}h^{2i_2+\mu}\Gamma(i_1+\mu)}\cdot$$

$$\cdot\frac{H^{2i_2}}{\Gamma\left(i_2+\mu+\frac{1}{2}\right)i_1!i_2!}B_{\frac{(1+k)\Omega_2z^{\alpha}}{2h\Omega_1+(1+k)\Omega_2z^{\alpha}}}(i_1+\mu,2i_2+2\mu) \quad (20)$$

The average fade duration(AFD)of wireless communication system [5, 8] with dual branch SIR based is:

$$AFD=\frac{F_z(z)}{N_z(z)}=\frac{\sqrt{\pi}\sum_{i_1=0}^{+\infty}\sum_{i_2=0}^{+\infty}\frac{(\mu k)^{i_1}H^{2i_2}}{2^{2i_2+2\mu}h^{2i_2+2\mu}}}{\pi\sqrt{2}\left(\Omega_1+z^{\alpha}\Omega_2\right)^{\frac{1}{2}}\sum_{i_3=0}^{+\infty}\sum_{i_4=0}^{+\infty}\frac{\mu^{i_3+\frac{1}{2}}(k+1)^{i_3+\mu}}{\Gamma(i_3+\mu)}}\cdot$$

$$\cdot\frac{\Gamma(i_1+2i_2+3\mu)B_{\frac{(1+k)\Omega_2z^{\alpha}}{2h\Omega_1+(1+k)\Omega_2z^{\alpha}}}(i_1+\mu,2i_2+2\mu)}{\Gamma(i_1+\mu)\Gamma\left(i_2+\mu+\frac{1}{2}\right)i_1!i_2!}\cdot$$

$$\cdot\frac{k^{i_3}H^{2i_4}z^{\alpha i_3+\alpha\mu-\frac{\alpha}{2}}\Omega_1^{2i_4+2\mu-\frac{1}{2}}\Omega_2^{i_3+\mu-\frac{1}{2}}\Gamma\left(i_3+2i_4+3\mu-\frac{1}{2}\right)}{\Gamma\left(i_4+\mu+\frac{1}{2}\right)\left((1+k)z^{\alpha}\Omega_2+2h\Omega_1\right)^{i_3+2i_4+3\mu-\frac{1}{2}}i_3!i_4!} \quad (21)$$

III. NUMERICAL RESULTS

Level crossing rate of a (α-k-μ)/(η-μ) random process versus normalized crossing threshold for several values of *k* and α is shown in Fig. 1. Level crossing rate decreases for negative values of *z*, and the system has better performance when the *k* and α parameters increase. This is the consequence of less probable deep fades that cross low thresholds less frequently. For positive values of *z*, parameter *k* has negligible influence on LCR, while the increaseing α decreases LCR. Generally speaking, the system is more sensitive to changes of the non-linear parameter α.

In Fig. 2, cumulative distribution function, or outage probability, versus threshold value is presented, for several values of α-k-μ and η-μ multipath fading parameter. When the parameters *k* and μ increase for negative dB threshold values CDF decreases, while for positive values of *z*[dB], CDF saturates at outage probability of one, thus indicating total loss of connectivity – as expected for very high thresholds.

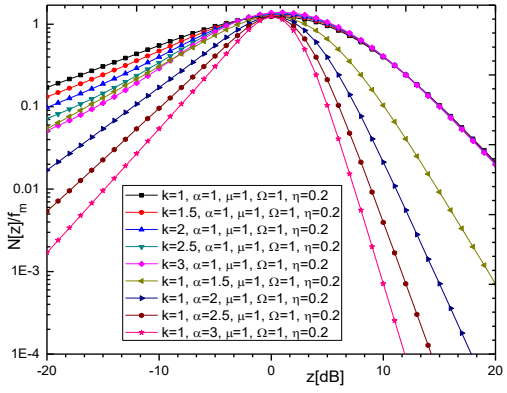


Fig. 1. Average level crossing rate of the ratio of α-k-μ and η-μ random variables

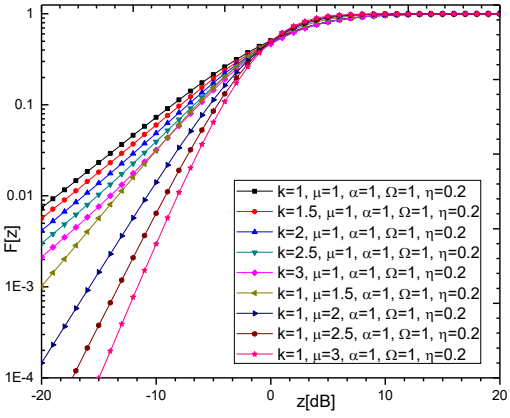


Fig. 2. Cumulative distribution function of the ratio of α-k-μ and η -μ random variables

Obtained expression for cumulative distribution rapidly convergence since 12-15 terms need to be summed in order to

rech accracly on 5th significant digit. This is illustrated by the numerical data shown in Table 1.

TABLE 1. NUMBERS OF TERMS THAT SHOULD BE ADDED IN EXPRESSION (10-15) IN ORDER TO REACH ACCURACY AT 5TH SIGNIFICANT DIGIT.

	z=-10 dB	z=0 dB	z=10 dB
k=1, μ=1, α=1, Ω=1, η=0.2	13	13	13
k=1.5, μ=1, α=1, Ω=1, η=0.2	13	14	13
k=2, μ=1, α=1, Ω=1, η=0.2	13	14	14
k=2.5, μ=1, α=1, Ω=1, η=0.2	13	14	14
k=3, μ=1, α=1, Ω=1, η=0.2	13	13	14
k=1, μ=1.5, α=1, Ω=1, η=0.2	14	15	15
k=1, μ=2, α=1, Ω=1, η=0.2	14	17	17
k=1, μ=2.5, α=1, Ω=1, η=0.2	15	17	18
k=1, μ=3, α=1, Ω=1, η=0.2	14	19	20

In Fig. 3, AFD of the ratio α-k-μ and η-μ random variables is presented, when *k* and α change. Conclusions about system performance are more obvious in Fig. 3, since the lower values of fade duration are certainly better. This situation has sense when the average signal envelope is actually above the set threshold.

Better performance is expected in cases where the parameter α increases, resulting in lower AFD. Changes to the *k* parameter also affect the change of the AFD function. Due to the increase of the parameter *k*, the AFD function decreases.

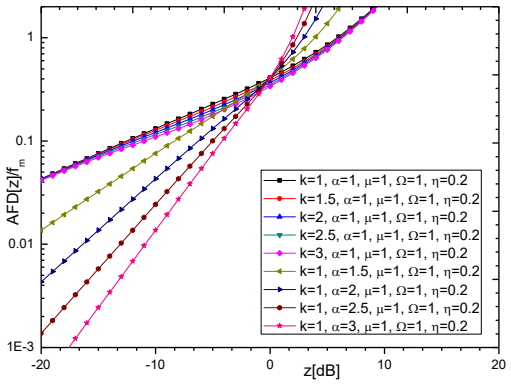


Fig. 3. Average fade duration - AFD of the ratio of α-k-μ and η-μ random variables

IV. CONCLUSION

In this paper, a research on the wireless fading channel was carried out for an environment where multipath fading can be modelled as alpha-k-mu, and interference is modelled as an eta-mu random process. The channel model is denoted as (α-k-μ)/(η-μ). We consider the case where we have fading and interference, as dominating influences, while the influence of Gaussian noise is neglected. Therefore, channel disturbance is considered to be the dominant impairment to the performance of the wireless radio system.

We have derived probability density function and cumulative distribution function of (α-k-μ)/(η-μ) random variable, as well as level crossing rate and average fade duration (of α-k-μ)/(η-μ) random process. By using these results, level crossing rate (LCR) of different, less general random processes can be calculated. The results enable analysis of influence of the α-k-μ fading severity parameter, Rician factor and nonlinearity parameter of the α-k-μ fading, η-μ faded interference nonlinearity parameter and η-μ fading severity parameter on level crossing rate and average fade duration at the wireless receiver.

ACKNOWLEDGEMENT

The paper is supported in part by the projects III44006 and TR32051 funded by Ministry of Education, Science and Technological Development of Republic of Serbia.

REFERENCES

[1] P. Spalević, S. Panić,"Analysis of wireless transmission Improvement in specific propagation environments: Monograph", Kosovska Mitrovica, ISBN 978-86-80893-52-5, 2014.

[2] Yacoub, M. D. (2007). The k-μ distribution and the η-μ distribution. *IEEE Communications Letters*, vol.9, no. 10, pp. 871-873.

[3] S. Panić, M. Stefanović, J. Anastasev, P. Spalević, Fading and Interference Mitigation in Wireless Communications, VSA: CRC Press, 2013.

[4] I Gradshteyn, I Ryzhik, Tables of Integrals, Series, and products, 1st edn, (Academic Press, New York, 1980).

[5] Z. Hadzi-Velkov, "Level Crossing Rate and Average Fade Duration of Dual Selection Combining with Cochannel", IEEE Transactions on Wireless Communications, Volume 6, Issue 11, 2007.

[6] Interference and Nakagami Fading Alan Jeffrey and Hui-Hui Dai, "Handbook of Mathematical Formulas and Integrals ", 4st edn, (Academic Press, New York, 2008).

[7] Da Costa, D. B. and Yacoub, M. D. (2007). The η-μ joint phase-envelope distribution, *IEEE Antennas and Wireless Propagation Letters*, vol. 6, no. 1, pp. 195-198.

[8] S. Suljović, D. Milić, V. Doljak, I. Marjanović, M.Stefanović, S. Milosavljević, "Level crossing rate of wireless system over cellular non linear fading channel in the presence of co-channel interference", INFOTEH-JAHORINA Vol. 16, March 2017.

[9] N. Beaulieu and X. Dong, "Average level crossing rate and average fade duration of MRC and EGC diversity in Rican fading", IEEE Transactions on Communications, vol. 51, No. 5, May 2003.

[10] B. Milosević, P. Spalević, M. Petrović, D. Vučković and S. Milosavljević, "Statistics of Macro SC Diversity System with Two Micro EGC Diversity Systems and Fast Fading", Electronics and Electrical Engineering, ISSN: 1392-1215, Vol. 96, No. 8, pp. 55-58, 2009.

[11] W. C. Y. Lee, "Statistical Analysis of the Level Crossings and Duration of Fades of the Signal from an Energy Density Mobile Radio Antenna", The Bell System Technical Journal, ISSN: 0005-8580, Vol. 46, pp. 412–448, February 1967.

[12] M. Abramowitz and I. A. Stegun, Handbook of Mathematical Function with Formulas, Graphs and Mathematical Tables, National Bureau Applied Mathematics, series 55, December 1972.

[13] M. Stefanović, D. Milović, A. Mitić, M. Jakovljević, "Performance analysis of system with selection combining over correlated Weibull fading channels in the presence of co-channel interference", AEU-International Journal of Electronics and Communications, vol. 62, Issue 9, pp. 695-700, October 2008.

[14] G. Petković, S. Panić, B. Jakšić, "Level crossing rate of macro-diversity with three micro-diversity SC receivers over Gamma shadowed Nakagami-m channel", Uiversity Thought Publication in Natural Sciences, vol. 6, No 1, pp. 55-59, , 2016.

TELECOMMUNICATION SYSTEMS AND TECHNOLOGY, AND REMOTE ECOLOGICAL MONITORING – ORAL SESSION

Use of infrared radiometry in temperature measurement of wild animals

Kalin Dimitrov[1] , Hristo Hristov[2] and Stanio Kolev[3]

Abstract – **The aim of our present work is to show the importance of infrared radiometry in conducting various animal studies in their natural living conditions. We want to show what difficulties exist when measuring from the air. We have explored numerical examples of temperature measurement at different solid angles and different wild animals.**

Keywords – **infrared radiometer, wild animals, solid angle**

I. INTRODUCTION

In modern times, infrared radiometry has entered many sectors: human and veterinary medicine, agriculture, livestock, energy, construction, security, defense and military, search and security activities, state border control, space research, etc.

Much of modern medicine such as rheumatology and dermatology uses infrared radiometry to provide diagnostic information. It is also used in the investigation of vascular dysfunction, wounds from burning, frosting as well as in the fight against cancer [1-4].

Infrared radiometry is a technique for remotely measuring temperature in a certain area of space. The aim is to capture, visualize and record changes in temperature [5-10]. Raising the surface temperature of the skin leads to an increase in body radiation. This is the result of increased metabolic activity, which in turn can be caused by inflammatory, metabolic and toxic factors. It has long been known that this may be a natural indicator of disease or a variety of deviations from normal life processes [11-14].

Infrared radiometry makes it possible to monitor and control the health status of animals in the wild [15-24]. A healthy organism is characterized by an even distribution of temperature between different parts of the body. Many pathological processes and illnesses arise through local changes in heat production, altering the pattern of blood flow in affected organs or tissues. Thanks to blood flow and the conductive transfer of heat from body depth to its surface, infrared images are thought to be able to reflect deeply the body processes [11,12].

In wildlife, radiometry can be used for early diagnosis and monitoring of severe contagious diseases such as foot and mouth rabies, for example. It was found that a rise in body temperature could occur before the onset of clinical symptoms. There are studies showing the possibilities of infrared radiometry to detect very serious and deadly diseases in animals such as plague, influenza, tuberculosis and others. It is essential for medicine, ecology, animal husbandry and all related sciences and economics to obtain quick and reliable information on the factors that affect animal populations and their interactions with the surrounding environment [13]. Infrared radiometry enables remote monitoring of the physical and physiological parameters of different animal species. It is a non-invasive, non-contacting and harmless method of visualization, both day and night, as the luminance in the visible spectrum is not essential. The ability to perform such observations on animals, in their natural habitat, from the air makes it a very preferred and important mechanism for various research.

[1]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.

[2]Hristo Hristov, E-mail: hristoveu@gmail.com

[3]Stanio Kolev, E-mail: skolev@tu-sofia.com

II. THEORY

The light propagation in a scattering and absorption medium can be strictly studied by using the classical electrodynamic theory and the partial coherence theory. We prefer to use the radiometric (energy) approach to the analysis of the respective phenomena, which is clearer in terms of physics and simpler in terms of mathematics [5,6,25]. This approach leads to the composition of the so-called propagation equation. This equation can be used only in the cases contained in the scope of proper application of radiometric concepts.

In order to analyze the situation, we use the spectral density of the elementary area emission in the whole half-space 2π sr

$$M_\lambda = \frac{(d^3\,\Phi)_{2\pi}}{dA\,d\lambda} \quad . \tag{1}$$

If the emitting body is in thermal equilibrium with the environment (1), the Planck law is used (Planck law for black body radiation)

$$M_\lambda^*(\lambda,T) = \frac{c_1}{\lambda^5\left[exp\left(\frac{c_2}{\lambda T}\right)-1\right]} \tag{2}$$

where the constants are defined [25] :

$$c_1 = 2\pi c^2 h = 3{,}74.10^{-16} Wm^2 \ ,$$
$$c_2 = ch/k_B = 1{,}44.10^{-2} mK \ .$$

In the expression, the * denotes that we refer to a black body.

For a better analysis of the situation, in the space we use a dimension which is related not only to the elementary radiation area like in (1) but also to the solid view angle and observation angle

$$L_\lambda = \frac{d^5 \Phi}{d\Omega \, dA \, d\lambda \, cos\theta} \quad . \tag{3}$$

If the radiation area is isotropic, then the relation between the formulae is simplified to

$$L^*_\lambda(\lambda, T) = \frac{1}{\pi} M^*_\lambda(\lambda, T) \quad . \tag{4}$$

In nature, bodies are more or less different from the idealization of the black body. Therefore, we use the emissivity coefficient

$$\varepsilon(\lambda, T) = \frac{M_\lambda(\lambda,T)}{M^*_\lambda(\lambda,T)} \quad . \tag{5}$$

When we are considering a situation on Earth [26-28], it is always necessary to take into account the influence of the atmosphere. As an example, we will study the atmospheric channel of one specific radiometric system situated between the emitting surface and the aperture of the receiving antenna (fig.1).

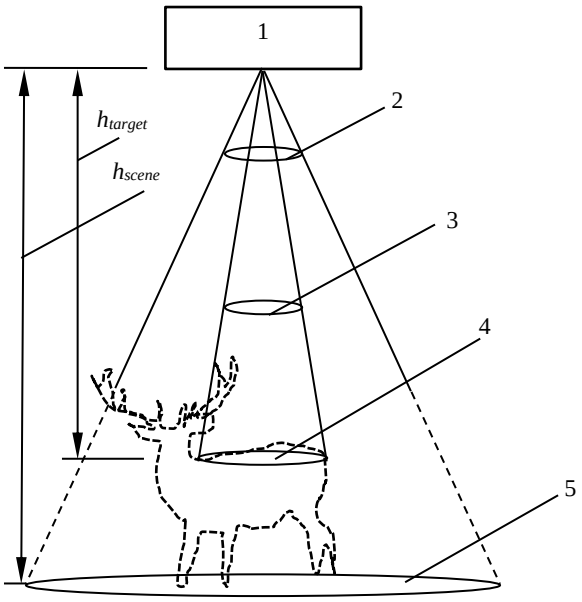


Fig. 1. General set-up of a wild animal radiometric investigation (1-radiometric device; 2-solid view angle of the radiometric device; 3-solid view angle of the animal investigated; 4-essential area of the wild animal surface which produces radiation (target); 5-radiation from the surface on which the animal is located (scene), h_{target} - distance to the target, h_{scene} – distance to the scene)

The propagation environment is characterized by volume coefficients of scattering and absorption, with an ensemble scattering indicatrice and the respective temperature. In this paper we will not consider the polarization effects during light propagation.

In the volume between the source and the measuring device, we can observe interactions of optical radiations with the atmospheric substance and energy transformations. The most essential ones among them are the scattering and absorption. The unified effect termed extinction (the reducing of energy) is often used.

Another important effect is that the substance in the volume between the source and the measuring device is a source of heat radiation whose energy is distributed along the spectral scale, depending on the temperature in that volume. One part of this radiation is directed to the receiver and added to the radiation of the object investigated.

The third essential effect is that, in the volume between the source and the receiver, it is possible for other radiations to enter, which are propagated along different spatial axes. Some of them can be determined by independent outside sources, while others are the result of multiple scattering of the radiations from the aerosol particles and the molecules of the gases.

The task of propagation can be formulated as a differential equation [25]

$$\frac{dL_\lambda}{dz} = -\left(\alpha^{(s)} + \alpha^{(a)}\right) L_\lambda + \alpha^{(a)} L^*_\lambda + \alpha^{(s)} L_{\lambda_{scattered}} \tag{6}$$

where $\alpha^{(s)}$ is the atmosphere scattering coefficient, $\alpha^{(a)}$ is the atmosphere absorption coefficient. $L_{\lambda_{scattered}}$ is the ingredient determined by the scattering from other sources. We will assume, in our case, that this ingredient is too low and rewrite (6) without it.

$$\frac{dL_\lambda}{dz} = -\left(\alpha^{(s)} + \alpha^{(a)}\right) L_\lambda + \alpha^{(a)} L^*_\lambda \, . \tag{7}$$

In order to derive the flux which would enter the photoreceiver, we have to solve (7) and multiply by the values of the aperture and the solid angle

$$\Phi_{\lambda,r} = A_r \, \Omega_r \, L_\lambda(Z) \, . \tag{8}$$

We derive the solution of the differential equation by assuming that the atmospheric scattering and absorption coefficients are constants in terms of space and spectrum. This is possible because the distance is relatively small and we are also using a small part of the spectrum.

We also assume that the sources of radiation are isotropic.

After these assumptions, we derive an expression for the optical flux, which enter the receiver (we use (1)..(8))

$$\Phi_r = A_r \, \Omega_r \, \tau_r \left\{\varepsilon_t \left\{exp\left[-\left(\alpha^{(s)} + \alpha^{(a)}\right)Z\right]\right\} U(T_t; \lambda_1, \lambda_2) + \right.$$
$$\left. + \frac{\alpha^{(a)}}{\alpha^{(s)} + \alpha^{(a)}} \left\{1 - exp\left[-\left(\alpha^{(s)} + \alpha^{(a)}\right)Z\right]\right\} U(T_a; \lambda_1, \lambda_2)\right\} \tag{9}$$

where

$$U(T_t;\lambda_1,\lambda_2)=\int_{\lambda_1}^{\lambda_2} L_\lambda^*(\lambda,T)d\lambda \quad (10)$$

and τ_r is the transparency of the receiver's optics.

We can summarize that the flux is formed by two main ingredients in (9): the one from the targets and the one from the atmosphere

$$\Phi_r = \Phi_t + \Phi_a \ . \quad (11)$$

III. Numerical investigation

In order to make the general theory useful, we will consider a specific example with a radiometric investigation of a deer. The solid angle of the radiometric device is fixed. The angle, however, at which we see the considerable part of the animal's body that we are interested in, changes depending on the height. For this reason, in order to make a more accurate assessment of the entering of different optical fluxes in the entrance aperture of the receiver, we modify (11) to

$$\Phi_{t,total}=\Phi_t\left(\Omega=\Omega_t\ ;Z=h_t\right)+$$
$$+\Phi_{scene}\left(\Omega=\Omega_r-\Omega_t\ ;Z=h_{scene}\right) \quad (12)$$

and

$$\Phi_{a,total}=\Phi_{a,t}\left(\Omega=\Omega_t\ ;Z=h_t\right)+$$
$$+\Phi_{a,scene}\left(\Omega=\Omega_r-\Omega_t\ ;Z=h_{scene}\right) \quad (13)$$

In order to derive a plausible numerical simulation, we use the literature about the respective wild animal, in this case the deer. Most of the objects investigated have a complex shape, which leads to measurement errors. Bodily surfaces also have a serious impact. The presence of fur has a significant influence on the surface temperature of the body and, accordingly, is of significant importance. Animals with thick fur are poorly visible, while those whose body is barely covered with fur are more suitable for research. The density of the fur, the individual features of the fur, its length and purity have a significant impact. The purity of the surface of the examined tissues is important, as is their coloring. For example, the presence of moisture, grease, dust adhesion or other physical particles on the surface of the skin affect the measurement results.

Behavioral factors also have an impact. Observations and studies of wildlife are accompanied by many problems due to the unique biological and behavioral traits of different species, such as hiding in hiding places. The behavior and reactions of wildlife species resulting from changes in environmental factors cannot be predicted. Serious influence is also caused by intrinsic factors such as sweat evaporation, vascular perfusion, local tissue metabolism. The stress factor is essential and can seriously affect the results of the research.

For this reason, radiometric analyses should be performed in a natural environment when the subject is adapted to the environment.

For the simulation, we select the following data [28,29]: $T_a = 293K$, $T_t = 313K$, $T_{scene} = 293K$, $\lambda_1 = 8\mu m$, $\lambda_2 = 13\mu m$, $\varepsilon_t = 0{,}99$, $\varepsilon_{scene} = 0{,}98$, $\alpha^{(s)} = 0{,}03km^{-1}$, $\alpha^{(a)} = 0{,}4km^{-1}$, $\Omega_r = 6.10^{-3}sr$, $A_r = 8.10^{-5}m^2$, $\tau_r = 1$.

Using (2) and (4) we perform numerical integration of expression (10). For this we use Scilab [30]. After that we perform calculations for (12) and (13). Part of the results are shown in the following Table I.

TABLE I
PART OF SIMULATION DATA

No	1	2	3	4
$h_{scene}[m]$	10,13	12	14	16
Φ_t $[W]$, T=313K	2,7027 .10^{-05}	1,8598 .10^{-05}	1,3304 .10^{-05}	9,9848. 10^{-06}
$\Phi_{all\ other}$ [W], T=293K	7,3266 .10^{-08}	6,1923 .10^{-06}	1,0035 .10^{-05}	1,2445 .10^{-05}

We are particularly interested in the comparison of the flux coming from the target with the sum of the fluxes from the atmosphere and the flux from the scene. The results are shown graphically in figure 2.

IV. Conclusion

Applying high-level infrared imaging opens up new serious possibilities. The method allows for large areas and many objects to be covered. It is a reliable mechanism for counting populations, detecting and registering habitat areas, exploring migratory processes, identifying the ways and rules for moving in large herds, studying the natural habitat of certain groups, and animal breeds in the wild. In this case, it is important to note that radiometry is used not for accurate temperature measurements but for detecting objects with temperatures significantly different from those of the surrounding environment (fig.2). It also provides us with a

good opportunity to find newborns or small ones hidden in the forest.

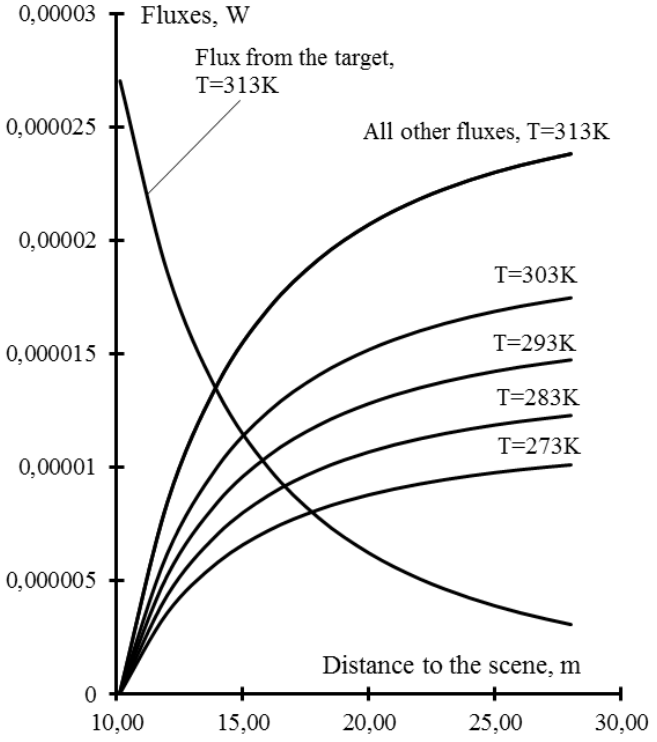


Fig. 2. Simulation for one type of target with temperature 313K and emissivity coefficient 0,99 (wild animal) and different temperatures for a scene with emissivity coefficient 0,98 and atmosphere.

REFERENCES

[1] M. Soroko, P. Cwynar, K. Howell, K. Yarnell, K. Dudek, D. Zaborski, "Assessment of Saddle Fit in Racehorses Using Infrared Thermography", Journalof Equine Veterinary Science, Elsevier, vol. 63, pp.30–34, 2018.

[2] B.B. Lahiri, S. Bagavathiappan, T. Jayakumar, J. Philip, "Medical applications of infrared thermography: A review", Infrared Physics & Technology, Elsevier, vol. 55, Issue 4, pp. 221-235, 2012.

[3] E. Faye, O. Dangles, S. Pincebourde, "Distance makes the difference in thermography for ecological studies", Journal of Thermal Biology, Elsevier, vol. 56, pp. 1–9, 2016.

[4] J. Cilulko, P. Janiszewski, M. Bogdaszewski, E. Szczygielska, "Infrared thermal imaging in studies of wild animals", European Journal of Wildlife Research, Springer, vol. 59, Issue 1, pp.17–23, 2012.

[5] M. Schlessinger, *Infrared Technology Fundamentals*, CRC Press, 1994.

[6] J. Miller, *Principles of Infrared Technology: A Practical Guide to the State of the Art*, Springer Science & Business Media, 2012.

[7] M. Diakides, J. Bronzino, D. Peterson, *Medical Infrared Imaging: Principles and Practices*, CRC Press, 2012.

[8] G. Zibordi, C. Donlon, A. Parr, *Optical Radiometry for Ocean Climate Measurements*, Academic Press, 2014.

[9] Z. Zhang, B. Tsai, G. Machin, *Radiometric Temperature Measurements: II. Applications*, Academic Press, 2009.

[10] W. Wolfe, *Introduction to Radiometry*, SPIE Press, 1998.

[11] G.Tattersall, "Infrared thermography: A non-invasive window into thermal physiology (Review) ", Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology, Elsevier, vol. 202, pp. 78-98, 2016.

[12] S. Hilsberg-Merz, *Zoo and Wild Animal Medicine (Sixth Edition), Ch.3 - Infrared Thermography in Zoo and Wild Animals*, Elsevier, 2008.

[13] A. Perez de Diego, P. Sanchez-Cordón, M. Pedrera, B. Martínez-López, J. Gomez-Villamandos, J. Sánchez-Vizcaino, "The use of infrared thermography as a non-invasive method for fever detection in sheep infected with bluetongue virus", The Veterinary Journal, Elsevier, vol.198, pp.182–186, 2013.

[14] D. Gates, R. Schmerl, *Perspectives of Biophysical Ecology*, Springer Science & Business Media, 2012.

[15] E. Hodnett, "Thermal Imaging Applications In Urban Deer Control", Wildlife Damage Management Conferences Proc., vol. 106, pp.141-148, 2005.

[16] K. Havens, E. Sharp, *Thermal Imaging Techniques to Survey and Monitor Animals in the Wild: A Methodology*, Academic Press, 2015.

[17] S. Jorgensen, *Thermodynamics and Ecological Modelling*, CRC Press, 2000.

[18] D. Hewitt, *Biology and Management of White-tailed Deer*, CRC Press, 2011.

[19] K. Havens, E. Sharp, *Thermal Imaging Techniques to Survey and Monitor Animals in the Wild: A Methodology*, Academic Press, 2015.

[20] E. Faye, O. Dangles, S. Pincebourde, "Distance makes the difference in thermography for ecological studies", Journal of Thermal biology, Elsevier, vol.56, pp.1-9, 2016.

[21] R. Kotrba, I. Knizkova, P. Kunc, L. Bartos, "Comparison between the coat temperature of the eland and dairy cattle by infrared thermography", Journal of Thermal Biology, Elsevier, vol.32, pp.355–359, 2007.

[22] N. Weissenbock, C. Weiss, H. Schwammer, H. Kratochvil, "Thermal windows on the body surface of African elephants (Loxodonta africana) studied by infrared thermography", Journal of Thermal Biology, Elsevier, vol.35, pp.182–188, 2010.

[23] C. Thompson, C. Scheidel, K. Glander, S. Williams, C. Vinyard, "An assessment of skin temperature gradients in a tropical primate using infrared thermography and subcutaneous implants", Journal of Thermal Biology, Elsevier, vol.64, pp.49-57, 2017.

[24] J. Beaver, C. Harper, R. Kissell Jr., L. Muller, P. Basinger, M. Goode, F. V. Manen, W. Winton, M. Kennedy, "Aerial Vertical-Looking Infrared Imagery to Evaluate Bias of Distance Sampling Techniques for White-Tailed Deer", Wildlife Society Bulletin, Wiley, vol. 38, No. 2, pp. 419-427, 2014.

[25] E.Ferdinandov, Basics of optoelectronics, Technika, 1993 (in Bulgarian).

[26] J. Hodgson, S. Baylis, R. Mott, A. Herrod, R. Clarkeet, "Precision wildlife monitoring using unmanned aerial vehicles", Scientific Reports, vol.6, 22574; doi: 10.1038/srep22574, pp.1-7, 2016.

[27] A. Entropa, A.Vasenev, "Infrared drones in the construction industry: designing a protocol for building thermography procedures", Energy Procedia, Elsevier, vol. 132, pp. 63-68, 2017.

[28] E. Dereniak, G. Boreman, *Infrared detectors and systems*, Wiley, 1996.

[29] A. Moen, "Surface temperatures and radiant heat loss from white-tailed deer",The Journal of Wildlife Management, Allen Press, vol. 32, No. 2, pp.338-344, 1968.

[30] C. Gomez, *Engineering and Scientific Computing with Scilab*, Springer Science & Business Media, 2012.

A Comparative Study of Routing Protocols for 6LoWPAN

Ravinder Beniwal[1], Kamelia Nikolova,[1] and Georgi Iliev[1]

Abstract – **In this paper a comparative study and analysis of various routing protocols designed for 6LoWPAN networks are performed. The demands of resource constrained devices imposing the restrictions in the choice of the routing mechanism are also considered. After comparison, some recommendations are done.**

Keywords – **Wireless sensor networks, IPv6, Routing protocols, 6LoWPAN, Constrained Devices.**

I. Introduction

Wireless sensor network (WSN) is one of the fastest growing area in networking today. WSN finds numerous applications in different areas like home automation, health-care, industrial and environment monitoring. Most of these applications can have real impact into monitored system if gathered data can be processed and controlled over Internet. In order to meet the requirements of resource constrained devices used in WSN and to give the possibility to interconnect them to the Internet, IPv6 low power personal area network (6LoWPAN) is introduced [1]. 6LoWPAN can be considered as an adaptation layer above the IEEE 802.15.4 link layer with the main aim to help packet fragmentation and reassembly. The major function of the adaptation layer is not only the TCP/IP header compression but routing also.

During the last decade, a number of different routing protocols developed especially for 6LoWPAN has been proposed in the literature [2]-[15]. Although, the number of routing protocols are proposed, some of them are not adopted and are not become an accepted standard. Different types of routing mechanisms are used in 6LoWPAN based on the topologies and applications running over it. Mobile nodes and dynamically adaptive topologies should be allowed in 6LoWPAN routing protocols. The main aim of this paper is to present, study, compare and evaluate different adaptive and modular approach to routing in 6LoWPAN networks and to classify them. After a carefully analysis and comparison of the possible routing protocols the best solutions and recommendations are presented.

This paper is organized into 5 sections. The 6LoWPAN technology constraints are considered in Section II. Different routing protocols are analysed in Section III. In Section IV, a classification is proposed and a comparison of the routing protocols is performed. Concluding remarks are presented in Section V.

[1]The authors are with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mails: ravin.beniwal29@gmail.com, ksi@tu-sofia.bg and gli@tu-sofia.bg.

II. 6LoWPAN Limitations

There are certain considerations for the 6LoWPAN while communicating with the other IP based networks. As the number of the nodes grows in the 6LoWPAN network, the auto configuration and statelessness are strongly required, thus 6LoWPAN network interconnects to the other IP networks by IPv6 [1].

The 6LoWPAN architecture is given in Fig. 1. As it can be seen the IEEE 802.15.4 physical and MAC layers are adopted and IEEE 802.15.4 frame format must be followed. The limited packet size must be taken into account while designing routing protocols for such type of network. As the size increases more than the 127 bytes, it poses challenges for the low end 6LoWPAN nodes, as they do not have enough RAM or memory to accommodate 1280 bytes IPv6 packet size [1].

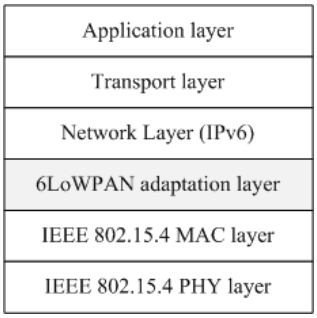


Fig. 1. 6LoWPAN architecture.

Because of the sustainable interest towards the WSN and also the services they offer, the number of the 6LoWPAN nodes increase significantly and these devices have limited input and display capabilities as well as can be located in hard to access locations. Thus the protocols used in 6LoWPAN devices should have minimal configuration, be easy to bootstrap and enable the network to self-heal of these nodes for inherent unreliable characteristics. The size constraint for the link layer protocol should also be considered. The management of network should have little overhead, but powerful enough to control dense deployment of the nodes in the network [1].

Service discovery protocols are required by the 6LoWPAN network to discover, control and maintain services provided by the nodes. Confidentiality and integration protection are required by the 6LoWPAN applications, which are provided by the all layers above the PHY layer. Small code size, low power consumption, low complexity and small bandwidth are the some of the prevailing constraints that can affect the choice of a particular protocol.

Location information of the nodes in the network is used to determine the place of the occurrence of the phenomenon and it also helps for the development of energy efficient routing protocols. Location information and IP addresses can be used to reduce the overhead of the transmission. Different challenges for the 6LoWPAN networks like failure of the node due to lack of power, physical damage or environmental interference leads to rerouting or re-organization of the network.

III. Routing Protocols

After a deep study of the existing literature, it can be seen that the developing of new routing protocols addressing all of the 6LoWPAN issues is still an open research area. The existing routing protocols could be classified based on different criteria as it was given in [2]-[15]. In this paper new classification is proposed leading to more clearly and understandable way to explain and compare the features of the existing routing protocols. Four main groups of routing protocols as it was given on Fig. 2 could be formulated.

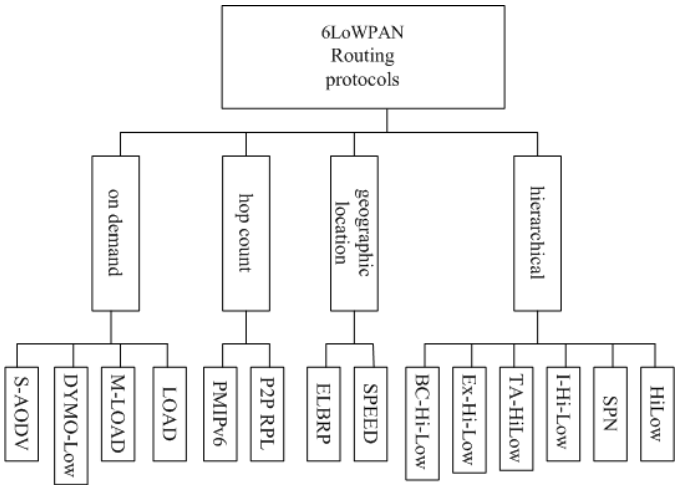


Fig. 2. Classification of 6LoWPAN routing protocols.

A. On demand routing protocols

6LoWPAN Ad-Hoc On-Demand Distance Vector Routing (LOAD) is a simplified on demand routing protocol, defined to be working on top of the adaption layer instead of transport layer [2]. The basic operations of LOAD are route discovery, managing data structures and maintaining local connections. For these operations, LOAD maintains the following two tables: the routing table (storing information such as destination, next hop and status) and the route request table (storing the temporary route information used in the route discovery process) [2]. In order to reduce the size of control messages and to simplify the route discovery process, the use of destination sequence number are omitted. Route Reply message (RREP) is generated only by destination of a route ensuring loop free condition. If a local repair fails, Route Error message (RERR) is forwarded only to the originator of the failed data delivery, thus no requiring to use the precursor list. LOAD uses number of hops and Link Quality Indicator (LQI) route metrics from source to destination to determine the strongest route. It also uses the Acknowledgement messages (ACK) to ensure guaranteed delivery of packets.

In order to overcome the energy consumption problem in LOAD, Multipath LOAD (M-LOAD) protocol is proposed in [3]. Energy consumption is high in LOAD protocol due to repeated broadcast of Repeat Request message (RREQ) for route discovery process. M-LOAD is designed to reduce the network overhead. It enhances the LOAD protocol to find multipath routes during path discovery process by implementing the Ad-hoc on-demand multipath distance vector routing (AOMDV) on LOAD.

Dynamic MANET on-demand routing for 6LoWPAN (DYMO-Low) is another routing protocol operating on the link layer directly and thus creates a mesh topology with the 6LoWPAN devices, so that IP can see the WPAN as a single link [4]. DYMO-Low provides an easy and effective method of implanting routing protocol. It uses both 16 bit short addresses and 64 bit extended addresses.

Sink Routing Table over Ad hoc On Demand Distance Vector (S-AODV) designed in [5] reduces the power consumption, provides load balancing and enhances the network lifetime. Routing table is maintained by the Sink node and it forwards the query packets to the specific internal node using the routing table. Response of the destined node to the query of sink node is given through the neighbouring node. It has setup phase and steady state phase, sink node broadcast its status to the nodes in the network and every node in the set-up phase establishes its path to the sink node through the optimal neighbour node. Sink node uses this information to construct a Sink Routing Table (SRT). Data is transferred in the steady-state phase between the sink node and destined common node [5]. With the help of this mechanism, the delay and the energy consumption is minimised during data forwarding in the network.

B. Hierarchical routing protocols

Hierarchical Routing in 6LoWPAN (HiLOW) is developed to increase the scalability and uses the 16 bit short address as interface identifier for memory saving and larger scalability [6]. It exhibits parent child model and every node in the network discovers its parent by sending the broadcast signal. If the node finds a 6LoWPAN parent node in its Personal Operating Space (POS), it gets associated by the parent node using 16 bit short address otherwise it configures itself as coordinator. The different child node receives a 16 bit short address from its parent node in the network. When a node wants to send a packet to destination, it determines the next hop node to forward the packet. If a link failure occurred in the network, no route recovery path mechanism is performed to repair the route resulting in unguaranteed packet delivery of packets in the network.

Step Parent Node (SPN) [7] is a new path recovery algorithm developed to improve the existing HiLow protocol in [6]. When a link break happens in the network, the child node of the failure parent node broadcasts a step parent node request message to his neighbouring nodes, which unicast a step parent node message reply to the sender, if he has the child nodes value less than its MC value (MC=4). If the

requesting node gets more reply messages, then it will check the Path Quality Indication (PQI) and address of different senders. It will get associated with that node, which has high PQI and not the descending of the sender. To improve the network scalability, 16 bit short addresses are used in this algorithm.

Improved HiLow (I-Hi-Low) [8] is used to increase the efficiency of the protocol. The current node broadcast "Hello" message in its Personal Operating Space (POS) to acquire information about its neighbours in this Improved High Low protocol. The current node "C" calculates its Parent address using the equation "AP= [(AC-1)/MC]" after receiving the packet from neighbouring node. Improved hierarchical routing protocol takes minimum hop counts to reach its destination while comparing with existing hierarchical routing protocols.

Light weight address allocation and addressing schemes are the main features of hierarchical routing protocol. A hierarchical tree is established between parent and child node to transmit the packets. Address allocation and routing mechanism are the main problems in the existing hierarchical routing protocol. A new mechanism is suggested in Bias Routing Tree Avoiding Technique for Hierarchical Routing Protocol for 6LoWPAN (TA-HiLow) [9] to avoid the bias routing tree that could happen if the child node do not attached to the parent node evenly. Attached child number information is transmitted to avoid the bias routing tree problem.

Extended Hierarchical Routing (Ex-Hi-Low) [10] mechanism configure a hierarchical routing tree in this protocol and if a sensor or parent node fails due some reasons the routing tree structure cannot be maintained. New child node sends a packet and destination node ID to parent and parent sends the path information of the destination to coordinator node. The coordinator node sends this information to router nodes to check their routing table for Neighbour Added Child (NAC) information and router sends the reply to the coordinator node. After this process the new child node sends the packet to the destination via his new parent node.

If a child node has more than one potential parent node and the child attaches to the first responding parent node, this situation leads to uneven or biased child distribution system and short span of life for the 6LoWPAN network.

The new developed protocol Bias Child Node Association Avoidance Mechanism for Hierarchical Routing Protocol in 6LoWPAN (BC-Hi-Low) avoids a biased hierarchical routing tree structure considering the potential parent node's depth, energy level and signal strength [11]. The potential parent node provides the child node with its existing child node count, by this count the new node selects its parent with less child nodes. The performance of this protocol [11] will be better if the parent node has same energy level, same depth and different child nodes, if it is same then it again leads to bias child node association. In this mechanism the new child is provided the depth and the average amount of power of the parent node. The average amount of the power is calculated by the equation Avg = CBP/(CC+2) where CBP is current battery power and CC is current child node.

C. Geographic location based routing protocols

SPEED Routing Protocol in 6LoWPAN Networks is designed for providing soft real time communication in 6LoWPAN networks. Geographic position and global addresses are used to identify the packets sent toward it destination nodes and the destination area is recognised by the central position and radius. Shortest path is used to send the nodes towards the destination and SPPED [12] supports for load balancing, soft real time and flow shaping mechanism for making this protocol an efficient solution in the 6LoWPAN networks for packet routing.

In Enhanced Location Based Routing Protocol for 6LoWPAN (ELBRP) [13] protocol the geographic location is used as a routing metric for transmission of packets in the 6LoWPAN network. The sink node broadcast its location information and address during the network initialisation, the other LER and RFD nodes of the network uses this location information, distance and LQI as a routing metric for communication. LER sends a RREP in unicast manner, in reply the source node sends the RREQ message and only the nearby LERs reply for that node and neighbour table information is filled in the source node. Between the LERs based on maximum distance, the best LER is chosen and this LER reply the RREP message. Each node of the network maintains a routing table and neighbour table. Routing table of node contains ER address, ER location, source address and source location and neighbour table of nodes stores the ER address, its location and LQI. Energy consumption of this protocol is also very low.

D. Hop count based routing protocols

Point to Point Routing Protocol for Low-Power and Lossy Networks (P2P RPL) [14] allows router to discover and establish paths to another router, based on a reactive mechanism. When a router S needs to discover a path to another router D, the first router S originates a message similar to an AODV- Route Request. This protocol has introduces a new destination-oriented directed acyclic graph (DODAG) Information Objects option that specifies an address which should be discovered and records the traversed path.

Different protocols for 6LoWPAN mobile sensor node (6LoWMSN), based on Proxy Mobile IPv6 (PMIPv6) [15] have been introduced. The PMIPv6 standard supports only single hop networks and can't be applied directly to multi hop based 6LoWPAN. This protocol does not support the mobility of 6LoWMSN and 6LoWPAN gateways and can't detect the attachment of PAN to 6LoMSN. The mobility in the multi hop based 6LoWPAN networks is introduced with the movement notification of a 6LoWMSN. Router solicitation (RS) and router advertisement (RA) messages are used to reduce the signalling cost over the wireless link with the attachment of 6LoMSN. PAN attachment detection scheme for 6LoMSNs is defined to apply the single hop based PMIv6 protocol to multihop based 6LoWPAN networks by using router solicitation (RS) and router advertisement (RA) messages.

TABLE I
ROUTING PROTOCOL COMPARISON

features	M-LOAD	S-AODV	SPN	BC-HILOW	HI-LOW	PMIV6	SPEED	ELBRP
RERR msg	use	use	no use	no use	no use	no use	no use	no use
energy usage	low	low	low	low	low	high	low	very-low
broadcasting rreq	high	reduced	high	reduced	high	reduced	reduced	reduced
sequence no	use	no use	no use	no use	no use	use	no use	no use
hop count	use	use	no use	no use	use	use	use	use
hello msg	no use	no use	no use	no use	no use	no use	no use	no use
node mobility	mobile	mobile	mobile	mobile	static	mobile	static	mobile
convergence to topology	fast	fast	slow	fast	slow	slow	slow	fast
PQI	no use	no use	use	no use	no use	no use	no use	use

IV. COMPARISON OF ROUTING PROTOCOLS

Because of the limited size of the paper, only some of the routing protocols are compared and the results are presented in the Table I. It is observed in our comparison that LOAD, DYMO-Low, M-LOAD, and S-AODV protocols utilize the RERR message to indicate the link breakage in the networks, while the other protocols not utilized this feature. All the protocols consumes very less energy and LOAD, M-LOAD, DYMO Low, Hi-Low and SPN broadcasts the RREQ for route discovery. DYMO-Low uses sequence number for freedom of loops while other protocols do not use this feature. Hello message is used only in DYMO-Low and I-Hi-Low for tracking the neighboring nodes constantly. Hi-Low, M-LOAD, S-AODV, I-Hi-Low and ELBRP uses the concept of hop count as a routing metric and the process of local repair is used in LOAD to determine alternate link for data forwarding while alternate path as identified in case of M-LOAD.

All the nodes support node mobility except the SPEED protocol and the mobility of sink is addressed in S-AODV in contrast to other protocols. In the protocols Hi-Low, SPN, I-Hi-Low, TA-Hi-Low, Ex-Hi-Low and BC-Hi-Low scalability analysis has been performed in comparison with other protocols. DYMO-Low has high routing delay as compared with other protocols. SPN and ELBRP utilized Path Quality Indication (PQI) as routing parameter as compared to other protocols. HiLow, SPN, and SPEED has slow convergence to varying topology in comparison with other protocols. P2P-RPL and PMIV6 has used multi hop as a routing metric while comparing with other protocols in the 6LoWPAN networks.

V. CONCLUSION

In this paper we have analyzed various routing requirements in the 6LoWPAN networks and also tried to understand the architecture. We have also tried to classify the routing protocols in the 6LoWPAN network and we found that four main groups of protocol could be formulated. We found many issues still need to be addressed in this area and the routing algorithm need to be optimized for different reasons.

REFERENCES

[1] N. Kushalnagaret at al, IPv6 over Low-Power Wireless Personal Area Networks (6LoWPANs): Overview, Assumptions, Problem Statement, and Goals, IETF RFC 4919, 2007.

[2] K. Kim at al., 6LoWPAN Ad hoc On-Demand Distance vector Routing (LOAD), IETF, draft daniel-6lowpan-load-adhoc-routing-03.txt, pp. 1-17, 2007.

[3] J. Chang, T. Chi, H. Yang, H. Chao, The 6LoWPAN ad-hoc on demand distance vector routing with multi-path scheme, IET Intern. Conf. on Frontier Computing. Theory, Technologies and Applications, pp. 204–209, August 2010.

[4] K. Kim, S. Park, I. Chakeres, C. Perkins, Dynamic MANET On demand for 6LoWPAN (DYMO-low) Routing, IETF, draft montenegro-6lowpan-dymo-low-routing-03, 2007.

[5] C. Zhongyu, L. Gang, S-AODV: Sink Routing Table over AODV Routing Protocol for 6LoWPAN, Second International Conference on Networks Security, pp. 340-343, 2010.

[6] K. Kim, S. Yoo, J. Lee, G. Mulligan, Hierarchical Routing over 6LoWPAN (HiLow), IETF, draft-daniel-6lowpan-hilow-hierarchical-routing-01.txt, pp. 1-12, 2007.

[7] G. Ee, Ch. Ng, N. Noordin, and B. Mohd. Ali, Path Recovery Mechanism in 6LoWPAN Routing, IEEE Conference on Computer and Communication Engineering, pp. 1-5, 2010.

[8] H. Yu, J. He, Improved Hierarchical Routing over 6LoWPAN, IEEE, pp. 377-380, 2011.

[9] H.-J. Lim, T.-M. Chung, The Bias Routing Tree Avoiding Technique for Hierarchical Routing Protocol over 6LoWPAN, IEEE Computer Society, pp. 232-235, 2009.

[10] Ch.-S. Nam, H.-J. Jeong, D.-R. Shin, Extended Hierarchical Routing, 5th Intern. Joint Conf. on INC, IMS and IDC, 2009.

[11] L. Chandra, K.-S. Chai, M. Abu Al-Haj, S. Ramadass, Bias Child Node Association Avoidance Mechanism for Hierarchical Routing Protocol in 6LoWPAN, IEEE, pp. 332-335, 2010.

[12] S. Bochhino, M. Petracca, P. Pagano, M. Ghibaudi, F. Lertora, SPEED Routing Protocol in 6LoWPAN Networks, IEEE 16th Conference on ETFA, pp. 1-9, 2011.

[13] M Rehenasulthana, P. Bhuvaneswari, N. Rama, Enhanced Location based Routing Protocol for 6LoWPAN, Intern. Journal of Comp. Networks & Communications, vol.4, No.3, May 2012.

[14] E. Baccelli, M. Philipp, The P2P-RPL Routing Protocol for IPv6 Sensor Networks: Testbed Experiments, 19th Intern. Conf. on Software, Telecommunications and Computer Networks, Sep 2011, Split, Croatia. pp.1 - 6, 2011.

[15] S. Gundavelli, K. Leung, V. Devarapalli, K. Chowdhury, B. Patil, Proxy Mobile IPv6. IETF, RFC 5213, August 2008.

Comparing the performance of SNMP to Network Telemetry streaming with gRPC/GPB

Ivan Ivanov[1]

Abstract – **This paper compares the performance of SNMP based network monitoring to Network Telemetry streaming with remote procedure call framework (gRPC) and Google Protocol Buffers (GPB). The study focuses on retrieving the** ifTable **from network elements using both SNMP and gRPC/GPB based approach. First, we evaluate the performance of SNMP using GetNext and GetBulk messages. Then, we get the same** ifTable **information encoded with Google Protocol Buffers using TCP and UDP for transport. The performance is then measured as a function of the number of retrieved objects. Several aspects are examined: bandwidth usage, round trip times and CPU.**

Keywords –**SNMP, gRPC, Network Telemetry, Google Protocol Buffers, Pipeline, XRV9k**

I. INTRODUCTION

Streaming network telemetry is a new paradigm in networking management and monitoring which hasn't been widely deployed yet. It utilizes the idea of "push the data" not "pull the data". To retrieve any information from a network device using SNMP, NMS (Network Management System) needs to first request this data in form of an SNMP request. Only then the data can be sent from the network device back to the NMS in form of SNMP response message/s. This is repeated every polling interval. To retrieve large amounts of data, SNMP polling relies on the GetBulk operation. It performs a continuous GetNext operation that retrieves all the columns of a given table (e.g. *ifTable*). The network device will return as many columns from the *ifTable* as can fit into a single packet. If the polling NMS detects that the end of the table has not yet been reached, it will do another GetBulk and will repeat the operation until the whole *ifTable* is fetched.

Streaming network telemetry gains efficiency over SNMP by eliminating the polling process altogether. Instead of sending SNMP requests with specific instructions that the network device must process every time, telemetry uses a configured policy on the device to know what data to collect, how often and to which NMS it should be sent.

This paper focuses on retrieving the *ifTable* from network devices using SNMP polling and at the same time streaming the same network information encoded with Google Protocol Buffers and compare the results.

II. RELATED WORKS

In literature several papers can be found that investigates the performance of SNMP [1] and compares the performance of SNMP to Web Services/XML-based monitoring systems [2]. Also, the performance of SNMP trap notification was directly compared to Web Services notifications using XML gateways [3]. Few papers have been published that discusses the performance of SNMP in large-scale deployments and provide analysis of the traffic patterns of large-scale monitoring systems [4].

Streaming network telemetry is a new approach for monitoring and managing communication networks. Early-release implementations of gRPC-based streaming telemetry are deployed by vendors like Cisco and Juniper. Several papers analyze different use cases of this approach [5][6].

This research was directly motivated by these publications and uses some of the proposed approaches, tools, prototypes and formulas. The purpose of this paper is to compare the performance of SNMP to streaming network telemetry using gRPC with Google Protocol Buffers. In this work, only SNMP versions 1 and 2c are used for measurements and analysis. Another study which focuses on SNMPv3 and compares the performance of the security features of SNMP to streaming network telemetry using gRPC is being worked upon and will be published as separate paper related to these works.

III. MEASUREMENT SET-UP

Within this study many of the measurements were performed on virtual network devices running on VMWare Workstation Pro 14.1.1 running on top of Windows 10. Used are the following virtual images and releases:

- Cisco IOS XRv 9000 Router release 6.4.x
- Cisco Nexus 9000/3000 Virtual Switch release 7.0.x

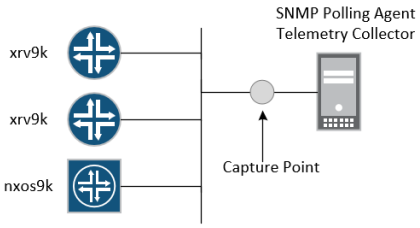


Figure 1. Measurement Set-up

In cases where additional *ifTable* rows were necessary, tunnels to other systems were created or dot1q sub-interfaces were added. ICMP traffic was generated between the systems to increase the counters of all interfaces. To measure bandwidth and delay, open source packet analyzer Wireshark was connected at the traffic capture point as shown Figure 1. The SNMP polling was done using the latest version of snmp daemon running on Ubuntu 16.04. Pipeline Telemetry Collection Service was used as network telemetry collector running on the same Ubuntu 16.04 host.

[1]Ivan Ivanov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: ivanov.ivan.iliev@gmail.com.

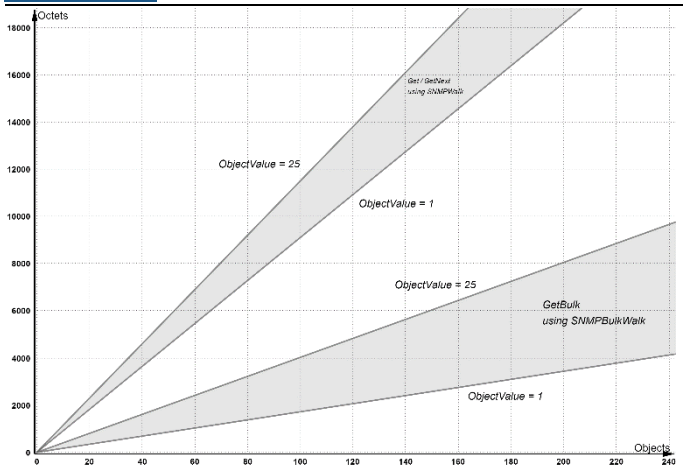


Figure 2. Theoretical SNMP bandwidth consumption

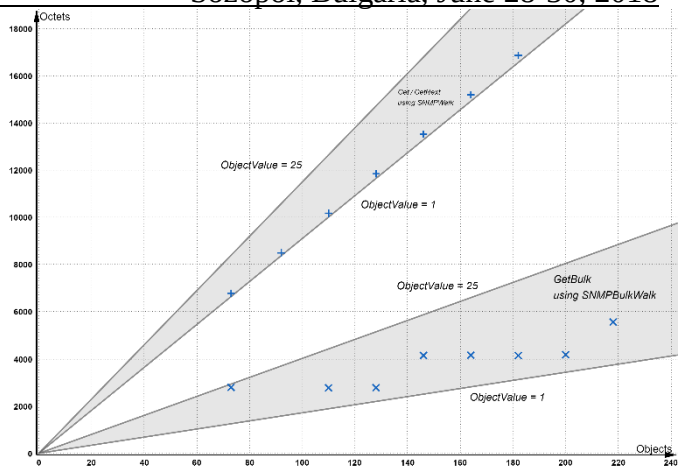


Figure 3. Measured SNMP bandwidth consumption

IV. BANDWIDTH USAGE

This section discusses and compares the bandwidth usage when retrieving data from network elements using SNMP polling and when receiving the same data via network telemetry service.

A. *SNMP messages and encoding*

Since SNMP is fully standardized protocol, the structure of a v1/v2c message is well defined. It consists of two parts: a header and a PDU. The header contains two fields: version (integer) and community (octet string). The PDU consists of five fields: PDU type (integer), Request ID (integer), Error Status / Non-repeaters (integer), Error Index/Max-repetitions (integer) and varlist (sequence). To fully understand the details of an SNMP frame, it is best to be considered as a set of nested fields. The main piece of information is the Object Identifier (OID), which identifies exactly the value to Get (read) or Set (write).

The message and it elements are defined as ASN.1 constructs. In SNMP there are two different sets of data: primitive data and complex data. The length of these data types is variable, so Basic Encoding Rules (BER) is used to solve this problem and transmit the message on the wire.

The most common ASN.1 types are INTEGER, OCTET STRING, OBJECT IDENTIFIER and SEQUENCE. In most cases, both their ASN.1 type part and ASN.1 length part takes a single octet. Therefore, the length of any of these common types depends directly on the length of the ASN.1 value part. The number of octets needed for the value part varies:

- INTEGER requires between one and five octets
- OCTET STRING requires the same number of octets as the length of the string.
- OBJECT IDENTIFIER requires the same number of octets as its length minus one.
- SEQUENCE is a construct for other types and does not require any octets for its value part.

Using the formulas derived in [2] "Comparing the Performance of SNMP and Web Services-Based Management" IEEE 2004, we can calculate the SNMP bandwidth usage in the next section.

B. *Theoritical SNMP bandwidth consumtion*

For each retrieval operation, two SNMP messages are required: a request and a response. The number of octets for the complete operation can be expressed as:

$$L_{DataRetrieval} = L_{request} + L_{response} \quad (1)$$

In an SNMP request, the BER encoding of the object value requires only two octets, because the L_{Value} is NULL. Therefore, the length of a request and response messages can be expressed as:

$$L_{Request} \approx 29 + n \cdot (5 + OID_{length}) \quad (2)$$

$$L_{Response} \approx 29 + n \cdot (5 + OID_{length} + L_{ObjectValue}) \quad (3)$$

For all measurements, all retrieved objects were from the *ifTable*, therefore the OID_{length} will be equal to 11 and we can rewrite (2) and (3) as:

$$L_{Get} \approx 58 + n \cdot (32 + L_{ObjectValue}) \quad (4)$$

$$L_{Bulk} \approx 74 + 16 \cdot n + n \cdot L_{ObjectValue} \quad (5)$$

Using (4) and (5), it is now possible to graphically represent the SNMP's bandwidth requirements as a function of the number of retrieved objects.

To verify the theoretical bandwidth projection on Figure 2, hundreds of MIB objects were retrieved from network elements. Retrieval included all rows of the *ifTable*. The results are shown on Figure 3. After all measurements were completed, all results fall into the expected areas.

C. *Streaming Network Telemetry encoded with Google Protocol Buffers (GPB).*

Telemetry describes how information from network elements can be collected using automated communication processes and transmitted to one or more telemetry collectors.

Network Telemetry is a new approach for network management and monitoring in which data is streamed from network elements continuously using a push model and provides near real-time access to operational statistics (e.g. *ifTable* for this paper).

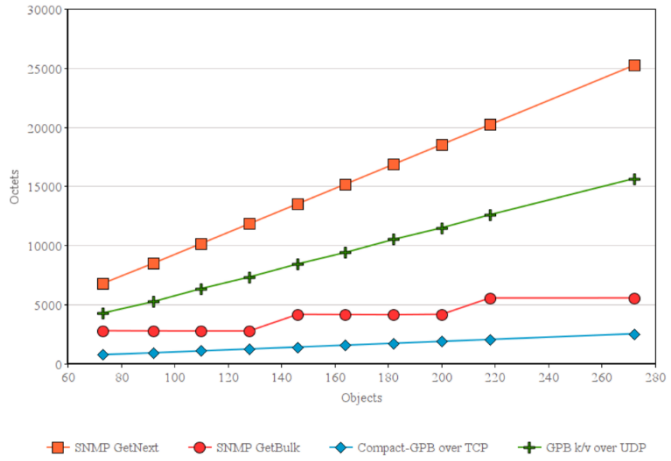


Figure 4. Measured bandwidth usage of SNMP vs Streaming Telemetry

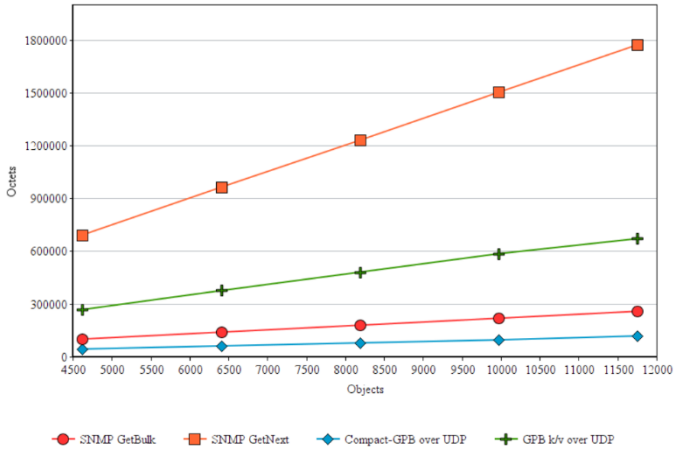


Figure 5. Large scale bandwidth usage of SNMP vs Streaming Telemetry

Since there is no standard yet for streaming network telemetry encoded with Google Protocol Buffers, it is not possible to accurately calculate the upper and lower bounds for the bandwidth needed to fetch the *ifTable* data. It had to be measured. In fact, the bandwidth required depends on specific Google Protocol Buffer definitions, which varies from case to case. This paper therefore only discusses the bandwidth requirements of our prototypes. The discussion focuses on the prototypes that receive the entire *ifTable* within a single interaction.

Interface statistics sent with telemetry using Google Protocol Buffers represent a superset of SNMP interface statistics since the network devices store 36 internal statistics for every interface and the *ifTable* has only 18 statistics per interface.

Every network device has a big number of internal databases which store raw data used for operational tasks that the device is performing. Before this raw information gets available for exporting out of the device, it has to be indexed and mapped to a data model. In the case of SNMP, the information is organized hierarchically using Management Information Bases (MIB) and Object Identifiers (OID). SNMP imposes a very tight model when it comes to indexing and exporting. In the case of the *ifTable*, each column of the table represents a different parameter for a given interface, indexed by the *ifIndex* as show in Table I.

TABLE I
IFTABLE EXAMPLE

ifIndex	ifDescription	ifType	ifSpeed	ifMTU	...
11	Loopback0	6	10000000	1500	...
12	GigabitEthernet0/0/0/1	24	1000000000	1500	...
13	FastEthernet0/0	6	100000000	1514	...

In the case of network telemetry, the internal raw data is mapped to an open-source data modeling language YANG [7]. The language, being protocol independent, can then be converted into any encoding format, e.g. XML, JSON or GPB, that the network configuration protocol supports. In our case, two types of message encoding with Google Protocol Buffers are used: Compact-GPB and GPB key-value (GPB k/v). In compact GPB, the "key" that the network device includes in the packet is just an integer.

For the interface statistics, the telemetry collector will get data that looks like this:

1: GigabitEthernet2/0/1
2: 10000
3: 1500
4: 4243242
5: 43243

Obviously that number 1 stands for the interface name, but what about 2, 3, 4, 5 etc.? To decode these keys, the telemetry collector needs a Google Protocol Buffers definition file called ". proto". With compact GPB, a ".proto" file must be generated on the network element for every path that is to be streamed and uploaded to the network collector.

With this ".proto" file, the network telemetry collector can determine that key (or "field number") "4" means packets_received, "5" means bytes_received, and so on.

This encoding is compact. It is far more efficient to send integers like "34" across the wire than strings like "MulticastPacketsReceived." And GPB is really good at sending integers on the wire: it uses the concept of "varints"[8] to serialize integers even more efficiently (i.e. a 64 bit integer doesn't actually need to take up 64 bits to be sent on the wire). From our measurements shown on Figure 6, it is obvious that the compact-GPB encoding uses the least amount of bandwidth to transmit the same amount of data (or more) than the other encoding methods and SNMP GetNext / GetBulk methods.

In the GPB key-value format, the key is sent as a string. Strings are much less efficient on the wire than varints but they are self-describing. This means that the network collector doesn't need a Google Protocol Buffers definition - ".proto" file for every path. It uses a single ".proto" file for all paths, then read the keys to figure out what the values refer to. This encoding method is easier to set-up on both sides – network telemetry collector and the network element itself but note how larger the data usage gets. For example, sending one instance of the statistics of 653 interfaces takes 610Kbytes of data. Sending the same interfaces' statistics encoded with Compact-GPB takes only 106Kbytes of data. At the same time, retrieving the same information with SNMP polling takes 258Kbytes. Therefore, we can conclude that getting the

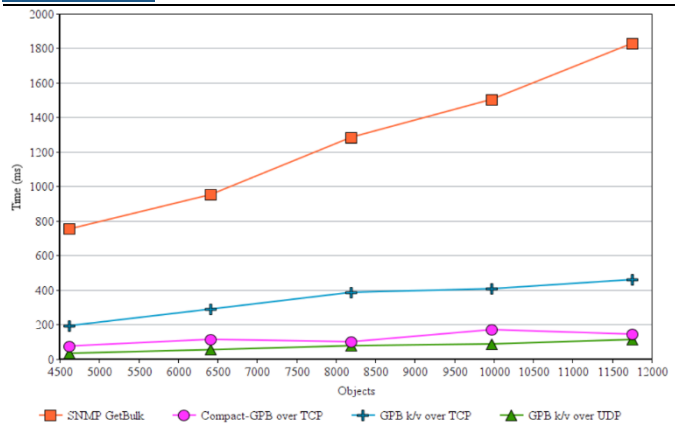


Figure 6. Round-trip delay of SNMP vs delay of Network telemetry

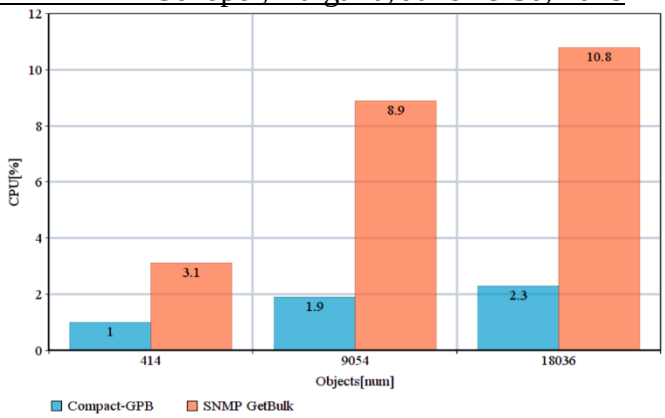


Figure 7. CPU usage of SNMP vs Network telemetry using Compact-GPB

IfTable statistics encoded with Compact-GPB is approximately 2.5 times more bandwidth efficient than SNMP polling with GetBulk. On the other hand, SNMP GetBulk is approximately 2.5 times more efficient than streaming network telemetry using GPB key-value pairs.

V. Delay and CPU usage

The strict semantics of the SNMP GetNext/GetBulk operations force the network device to traverse the *ifTable* column by column from lowest index value to highest. From a network device's perspective, that is not optimal. Network devices store their internal information in a way that is most efficient for their operational needs. In IOS XR, the internal data structure for interface statistics is indexed by interface name and is stored in a structure called a bag. The router's most efficient internal bulk data retrieval is to grab a whole bag of data at once. But the router cannot just send the bag in SNMP. Instead, it has to re-order the data into a table and walk the columns to fulfill the GetBulk request indexed by the *ifIndex*.

Telemetry collects data using the internal bulk data collection mechanisms, does some minimal processing to filter and translate the internal structure to a Google Protocol Buffer, and then pushes the whole thing to the network collector at the configured intervals.

By eliminating the process of re-ordering of the information like in the case of SNMP, streaming network telemetry is more process efficient and requires less CPU cycles. As seen on Figure 6, measured round-trip delays, including packetization, serialization and processing of network telemetry data is approximately 4 times lower than that of SNMP. Figure 7 shows the measured CPU usage at the time of retrieval of the interface statistics using both SNMP and Compact-GPB.

VI. Conclusion

This paper compared the performance of SNMP to network telemetry streaming encoded with Google Protocol Buffers. In particular, it investigated the bandwidth usage, delays and CPU usage.

Our measurements show that retrieving the *ifTable* statistics encoded with Compact-GPB is approximately 2.5 times more bandwidth efficient than SNMP polling with GetBulk. Measured round-trip delays, including packetization, serialization and processing of the telemetry stream is approximately 4 times lower than that of SNMP. Additionally, the measurements show that network telemetry is less CPU intensive than SNMP polling.

Network telemetry is still complex to set-up and is not standardized. It lacks compatibility between vendors and even between different platforms of the same vendor.

References

[1] C. Pattinson, "A Study of the Behaviour of the Simple Network Management Protocol," in Proceedings of DSOM2001, October 2001.

[2] A. Pras, T. Drevers, R. van de Meent and D. Quartel, "Comparing the performance of SNMP and Web services-based management," in IEEE Transactions on Network and Service Management, vol. 1, no. 2, pp. 72-82, Dec. 2004.

[3] W. Queiroz de Lima, R. S. Alves, R. L. Vianna, M. J. B. Almeida, L. M. R. Tarouco and L. Z. Granville, "Evaluating the Performance of SNMP and Web Services Notifications," 2006 IEEE/IFIP Network Operations and Management Symposium NOMS 2006, Vancouver, BC, 2006, pp. 546-556.

[4] J. Schonwalder, A. Pras, M. Harvan, J. Schippers and R. van de Meent, "SNMP Traffic Analysis: Approaches, Tools, and First Results," 2007 10th IFIP/IEEE International Symposium on Integrated Network Management, Munich, 2007, pp. 323-332.

[5] F. Paolucci, A. Sgambelluri, M. Dallaglio, F. Cugini and P. Castoldi, "Demonstration of gRPC Telemetry for Soft Failure Detection in Elastic Optical Networks," 2017 European Conference on Optical Communication (ECOC), Gothenburg, 2017, pp. 1-3.

[6] T. Choi, S. Yoon and S. Song, "Information fusion based agile streaming telemetry for intelligent traffic analytics of softwarized network," 2017 19th Asia-Pacific Network Operations and Management Symposium (APNOMS), Seoul, 2017, pp. 399-402

[7] YANG - A Data Modeling Language for the Network Configuration Protocol (NETCONF); Internet Engineering Task Force (IETF) M. Bjorklund, Ed.; Tail-f Systems; October 2010

[8] Google Protocol Buffers Encoding Integers - binary wire format https://developers.google.com/protocol-buffers/docs/enco

Optimal Pair-Wise SUS Scheduling Algorithm for Multiuser MIMO

Aleksandra Panajotović[1], Nikola Sekulović[2] and Daniela Milović[3]

Abstract – **In this paper we present an optimal pair-wise semi-orthogonal user selection (SUS) scheduling algorithm for multiuser multiple-input multiple-output orthogonal frequency division multiplexing (MU-MIMO-OFDM) system. Based on the perfect channel state information (CSI), zero-forcing beam forming (ZFBF) precoding is applied to cancel inter- and intra-user interferences. Presented simulation results demonstrate benefit of proposed user selection algorithm in comparison to generalized multicarrier semi-orthogonal user selection algorithm (GMSUS), both applicable to IEEE 802.11ac.**

Keywords – **IEEE 802.11ac, MU-MIMO-OFDM, User scheduling algorithm, ZFBF.**

I. INTRODUCTION

Wi-Fi standard IEEE 802.11n has launched the technique known as multiple-input multiple-output (MIMO) enabling the system to arrange multiple data streams on the same channel which results in higher channel capacity. In addition, orthogonal frequency division multiplexing (OFDM) is one more feature incorporated into IEEE 802.11n. It is effective technique to mitigate intersymbol interference in frequency-selective channel by transforming a broadband frequency-selective channel into series of non-interfering narrowband sub-channels [1]. Capacity increment realized by proposed MIMO technique is limited by minimum number of transmit (N_T) and receive antennas (N_R) [2]. Therefore, a new IEEE 802.11ac standard, known as Gbps Wi-Fi, suggests using multiuser MIMO (MU-MIMO) technique. In MU-MIMO, an access point (AP) transmits simultaneously to several compatible mobile stations (MSs), i.e. users over the same spectrum. Therefore, this technique does not increase the performance that MSs will see, but allows the network to increase its utilization by transmitting to multiple clients simultaneously.

In order to maximize capacity in MU-MIMO, the optimal strategy consists of dirty paper coding (DPC) in combination with user scheduling and power allocation [3]. Unfortunately, DPC technique is impractical because of complicated encoding and decoding schemes even for a moderate number of users. Therefore, some alternative and more practical precoding techniques have to be applied, such as zero-forcing beamforming (ZFBF) [4] and block diagonalization (BD) [5].

[1]Aleksandra Panajotović is with the Faculty of Electronic Engineeering at University of Niš, Aleksandra Medvedeva 14, Niš 18000, Serbia, E-mail: aleksandra.panajotovic@elfak.ni.ac.rs.

[2]Nikola Sekulović is with the College of Applied Technical Sciences, Aleksandra Medvedeva 20, Niš 18000, Serbia.

[3]Daniela Milović is with the Faculty of Electronic Engineeering at University of Niš, Aleksandra Medvedeva 14, Niš 18000, Serbia.

It is known that ZFBF cancels inter- and intra-user interferences, while BD just cancels inter-user interference leaving intra-user cancellation to subsequent beamforming or equalization processes. Analysis which of these techniques is the most appropriate from sum-rate point of view is done in [6].

The number of users that AP can serve simultaneously is equal to or less than the number of antennas in AP. So, a fundamental problem arising in MU-MIMO system is how AP should choose a subset of users to be served. The optimal schedule is found by exhaustive search, i.e. achieved sum rate is evaluated for all combination of users and the user combination providing the maximal sum rate is scheduled. However, in the case when number of the users, N_u, is large, exhaustive search cannot be used any longer, since the size of search space, $\sum_{i=1}^{N_T}\binom{N_u}{i}$, becomes prohibitively large. Therefore, authors in [4] propose semi-orthogonal user selection (SUS) algorithm which in combination with ZFBF gives the performance reasonably close to that of DPC under practical values of N_u. That algorithm is adapted to multi-stream multicarrier users transmitting using OFDM as generalized multicarrier semi-orthogonal user selection (GMSUS) algorithm applicable to IEEE 802.11ac [7-9]. Modified version of SUS algorithm suitable for massive MIMO is described in [10] as massive MIMO pair-wise SUS algorithm.

In this paper we propose a new algorithm, acceptable to IEEE 802.11ac, based on both pair-wise and optimal scheduling algorithms. Simulation results show that proposed algorithm outperforms performance of its counterpart also based on SUS. The price paid for performance improvement is a bit greater complexity from practical point of view.

This introduction ends with notational remarks. Vector and matrices are denoted by lower- and upper-case bold letters, respectively, while scalars are represented with non-bold letters. $(\cdot)^T$ and $(\cdot)^H$ denote transpose and complex transpose, correspondingly, $D(x)$ is a (block) diagonal matrix with x at its main diagonal, $|Q|$ is the cardinality of subset Q, $\boldsymbol{I}_L$ is $L \times L$ identity matrix, $\|\boldsymbol{a}\|$ represents the Euclidian norm of a vector $\boldsymbol{a}$, and R and C are the set of real and complex numbers, respectively.

II. SYSTEM MODEL

We consider the downlink of MU-MIMO-OFDM system with N_T transmit antennas at AP and N_u users each equipped with N_R ($N_T \geq N_R$) receive antennas. The system operates over N_c OFDM subcarriers, out of which N_d are used to transmit data while rest of them correspond to pilots and guard band.

Let $\mathbf{H}_u[q] \in C^{N_R \times N_T}$ represents MIMO propagation channel between AP and *u*-th user over *q*-th subcarrier. Singular value decomposition (SVD) applied on $\mathbf{H}_u[q]$ results into

$$\mathbf{H}_u[q] = \mathbf{U}_u[q]\boldsymbol{\Sigma}_u[q]\mathbf{V}_u^H[q], \quad (1)$$

where $\mathbf{U}_u[q] = \left[\mathbf{u}_{u,1}[q], \text{K}, \mathbf{u}_{u,N_R}[q]\right] \in C^{N_R \times N_R}$ and $\mathbf{V}_u[q] = \left[\mathbf{v}_{u,1}[q], \text{K}, \mathbf{v}_{u,N_T}[q]\right] \in C^{N_T \times N_T}$ are unitary matrices containing the left and the right singular vectors of $\mathbf{H}_u[q]$ and $\boldsymbol{\Sigma}_u[q] = D\left(\sigma_{u,1}[q], \text{K}, \sigma_{u,N_R}[q]\right) \in R^{N_R \times N_T}$ is a diagonal matrix which elements on main diagonal are singular values of $\mathbf{H}_u[q]$. In order to eliminate interference ZFBF post-processing is applied at receiver, so now the equivalent channel matrix for *u*-th user on *q*-th subcarrier is defined as

$$\tilde{\mathbf{H}}_u[q] = \mathbf{U}_u^H[q]\mathbf{H}_u[q] = \boldsymbol{\Sigma}_u[q]\mathbf{V}_u^H[q]. \quad (2)$$

Since $N_R \le N_T$ and $\mathbf{H}_u[q]$ is full rank matrix, $\boldsymbol{\Sigma}_u[q]$ contains N_R non zero values, so there is a potential to transport N_R spatial streams. By treating each spatial stream as a virtual user with single antenna, now the system can be considered as a multiuser multiple-input single-output (MU-MISO) with $K = N_u N_R$ virtual users. Now, equivalent channel gain corresponding to *k*-th virtual user on *q*-th subcarrier is $\tilde{\mathbf{h}}_k[q] = \sigma_k[q]\mathbf{v}_k^H[q]$, where $k = n + (u-1)N_R$ and $n \in \overline{1, N_R}$. Assuming perfect frequency synchronization between transmitter and receiver and a cyclic prefix duration exceeding the channel delay spread, the received signal at the *k*-th virtual user on subcarrier *q* for an arbitrary OFDM symbol may be written as

$$\tilde{y}_k[q] = \tilde{\mathbf{h}}_k[q]\mathbf{x}_k[q] + \eta_k[q], \; k = \{1, \text{K}, N_u N_R\}, \quad (3)$$

where $\mathbf{x}_k[q] \in C^{N_T \times 1}$ is the vector of transmitted symbol from AP antenna on subcarrier *q*. $\tilde{\eta}_k[q]$ is *n*-th element of vector $\tilde{\boldsymbol{\eta}}_u[q] = \mathbf{U}_u^H[q]\boldsymbol{\eta}_u[q]$, where $\boldsymbol{\eta}_u[q]$ is a zero-mean circularly symmetric complex Gaussian vector with covariance matrix $\mathbf{R}_\eta = \sigma_\eta^2 \mathbf{I}_{N_R}$. Since $\mathbf{U}_u[q]$ is unitary matrix, it holds $\tilde{\boldsymbol{\eta}}_u[q]$: $CN\left(0, \sigma_\eta^2 \mathbf{I}_{N_R}\right)$.

In downlink scenario with large number of users, the AP serves users with favourable channel conditions. Let $Q = \left\{u_1, \text{K}, u_{|Q|}\right\}$ be the subset of those virtual users which be served by AP in a given time slot. Now, let define matrix collecting the channel coefficient of selected virtual users on subcarrier *q* as $\tilde{\mathbf{H}}_Q[q] = \left[\mathbf{h}_{u_1}^T[q] \text{K} \, \mathbf{h}_{u_{|Q|}}^T[q]\right]^T \in C^{|Q| \times N_T}$. To totally suppress multiuser interference, linear precoding has to be carried out as

$$\mathbf{x}[q] = \mathbf{W}_Q[q]\mathbf{P}_Q^{1/2}[q]\mathbf{s}_Q[q], \quad (4)$$

where $\mathbf{W}_Q[q] = \left[\mathbf{w}_{u_1}[q] \text{K} \, \mathbf{w}_{u_{|Q|}}[q]\right] \in C^{N_T \times |Q|}$ is ZFBF precoding matrix on subcarrier *q*, $\mathbf{P}_Q[q] = D\left(P_{u_1}[q], \text{K}, P_{u_{|Q|}}[q]\right) \in R^{|Q| \times |Q|}$ is power allocation matrix which satisfies $\sum_{u_i \in Q} \left\|\mathbf{w}_{u_i}[q]\right\|^2 P_{u_i}[q] = \frac{P_T}{N_d}$, with P_T being total power available for all subcarriers and $\mathbf{s}_Q[q] = \left[s_{u_1}[q] \text{K} \, s_{u_{|Q|}}[q]\right]^T \in C^{|Q| \times 1}$ being the vector which contains the information symbols sent to users from selected group *Q*. After pre- and post-processing, the received signal for selected virtual user $u_i \in Q$ on subcarrier *q* can be presented as

$$\tilde{y}_{u_i}[q] = \sqrt{P_{u_i}[q]}\, s_{u_i}[q] + \tilde{\eta}_{u_i}[q], \quad (5)$$

where optimal value of $P_{u_i}[q]$, obtained to provide maximum capacity, can be found using waterfilling method [4].

III. Optimal pair-wise SUS algorithm

The optimal scheduled user group can be found through exhaustive search. Such approach is not acceptable when number of users is large, especially when users are equipped with more than one antenna. Therefore, design of applicable scheduling algorithm is important issue in MU-MIMO systems. In [10], authors point out that the complexity of pair-wise SUS algorithm is smaller than its traditional counterpart [4]. For that reason, in this paper we propose optimal pair-wise SUS algorithm based on both pair-wise SUS algorithm and exhaustive search. The steps of the optimal pair-wise SUS algorithm are:

Step 1: Initialization

$$Q_0 = \{1, \text{K}, K\}, \quad (6)$$

$$i = 1. \quad (7)$$

Step 2: Determine the degrees of orthogonality, $\beta_{l,p}$, between all virtual user pairs $l \ne p$:

$$\beta_{l,p} = \frac{\left|\tilde{\mathbf{h}}_l^H \tilde{\mathbf{h}}_p\right|}{\left\|\tilde{\mathbf{h}}_l\right\| \left\|\tilde{\mathbf{h}}_p\right\|}. \quad (8)$$

Step 3: Find the pair P_i from Q_{i-1} with the smallest degree of orthogonality

$$P_i = \{l, p\} = \arg \min_{l, p \in Q_{i-1}} \max \beta_{l,p}. \quad (9)$$

Step 4: Select *i*-th virtual user to be eliminated as follows

$$\pi(i) = \arg \min_{r \in P_i} \left\|\tilde{\mathbf{h}}_r\right\| \quad (10)$$

$$Q_i = \left\{ m \in Q_{i-1} \middle| m \neq \pi(i) \right\}, \quad (11)$$

$$i \leftarrow i+1. \quad (12)$$

If $|Q_{i-1}| > N_T$, go to Step 3.

Step 5: Apply exhaustive search, i.e. calculate the realized throughput for all combination of virtual users selected in Step 4 and find the combination providing maximal throughput.

$$Q_{opt} = \arg \max_{S \subset Q_{i-1}} T(S). \quad (13)$$

In IEEE 802.11ac system, only finite set of transmission modes, which consist of a combination of modulation alphabet and coding rate (MCS), is available [9]. Consequently, once the group of virtual users has been selected, fast link adaptation (FLA) carries out a procedure to allocate MCS to users in order to maximize system throughput while satisfying predetermined quality of service (QoS) constraints. Usually, QoS constraint is in the form of an outage probability of target packet error rate (PER_0). Since system packet error rate (PER) depends on the allocated MCS, the received SNR, packet length and channel realization, derivation and evaluation of analytical expression for PER is a awkward task. This problem can be solved using a look-up table which maps all those parameters onto a single link quality metrics (LQM) which is then associated to PER value. In this paper, LQM known as effective SNR (SNR_{eff}) is used [11]. Namely, the optimal MCS for particular conditions in the channel is determined using SNR_{eff}.

IV. Numerical results

In order to show advantage of the proposed optimal pair-wise SUS scheduling algorithm in throughput point of view, this section presents simulation results obtained using parameters from IEEE 802.11ac standard. System operates on 5.25GHz carrier frequency with bandwidth of 20MHz that is divided into N_c = 64 subcarriers out of which N_d = 52 are used to carry data while the rest correspond to pilot signals and guard intervals. The AP has N_T = 4 transmit antennas, while all users are equipped with N_R = 2 receive antennas. Channel profile B from [12] is used in the simulations testbed to generate a space-time-frequency-selective fading channel. The values of parameters for FLA are taken from Table I in [9].

Figures 1 and 2 present the system throughput as a function of average received SNR for MU-MIMO-OFDM in B channel which characterizes environment with little-to-moderate frequency selectivity. System throughput results achieved serving homogenous users selected by optimal pair-wise SUS algorithm are compared with ones achieved using GMSUS algorithm [9]. The GMSUS algorithm is built on SUS algorithm described in [4]. Namely, it is enhanced version of SUS algorithm suitable for IEEE 802.11ac. Parameter θ defines degree of orthogonality between selected virtual users. It has been shown that θ in range [0.4-0.6] maximizes sum-rate capacity [9]. In adaptive GMSUS algorithm, appropriate new value of θ that maximizes system throughput is selected

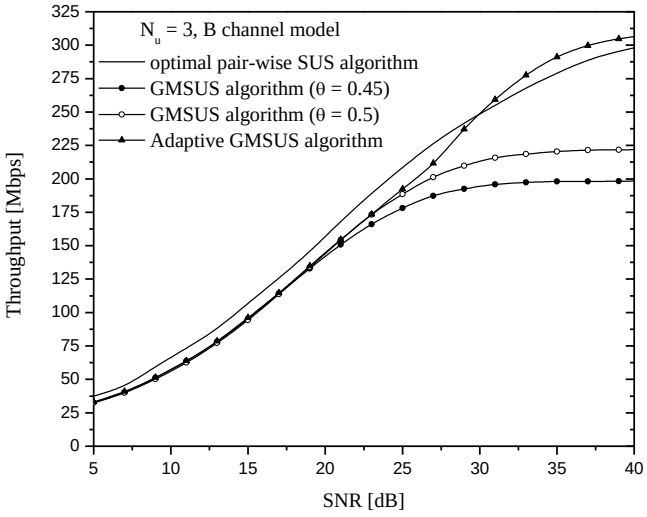


Fig. 1. Throughput comparison for different user scheduling algorithms for N_u = 3 users

for each considered value of SNR. It is evident that optimal

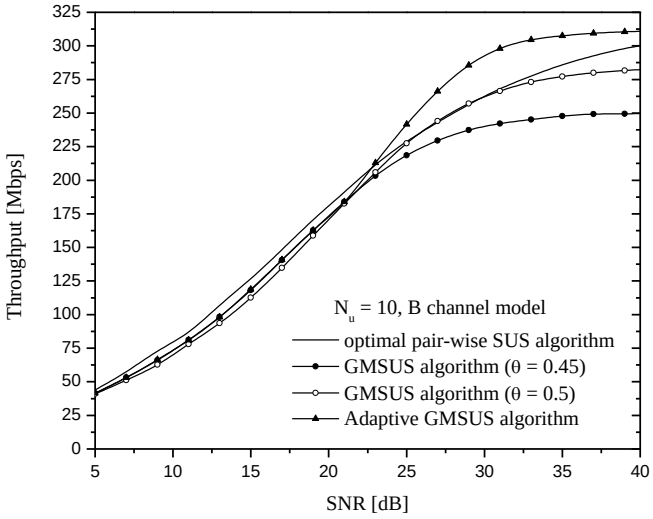


Fig. 2. Throughput comparison for different user scheduling algorithms for N_u = 10 users

pair-wise SUS algorithm provides the highest throughput for small and moderate SNR, while for large SNR the best throughput performance is achieved using adaptive GMSUS. It is known, that maximal data rate per spatial stream in IEEE 802.11ac system is 78Mbps. Therefore, maximal throughput that can be realized in considered IEEE 802.11ac system is 4*78Mb/s = 312 Mb/s. It has to be highlighted that maximal throughput can be achieved with optimal pair-wise SUS algorithm for large SNR even in a case when system should serve small number of users. In addition, it can clearly be noticed that existence of more users in the system, for all

considered algorithms, leads to higher throughput thanks to the increased multiuser diversity that a larger N_u brings along.

V. Conclusion

This paper has proposed the optimal pair-wise SUS scheduling algorithm. The simulated throughput results have shown that this user scheduling algorithm can be accepted for application in MU-MIMO-OFDM systems complaint with IEEE 802.11ac standard. Namely, throughput performance of that system outperforms those ones realized using GMSUS and adaptive GMSUS algorithm for small and moderate SNR and achieves possible maximum for large SNR. The advantage of proposed algorithm is more significant for the case when AP serves small number of users. Further work will concentrate on modification of exhaustive search part of optimal pair-wise SUS algorithm in order to achieve performance improvement with slightly less complicated approach.

Acknowledgement

This work has been funded by Ministry for Education, Science and Technological Development of the Republic of Serbia under the projects TR-32052, III-44006, TR-32051.

Professors G. Femeinas and F. Riera-Palou deserve sincere gratitude for invaluable guidance and support for research of MU-MIMO systems.

References

[1] R. van Nee, R. Prasad, *OFDM for Wireless Multimedia Communications*, Artech House, 2000.

[2] I. E. Teletar, "Capacity of Multi-antenna Gaussian Channels", Eur. Trans. Telecommun.,vol. 10, no. 6, pp. 585-595, 1999.

[3] D. S. Gesbert, M. Kountouris, R. W. Health Jr., C. –B. Chae and T. Sälzer, "Shifting to MIMO Paradigm", IEEE Sig. Proc. Mag., vol. 24, no. 5, pp. 36-46, 2007.

[4] T. Yoo, A. Goldsmith, "On the Optimaplity of Multiantenna Broadcast Scheduling Using Zero-Forcing Beamforming", IEEE JSAC, vol. 24, no. 3, pp. 528-541, 2006.

[5] Q. Spencer, A. Swindlehurst and M. Haardt, "Zero-forcing Methods for Downlink Spatial Multiplexing in Multiuser MIMO Channels", IEEE Trans. Sig. Proc., vol. 52, no. 2, pp. 461-471, 2004.

[6] R. Chen, Z. Shen, J. G. Andrews and R. W. Heath, "Multimode Transmission for Multiuser MIMO System with Block Diagonalization", IEEE Trans. Sig. Proc., vol. 56, no. 7, pp. 3294-3302, 2008.

[7] M. Essaouli, F. Rierra-Palou and G. Femenias, "Opportunistic Multiuser MIMO for OFDM Networks", MC-SS 2011, Conference Proceedings, pp.1-5, Herrsching, Germany, 2011.

[8] M. Essaouli, F. Rierra-Palou and G. Femenias, "A fair MU-MIMO Scheme for IEEE 802.11ac", ISWCS 2012, Conference Proceedings, pp.1049-1053, Paris, France, 2012.

[9] A. Panajotovic, F. Rierra-Palou and G. Femenias, "Adaptive Uniform Channel decomposition in MU-MIMO-OFDM: Application to IEEE 802.11ac", IEEE Trans. Wirel. Commun., vol. 14, no. 5, pp. 2896-2910, 2015.

[10] S. Dierks, N. Jünger, "Scheduling for Massive MIMO with Few Excess Antennas", WSA 2016, Conference Proceedings, pp. 40-44, Munich, Germany, 2016.

[11] G. Martorell, F. Riera-Palou and G. Femenias, "Cross-Layer Fast Link Adaptation for MIMO-OFDM based WLANs", Wirel. Pers. Commun., vol. 56, no. 3, pp. 599-609, 2011.

[12] J. Kermoal, L. Schumacher, K. Pedersen, P. Mogensen and F. Fredeiksen, "A Stochastic MIMO Radio Channel Model with Experimental Validation", IEEE JSAC, vol. 20, no. 6, pp. 1211-1226, 2002.

Influence of the Doping Additives on the Properties of the Synthesized Barium Titanate

Ivaylo Lazarov[1] and Anka Zheglova[2]

Abstract – **In this paper are investigated influence of the different doped additives on the properties of the synthesized barium titanate. For composition $BaTiO_3$ + 1mol%H_3BO_3 have been obtained the highest values for the density, relative shrinkage and the relative permittivity at the temperatures of sintering 1200ºC and 1250ºC. The adding of the 0.05mol%Bi_2O_3 and 0.05mol%SnO_2 increases the values of the relative permittivity. For these compositions is observe the diffuse phase transitions in the temperature dependencies of the dielectric permittivity.**

Keywords –**Barium titanate, Doping additives, Dielectric permittivity.**

I. INTRODUCTION

Barium titanate is one of the most studied ferroelectric materials due to its excellent properties and high values of the dielectric permittivity. $BaTiO_3$ – based ceramics are widely uses in manufacture of multilayer capacitors (MLCC)[1], thermistors[2], electric-field tunable devices[3], sensors[4] and piezoelectric actuators[5]. Barium titanate is typical ferroelectric material with perovskite tetragonal microstructure (ABO_3) at room temperature and the Curie of temperature about 120ºC. When heated above the Curie of temperature, it passes from ferroelectric to parra-electric phase. It is structure changes from tetragonal to cubic.

Much efforts have been developed to improve its dielectric properties such as increase the dielectric permittivity and decrease the dielectric losses.In order toobtain the required dielectric properties of modified barium titanate ceramics are added various doped additives suchas MnO_2[6], NbO_2[7], Sn[8], W[9], Ca[10], Zr[11].The doped additives influence also on the homogeneity, the particle size, the density, the structural defects and Curie of temperature.

The aim of this study is to determine the influence of the doping additives on the properties of the synthesized barium titanate.

II. EXPEREMENTAL

Samples of the study compositions were prepared using solid-state reactionmethodtaking commercial barium titanate $BaTiO_3$. As doped additives are used boric acid H_3BO_3, bismuth trioxide Bi_2O_3 and stannic oxide SnO_2 such as the starting powders are in stoichiometric proportions. The starting powders in the required amounts have been homogenized in 3% of polyvinyl alcohol solution for 4 hours at planetary mill Pulvirisete 5. After drying the materials have been formed in the form of disks with diameter 11mm and thickness of 2-3mm., by compression at a pressure 300MPa. Synthesis was carried out in air for 2 hours at five different temperatures – 1050ºC, 1100ºC, 1150ºC, 1200ºC and 1250ºC. After a mechanical treatment, silver electrodes have been formed by applying a coating of silver paste on the ceramic disks, calcined at the temperature of 900ºC for 1h.

[1]Ivaylo Lazarov is with the Faculty of Electrical Engineering and Electronicsat Technical University of Gabrovo, 4 H. Dimitar Str, Gabrovo 5300, Bulgaria, E-mail:iv.lazarovt@mail.bg.

[2]Anka Zheglova is with the Faculty of Electrical Engineering and Electronicsat Technical University of Gabrovo, 4 H. DimitarStr, Gabrovo 5300, Bulgaria.

III. ANALYSIS OF THE RESULTS

A. Density and relative shrinkage

The results of the measured density at different sintering temperatures are presented in Fig.1 and are given in Table I. The density increases for compositions $BaTiO_3$ + 1mol%$H_3BO_3$and $BaTiO_3$+0.05mol%$SnO_2$with increasing the temperature of sintering. The density has maximum values for composition $BaTiO_3$+0.05mol%Bi_2O_3 at temperature of sintering 1200ºC. At the temperature of sintering 1250ºC the density decreases for this composition.

TABLE I
THE VALUES OF THE DENSITY AT DIFFERENT TEMPERATURES OF SINTERING FOR STUDY COMPOSITIONS

Composition	$t_{sintering}$, [ºC]	Density ρ, [g/cm³]
$BaTiO_3$+1mol%H_3BO_3	1050	3,41
	1100	3,67
	1150	4,12
	1200	5,52
	1250	5,61
$BaTiO_3$+0.05mol%Bi_2O_3	1050	3,9
	1100	5,18
	1150	5,25
	1200	5,28
	1250	4,18
$BaTiO_3$+0.05mol%SnO_2	1050	3,27
	1100	3,28
	1150	3,332
	1200	3,98
	1250	5,16

The highest values of the density have been obtained for composition $BaTiO_3$ + 0,05mol%Bi_2O_3 at low temperature of sintering. At high temperatures of sintering were obtained the highest values of the density for compositions $BaTiO_3$+1mol%H_3BO_3.

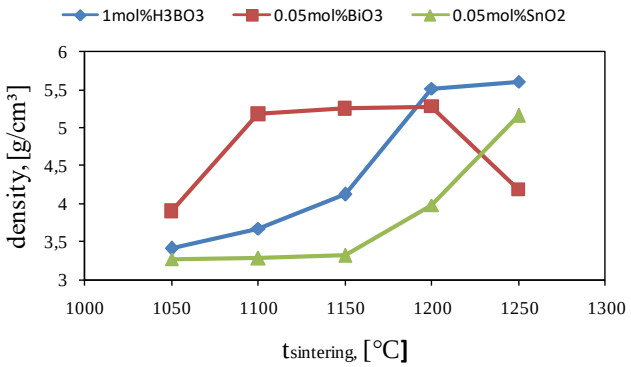


Fig. 1.Dependence of the density of the ceramic materials on the temperature of sintering

The results for obtained values of the relative shrinkage for study compositions are given in Table II.

TABLE II
THE VALUES OF THE RELATIVE SHRINKAGE AT DIFFERENT TEMPERATURES OF SINTERING FOR STUDY COMPOSITIONS

Composition	$t_{sintering}$, [°C]	relative shrinkage, [%]
$BaTiO_3$+1mol%H_3BO_3	1050	-1,82
	1100	-3,64
	1150	-7,27
	1200	-16,36
	1250	16,36
$BaTiO_3$+0.05mol%Bi_2O_3	1050	-4,55
	1100	-13,6
	1150	-13,6
	1200	-14,56
	1250	-7,27
$BaTiO_3$+0.05mol%SnO_2	1050	0
	1100	0
	1150	0
	1200	-5,45
	1250	-13,64

For composition $BaTiO_3$+0.05mol%SnO_2 do not observe relative shrinkage at low temperature of sintering. The relative shrinkage increases with increasing the temperature of sintering for composition $BaTiO_3$+1mol%H_3BO_3. For composition $BaTiO_3$+0.05mol%Bi_2O_3 have been obtained the maximum values of the relative shrinkage at the temperature of sintering 1200°C.

The highest values of the relative shrinkage have been obtained for composition $BaTiO_3$+1mol%H_3BO_3 at the high temperatures of sintering.

Fig.2 shows dependence of the relative shrinkage versus temperature of sintering.

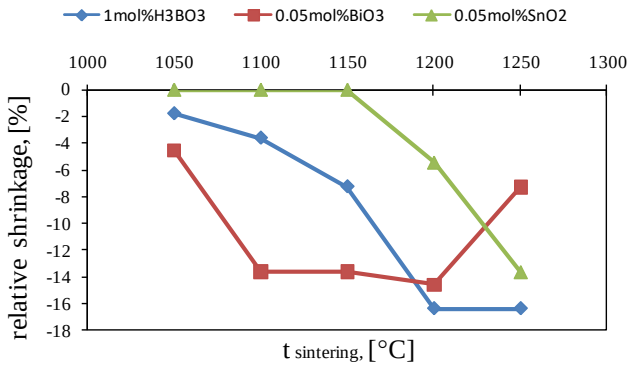


Fig. 2.Dependence of the relative shrinkage of the ceramic materials on the temperature of sintering

B. *Dielectric permittivity*

In the Fig.3 are showed the variation of the dielectric permittivity as a function of the temperature of sintering.

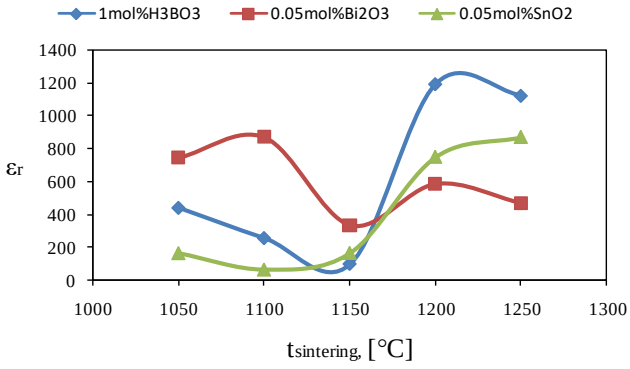


Fig.3. Dependence of the relative permittivity on the temperature of sintering

The highest values of the relative permittivity have been obtained for composition from doped with 1mol%H_3BO_3barium titanate at the temperatures of sintering 1200°C and 1250°C.The adding of 0.05mol%Bi_2O_3 to barium titanate reduces the temperature of sintering. The highest values of the dielectric permittivity have been obtained for this composition at low temperature of sintering.

The values of the relative permittivity at the Curie temperature and the Curie temperature for composition $BaTiO_3$+1mol%H_3BO_3 are given in Table III.

Fig.4.shows the temperature dependencies on the relative permittivity for doped with 1mol%H_3BO_3barium titanate.

It is found that dielectric permittivity have been the highest values for samples sintered at temperatures 1200°C and 1250°C. It is observed dielectric peaks in the temperature dependencies of the relative permittivity about 110°C. The diffuse phase transitions have been obtained for samples sintered at temperature 1050, 1100 and 1150°C.

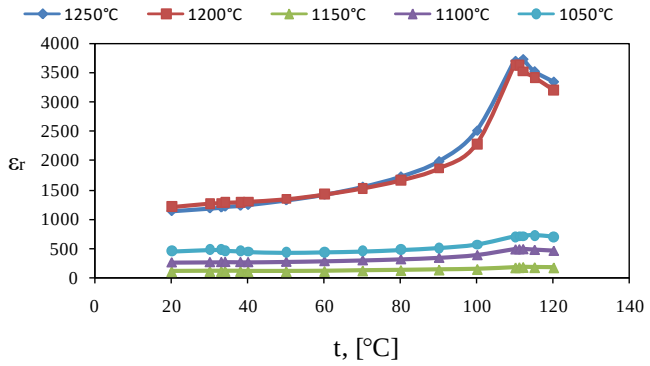


Fig.4. Temperature dependencies on the relative permittivity for composition $BaTiO_3$+1mol%H_3BO_3 at different temperatures of sintering

Table III

THE VALUES OF THE DIELECTRIC PERMITTIVITY AT THE CURIE TEMPERATURE AND THE CURIE TEMPERATURE FOR DOPED WITH 1MOL%H_3BO_3 BARIUM TITANATE

$t_{sintering}$, [°C]	max ε_r	Curie temperature, [°C]
1050	467	113
1100	487	112
1150	101	112
1200	3627	111
1250	3729	112

The temperature dependencies for composition $BaTiO_3$+0.05mol%Bi_2O_3 are shown in Fig.5.

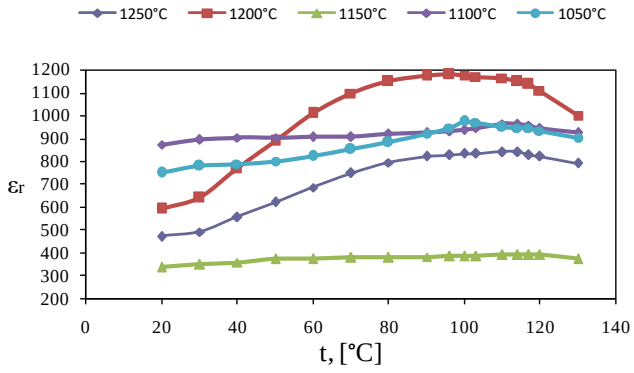


Fig.5. Temperature dependencies on the relative permittivity for composition $BaTiO_3$+0.05mol%Bi_2O_3 at different temperatures of sintering

The adding of 0.05mol%Bi_2O_3 leads to reduce the temperature of sintering. It is observed the diffuse phase transitions for the temperature dependencies at all the temperature of sintering. The values of the relative permittivity at high temperature of sintering are lower than for composition from doped with 1mol%H_3BO_3 barium titanate.

The values of the relative permittivity at the Curie temperature (max ε_r) and the Curie temperature for composition $BaTiO_3$+0.05mol%Bi_2O_3 are given in Table IV.

Table IV

THE VALUES OF THE DIELECTRIC PERMITTIVITY AT THE CURIE TEMPERATURE AND THE CURIE TEMPERATURE FOR DOPED WITH 0.05MOL%BI_2O_3 BARIUM TITANATE

$t_{sintering}$, [°C]	max ε_r	Curie temperature, [°C]
1050	972	103
1100	958	120
1150	420	117
1200	1179	96
1250	841	114

The temperature dependencies for composition $BaTiO_3$+0.05mol%SnO_2 are shown in the Fig.6.

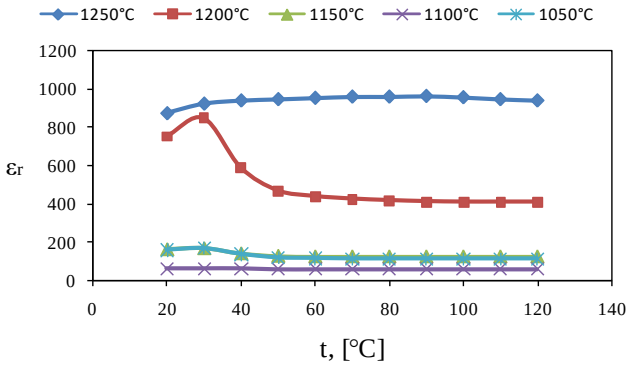


Fig.6. Temperature dependencies on the relative permittivity for composition $BaTiO_3$+0.05mol%SnO_2 at different temperatures of sintering

The highest values of the dielectric permittivity have been obtained for samples, sintered at temperature 1250°C. It is observed diffuse phase transitions. For samples sintered at temperature 1200°C have been obtained lower the values of the dielectric permittivity, but on the temperature dependence have been found dielectric peak at the temperature near to the room temperature. The values for the dielectric permittivity are the lowest than all studies compositions.

The values of the relative permittivity at the Curie temperature (max ε_r) and the Curie temperature for composition $BaTiO_3$+0.05mol%SnO_2 are given in Table V.

Table V

THE VALUES OF THE DIELECTRIC PERMITTIVITY AT THE CURIE TEMPERATURE AND THE CURIE TEMPERATURE FOR DOPED WITH 0.05MOL%SNO_2 BARIUM TITANATE

$t_{sintering}$, [°C]	max ε_r	Curie temperature, [°C]
1050	171	32
1100	62	40
1150	169	31
1200	865	31
1250	962	89

IV. CONCLUSION

Doped BaTiO3 ceramics were prepared by a solid state reaction method are investigated. The adding of 1mol%H_3BO_3 leads to obtain the higher values of the density, the relative shrinkage and the dielectric permittivity at temperatures of sintering 1200°C and 1250°C.The diffuse phase transitions have been obtained for doped with 0.05mol%Bi_2O_3 and 0.05mol%SnO_2 barium titanate. The adding of 0.05mol%Bi_2O_3 reduces the temperature of sintering.

REFERENCES

[1] C.Ma, X.H.Wang, R.Z.Chen, L.T.Li, Z.I.Gui, "The structure and dielectric properties of low temperature sintering barium titanate based x7r ceramics", Journal of Electroceramics, vol.21, pp.242-245, 2008.

[2] M.Wegman, R.Brönnimann, F.Clemens, T.Graurle, "Barium titanate-based PTCR Thermistors fiber:Processing and properties", Sensors and Actuators A:Physical, vol.135, no. 12, pp. 394-404, 2007.

[3] T.Wang, X.M.Chen, H.H.Zheng,"Dielectriccharacterristics and tenability of barium stannatetitanate ceramics", Journal of Electroceramics, vol.11. pp. 173-178, 2013.

[4] J.Wang, H. Wan, Q. Lin," Properties of a nanocrystalline barium titanate on silicon humidity sensor", Measurement Science and Technology, vol. 14, no 2, pp.172-175, 2003.

[5] J.Gao, D.Xue, W.Liu, C.Zhou, X.Ren, "Recent progress on $BaTiO_3$-based piezoelectric ceramics for actuators application", Actuators, vol.6, pp. 1-20, 2017.

[6] D.Sitko, W.Bak, M.Garbartz-Glos, M.Antonova, I.Jankowska-Sumara, "Effect of MnO_2 doping on the dielectric properties of barium titanate ceramics", Ukrainian Journal of Physical-Optics, vol.13, pp.34-43, 2012.

[7] Y.Yan, S.R.Zhang, X.H.Zhou, B.Tang, "Effects of Nb_2O_5 doping on the microstructure and the dielectric temperature characteristics of barium titanate ceramics", Journal of Marerials Science, vol.44, no.14, pp.3751-3757.

[8] S.A.Solopan, A.G.Belous, O.I. V'yunov, L.L.Kovalenko, "$BaTi_{1-x}Sn_xO_3$ solid solution: Solid –phase and sol-gel syntheses and characterization", Russian Journal in Organic Chemistry, vol.53, no. 2, pp. 197-203.

[9] S.Devi, S.Jain, A.K.Jha, "Effect of sintering temperature on dielectric properties of tungsten doped barium titanate",17th IEEE International Symposium on the applications of ferroelectrics, vol.3, pp.1-5 , 2008.

[10] M.R.Panigrahi, S.Panigrahi, "Synthesis and microstructure of Ca-doped $BaTiO_3$ceramics prepared by high-energy ball milling", Physica B: Condensed Matter., vol. 404, no. 21, pp. 4267-4272, 2009.

[11] M.Aghayana, A.Khorsand.Zak,M.Behdania, A.ManafHashimc, "Sol-gel combustion synthesis of Zr-doped $BaTiO_3$nanopowders and ceramics: Dielectric and ferroelectric studies", Journal Ceramic International, vol.40, pp. 16141-16146, 2014.

INFORMATICS AND COMPUTER SCIENCE, AND ENGINEERING EDUCATION – ORAL SESSION

Development of a Smart City Portal Based on Open Data Sources and Open Web Map Services

Leonid Stoimenov[1], Nataša Veljković[2] and Miloš Bogdanović[3]

Abstract –**The goal of the paper is to examine the potential of developing a smart city portal based on open data sources and open Web map services for the purposes of introducing smart solutions in less developed countries. In the paper we will introduce a smart city solution for Niš city named GinisLS, explain the concept and architecture behind it, and demonstrate its usage.**

Keywords –**Smart cities, open data, open web map services.**

I. Introduction

Technology has radically changed the way of life in the last decade, imposing itself as default companion in everyday activities. Today's cities are the engines of the new data economy, based on digital services such as on-demand transport, intelligent water management, responsive lighting, and distributed energy resources [1]. Such digital services, as a part of smart city environment, are rapidly replacing the legacy infrastructures and service delivery models [1]. Despite of this, probably the most valuable resource of a city are its residents [2].

Many definitions of smart city exist, but none has been universally acknowledged yet. There is a high level of agreement in the literature that there is no common definition of a smart city yet [3, 4]. According to literature, smart city and digital city concepts are mostly used and developed from 1993 onwards [4].The concept of smart city embraces definitions depending on the meanings of the word "smart", as intelligent city, knowledge city, ubiquitous city, sustainable city, digital city, etc.

According to British Standards Institution, smart city refers to the effective integration of physical, digital and human systems in the build environment to deliver a sustainable, prosperous and inclusive future for its citizens [5]. Giffinger et al. explained that smart cities use information and communication technologies (ICT) to be more intelligent and efficient in the use of resources, resulting in cost and energy savings, improved service delivery and quality of life, and reduced environmental footprint [6].

But what does it take to create a smart city? First of all, people that would recognize the importance of the concept and more prominently that would understand its significance in the context of the not so far away future. Secondly, strategies and well defined goals based on well assessed current state. Thirdly, experts that would implement adopted strategies, and fourthly: finances. The stronger the city is, economically and financially, the easier it is to direct its capacities towards planning better, smarter and more efficient resource usage.

As a consequence, becoming smart comes with a high price. Investments are enormous and custom solutions are extremely expensive. Developing countries around the globe share the same struggles and anticipations. Lacking in budgetary resources, they have a significantly limited investment power and are therefore far behind advanced world cities. Smart city concept requires high investments in not just one, but several areas: economy, environment, government, people etc. For cities in developing countries, such investments are very low on priority list, slowing them down in achieving any noticeable advancement and gaining any competitiveness value.

The paper is organized as follows. Section 2 describes the usage of open source technologies and their contribution to the smart city concept. In Section 3, GinisLS smart city portal is introduced, its architecture and usage are described. Conclusion is given in Section 4, followed by the list of cited papers.

II. Literature Review

Popularity of smart city idea has sparked growing skepticism across research and industry communities in the idealization of the smart city as a vendor-oriented vision of ICT-led urban growth [1]. Also, there are research groups which attempt to answer question whether it is the smart city idea utopia (according to companies and municipalities which promotes the smart city as a revolutionary utopia) or is it, on the contrary, an expression of the neo liberal ideology [3]. However, these concerns are also accompanied by growing recognition that, whether or not cities are 'smart', the proliferation of data-driven platforms requires governments to play a much more active role in the management of their cities' data assets [7].

The significance of smart city for developing countries is major: healthcare problems rise in last ten years (cancer, heart disease etc.), water availability and water consumption rise, rapid rise of waste generation and the rise of carbon footprint.

In the light of smart city requirements, challenges that are defined for developed countries need to be differently seen for developing ones. The final goal is the same for both country

[1]Leonid Stoimenov is with the Faculty of Electronic Engineering at the University of Niš, A. Medvedeva 14, 18000 Ni, Serbia, E-mail: leonid.stoimenov@elfak.ni.ac.rs.

[2]Nataša Veljković is with the Faculty of Electronic Engineering at the University of Niš, A. Medvedeva 14, 18000 Ni, Serbia, E-mail: natasa.veljkovic@elfak.ni.ac.rs.

[3]Miloš Bogdanović is with the Faculty of Electronic Engineering at the University of Niš, A. Medvedeva 14, 18000 Ni, Serbia, E-mail: milos.bogdanovic@elfak.ni.ac.rs.

groups, developed and developing, but the roadmap towards achieving it is significantly different.

Increasingly popular open technologies, and particularly Internet enabled ones, offer a whole new perspective on becoming smart. Growing open data sources and catalogues represent valuable information storages that are freely available and could be successfully utilized in building smart cities portals. Open web map services, such as Google Maps and Open Street Maps (OSM), simplify expanding such portals with geographical dimension and spatial representation of gathered data. The open market and open initiative truly are a great source of possibilities for those not being able to purchase highly proficient but indeed expensive commercial solutions. Cities from developing countries are obvious candidates that with minimal investments could make a significant progression with open technologies on their way towards transforming into more intelligent environment.

Smart Governance is seen as a combination of political participation, services for citizens and functioning of the administration [8]. Smart Government is at service to its citizens 24 hours a day, offering services that would enable citizen to quickly obtain requested information without having any knowledge on back-office systems involved in providing the response. For Government to become smart it is not necessary to invest millions in technology, it is rather needed to be astute in making the best possible use of existing resources [9]. Being innovative and open for new technology solutions and concepts are leading drivers for introducing smart governance in developing countries. Open technologies again provide a colourful palette of possible solutions. Internet of things offers vast of open source Content Management Systems that could efficiently be used in developing smart government portal and integrating it with administrative bodies in back-office domain. Opening up data and enabling free access to governmental information is a trend and it has imposed as an important feature of smart government.

Open data platforms are also available (CKAN, OGDI etc.) offering tools for easy data upload and maintenance. Open map services, such as Google Maps and Open Street Maps, enable addition of geographical dimension to smart government portal and representing governmental data in spatial context, considerably increasing data understandability. Social and geo-social networks, such as Twitter, Facebook, Four Square and others, represent valuable open sources of users' opinions, experiences and comments that could be connected with other available governmental data and easily utilized in building smart city guides. Smart Government strongly stands behind smart city, addressing financial, environmental and service challenges that public offices are facing with.

The successful examples of open source technologies joint interaction for the creation of smart city portals can be found, namely: (1) Padova Smart city portal enables harvesting environmental data from wireless Internet of Things (IoT) nodes installed on street light poles and connected to the Internet through a gateway unit [10]; (2) PlanIT Valley, a project started in Portugal to enable the enhanced monitoring of the vital signs of urban life, the condition and performance of vehicles and infrastructure. It is a vision to build the world's greenest city and establish a working template for low carbon emission for cities [11]; (3) Smart Amsterdam resulted as a partnership between businesses, authorities, research institutions and the people of Amsterdam (https://amsterdamsmartcity.com), with focus on carbon emissions reduction, energy efficiency and behavioral change, and (4) Smart Barcelona is centered on the notion 'city of people' with an aim to improve citizens' quality of life [11].

III. Open Street Maps and Open Data sources for Smart Government Portal implementation

In our research so far, we have started to develop a smart government portal for the city of Niš, using OSM and available open data sources. Niš is a city in southeast Serbia, which is one of the Western Balkans' developing countries, and as such it could be used as a perfect example for demonstrating previously described concepts and solution approaches [12].

Our solution, named GinisLS Smart Nis Portal (GinisLS), represents a Web Geographic Information System (GIS) portal that integrates freely available maps provided by OSM, as well as custom maps provided by Republic Geodetic Authority of Serbia, with local databases and open data sources. Being deployed and publicly available, it represents a fully functional GIS portal that significantly contributes to city's efficiency, increases availability of city's services to citizens and helps citizens build trust regarding their government. The development of GinisLS portal has been

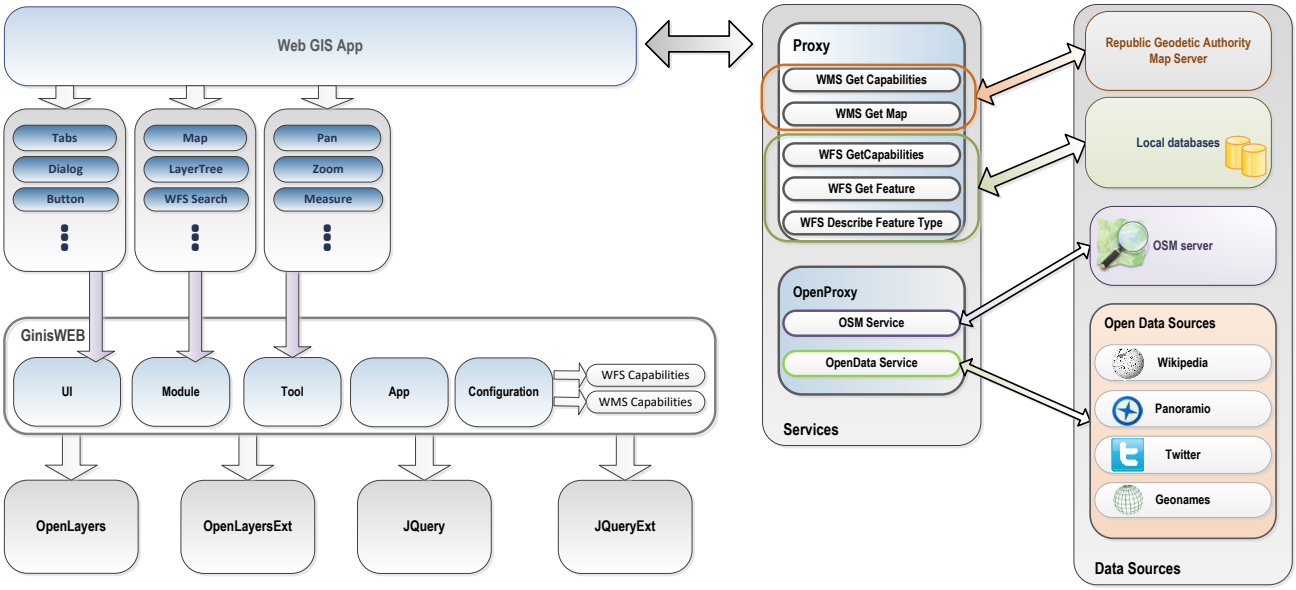


Fig. 1.GinisLSWebGIS architecture

driven by previously described and agreed main challenges for smart city development, as a practical attempt of demonstrating how open technologies indeed help implementing some of the smart city concepts aspects.

GinisLS Web GIS portal is designed to enable simple linking and integrating with standard types of spatial Web services. Built upon standard client-server architecture, it is comprised of a Web server component, GIS server and a Web GIS client [13]. GIS server is responsible for providing geo-referenced data and maps, and it communicates with OSM for obtaining maps of Niš region, but it also communicates with other Web Map Services (WMS) and Web Feature Services (WFS) for obtaining other providers' maps, concretely Republic Geodetic Authority of Serbia.

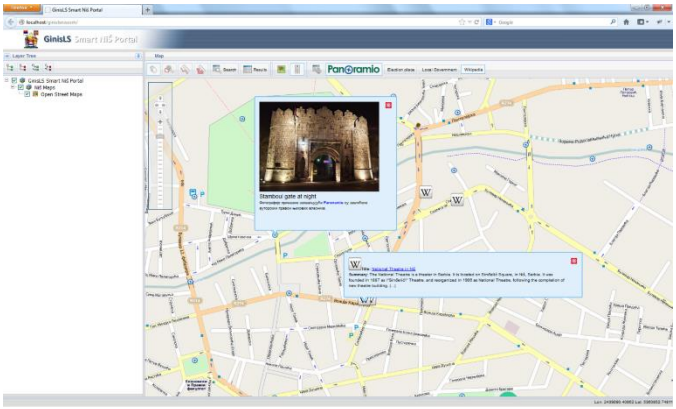


Fig. 2. Wikipedia and Panoramio tools

GinisLS Web GIS client, depicted on Fig 1, is a Web application specialized for visual representation and querying the data obtained from GIS server. This client relies on Open Layers framework and JQuery Javascript library. Because of its modular architecture, it successfully combines the advantages of centralized data control with advanced user interface, at the same time representing highly scalable solution that could be easily modified and extended with additional functionalities.

GinisLS client communicates with external data sources through Proxy and Open Proxy services. Proxy service is responsible for obtaining maps from custom providers and feature information from local data sources. OpenProxy service consists of two components: OSM and Open Data services, and is responsible for communicating with open data sources. OSM service requests maps from OSM server, while Open Data service communicates with Wikipedia, Geonames, Panoramio and Twitter for obtaining open available data.

GinisLS portal integrates the existing local data sources with external ones, such as Wikipedia, YouTube videos, Panoramio images, Geonames services and others, offering the possibility of custom geo-referencing and pinpointing dynamically obtained information to a concrete location. It provides data organized in different layers: maps, touristic information, investment locations, election places, Panoramio images, Wikipedia sources and others (Fig 2).

Touristic representation of a city is an important aspect that directly influences the promotion of city's offerings and its rating among potential visitors. Allowing tourists to have an insight into city's cultural heritage and make an impression of its atmosphere, most certainly contributes to growth of touristic visits. This feature could also be observed as one aspect of city's smart economy as increased touristic visits directly influence increased city's income, thus helping decrease the urban poverty.

GinisLS portal provides an excellent tourism support by enabling users to visualize touristic sights: theatres, movies, hotels, libraries, museums, national monuments, sport clubs and others, available from local databases. By including Local Government layer, map area is being populated with suitable markers pinpointing to exact locations of contained features. As shown in Fig 3. each marker is clickable and provides additional information regarding the object it is related to: object's name, type, address, contact, visiting hours, brief description.

Smart economy city aspect is even more addressed by offering the possibility of reviewing potential and ongoing investment locations. By simply including the Investments layer, a set of markers spreads on the map area indicating the exact locations of potential construction plots. For each location, user will be provided with additional information by clicking on a marker: location status (available, occupied, in the process of construction, finished etc.), location description, obtained permits, general construction plan information, as well as related YouTube video materials. This feature offers benefits to both investors and citizens.

Investors could obtain an insight regarding possible

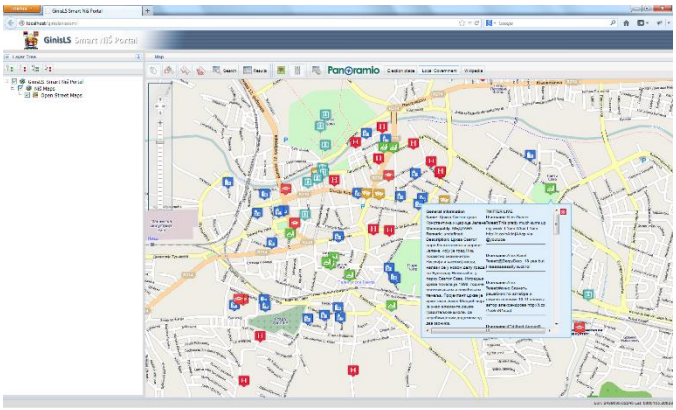


Fig. 3. Visualization of city's touristic sights

construction locations, while citizens could keep up with urban plans that affect their living surrounding and at the same time be updated about possible buying locations. City's Administration becomes more efficient by utilizing this feature as assistance tool for overseeing construction resources with the purpose of better planning of urban land exploitation for housing requirements.

Open data sources truly represent significant sources of freely available information that cities could quite efficiently utilize for bringing government closer to citizens. Users are valuable volunteers that on daily bases upload vast amount of newly generated content contributing to growth of open communities such as Wikipedia, Geonames, Twitter and Panoramio. Wikipedia is the world most used online

encyclopaedia and a source of large number of user edited articles related to almost any known topic. Geonames, the best known worldwide geographical database, provides suitable Web services towards Wikipedia, enabling simple articles obtaining based on provided spatial context.

GinisLS provides Wikipedia tool that populates visible map area with Wikipedia articles related to available city objects and pinpoints them to objects' locations. After clicking on a marker, a pop-up window appears with article preview, offering the possibility of reading full article on official Wikipedia page. In such way, a user could easily obtain additional information regarding objects of interest and from a popular and reliable source.

Having in mind the increased popularity of Twitter, micro-blogging network community, we have included tweets as additional touristic sights' description, thus enabling users to not only get general data and historic facts on objects of interest, but also to have an insight into other experiences and impressions, regarding these objects (Fig 3). That kind of information highly increases data reliability as well as users' trust in such provided context, but more importantly it represents a way for including youth into portal's content creation, recognizing their social importance and acknowledging their opinions. This aspect contributes to smart people and smart living concepts as well.

Panoramio is one of the greatest photo-sharing communities that enable users to upload their photos and link them to geo locations. It provides a suitable API for accessing the imagery content and thus enables free of charge data utilization. We have used the API and implemented Panoramio tool that populates visible map area with markers referencing available Panoramio photos, similar to addition of Wikipedia articles. By clicking on Panoramio map marker, a pop-up window opens displaying a photo with available related description and linking to an official Panoramio Web page where it can be previewed in original size. Wikipedia and Panoramio tools increase portal's reliability, data understandability and provide better context for users browsing the portal. In such way they contribute to better educating citizens about city's locations and indirectly add to smart people city's aspect.

IV. Conclusion

As been thoroughly discussed previously in the paper, there are six city aspects that most accurately define 'smart city' concept: economy, governance, people, mobility, environment and living, each addressing specific city issues. Vast of available open technologies truly make their implementation easier, which is especially convenient for developing countries. In this paper, we have presented a part of open data sources and open Web map services and their potential of being used for introducing smart solutions in developing countries. Our smart city solution, named GinisLS, uses open data sources and free mapping libraries. It represents a fully functional GIS portal capable of improving local government efficiency and increase availability of local government services with minimal costs. Although we believe our solution makes a step forward in a city becoming smarter, to cite Boyd Cohen, opening twenty databases for public use does not make a city smart. To transfer this statement into the scope of this paper, utilizing open Internet enabled technologies to build a smart portal does not mean that the city has classified into smart cities group. Going smart is a long-term process that needs to begin with years of planning and strategizing, as it requires a thorough cooperation between all city's authorities and a 180° shift from what is being known as normal city functioning. Nevertheless, initiatives, such are those described previously in the paper, most definitely place a city on the road towards becoming smarter.

References

[1] S. Barns, “Smart cities and urban data platforms: Designing interfaces for smart governance”, City, Culture and Society, vol 12, pp. 5-12,2018.

[2] G. R. Ceballos and V. M. Larios, “A model to promote citizen driven government in a smart city: Use case at GDL smart city”, 2016 IEEE International Smart Cities Conference (ISC2), Conference Proceedings, pp. 1-6, Trento, 2016.

[3] G. Grossi, D. Pianezzi, “Smart cities: Utopia or neoliberal ideology?”, Cities, vol. 69, pp 79-85, 2017.

[4] A. Cocchia, “Smart and Digital City: A Systematic Literature Review”, In: R. Dameri, C. Rosenthal-Sabroux(eds) Smart City, Progress in IS. Springer, Cham, 2014.

[5] BSI:182, Draft PAS 182:2014. Smart city concept model – Guide to establishing a model for data interoperability, Draft 2.0 for public consultation, In Institution, T.B.S. (Ed.) London, 2014.

[6] R. Giffinger, C. Fertner, H. Kramar, R. Kalasek, N. Pichler-Milanović and E. Meijers, Smart cities: ranking of European medium-sized cities, Final report, Retrieved 3 May, 2018 from: http://www.smart-cities.eu/download/smart_cities_final_report.pdf.

[7] C. Pettit, S.N. Lieske and M. Jamal, CityDash: Visualising a Changing City Using Open Data. In: Geertman S., Allan A., Pettit C., Stillwell J. (eds), Planning Support Science for Smarter Urban Futures. Lecture Notes in Geoinformation and Cartography. Springer, Cham, 2017.

[8] European Smart Cities, Smart Cities: Ranking of European medium sized cities, Retrieved 18 June, 2013 from http://www.smart-cities.eu/index2.html.

[9] A. Di Maio, Smart cities are not intelligent: They are astute, Retrieved 3 May, 2018 from http://blogs.gartner.com/andrea_dimaio/2012/12/14/smart-cities-are-not-intelligent-they-are-astute/.

[10] A. Zanella, N. Bui, A. Castellani, L. Vangelista and M. Zorzi, “Internet of Things for Smart Cities”, in IEEE Internet of Things Journal, vol. 1, no. 1, pp. 22-32, Feb. 2014.

[11] Angelidou, M., “Smart city policies: A spatial approach”, Cities, vol. 41, pp.S3-S11, 2014.

[12] M. Bogdanovic, A. Stanimirovic, N. Davidovic and L. Stoimenov, “Web-based GIS solution for electric power supply network data management”, 12th AGILE international conference on geographic information science, AGILE 2009, Conference Proceedings, 2009.

[13] B. Cohen, What exactly is a smart city?,CoExist. Retrieved 3 May, 2018 from http://www.fastcoexist.com/1680538/what-exactly-is-a-smart-city.

The Challenges of Engineering Education in a Current Project Environment

Vangel Fustik[1], Atanas Iliev[2], Nevenka Kiteva Rogleva[3], Dimitar Dimitrov[4] and Natalija Petrova[5]

Abstract – **As the world business and social environment continues to change, today's engineering education is faced with the challenges that need new approach in organizing it's curricula. The paper analyze the nature of challenges and needed topics or contents of engineering curricula enabling creative and practical knowledge and skills for running complex projects. The identified trends of current project environment, in which significant part of the business is located, are listed and commented to understand increased challenges in engineering education concerning project management and project risk management**

Keywords – **Engineering education, Project management, Project management competence, Undergraduate studies.**

I. Introduction

Engineering education is faced with challenges that need new approach in organizing a companies, running a business and creating the everyday working habits in a new project environment. A certain surveys [1] worldwide highlighted this phenomena indicated that top two skills desired from new hires are project management and business process management. In addition in Europe a survey for appointment data [2], found that while there was a modest increase in overall IT recruitment, the need for IT management positions with project management skills grew faster. This trend was attributed to increased levels of confidence, leading to the implementation of the new projects requiring management professionals, while lower-lever technical tasks were frequently outsourced.

Projects complement regular business processes and are often the main vehicle for strategy implementation in an organization. The processes are how work gets done on a daily operational basis, projects are important efforts to meet the organization's strategic objectives and to implement changes. However, as projects take a more crucial role in organizations, project management discipline has become a leadership competence. In today's workplace there could be hardly found managers who just "do their job". Especially in engineering environment managing projects alongside one's regular function is a daily reality, and project success has become an important factor in any manager's performance evaluation.

[1]Vangel Fustik, [2]Atanas Iliev, [3]Nevenka Kiteva Rogleva, [4]Dimitar Dimitrov, [5]Natalija Petrova are with the Faculty of Electrical Engineering & Information Technologies University of Skopje, Rugjer Boshkovikj, Skopje 1000, Macedonia, E-mails: [1]vfustic@feit.ukim.edu.mk,[2]ailiev@feit.ukim.edu.mk, [3]nkiteva@feit.ukim.edu.mk,[4]ddimitar@feit.ukim.edu.mk, [5]npetrova@feit.ukim.edu.mk.

II. Engineering Working Environment and Project management Maturity tracking

In a few recent years Prof. Ambler and Martinich members of IEEE – Engineering Management Society [3], have been conducted several focus groups to understand the needs of engineering companies with respect to their engineering managers. Their survey's conclusion was: "Recently appointed engineering managers lack the business skills and interpersonal skills necessary to successfully deliver their engineering projects. They are well trained as engineers, but they are not receiving the training they need to succeed as project managers." Upon their professional experience they are convinced that it is a true on a global scale. They have seen the problems in projects that result in China, India, Australia, Canada, Europe and the United States. Projects fail, they stated, not because of engineering challenges, but because in the projects engineering managers lack communication, collaboration, negotiation and business alignment skills. There are few maturity tracking elements that are worth mentioning and that could help to understanding a "new wave". Following such tracking elements and the identified trends, our curricula for project management is created in a way to consist practical examples that are selected from the project management practice. And, consequently, besides the new and worthy knowledge the students are obtaining, the lectures are more interesting and interactive. The students can explore themselves in a creative manner.

A. More Effective Project Teams

Remote or distant teams face the challenge of geography and diversity. To create a modern project team conditions is equal to the "art of management" in each company's mission control [4]. As an example, the following Project management practice made the effective teamwork possible: *Project Plan and Project Baseline, Project Management Control (Continuous Critical path), Human Relations Emphasis and Risk Management Techniques.*

B. The Risks management

The sub-discipline of risk management has advanced in areas of risk identification, analysis, planning and management [5]. New techniques are on the horizon and we should explore them in our maturity tracking, especially in academia.

C. Project competencies

Regularly in the beginning project practitioners focused on their subject matter expertise, such as financial analysis, telecommunications, power engineering, and IT design, or marketing plan. Those who became involved in projects were forced to include scheduling, status reports, and risk analysis in their competencies. In today's engineering challenging practice the project practitioners should move towards general business awareness. In that sense they should explore: *facilitation (not only meeting management), financial knowledge, leadership, creativity, problem solving and/or decision making*.

D. Basic training for a Project Manager

The Basic training should include the: *Roles of Project Manager, Types of Management Skills, Managing Former Peers, Building a Team, Effective Decision Making (Problem Solving), Authority Delegating*. Here, we will emphasize the importance of Types of Management Skills and to be aware why are important for the Project Manager. The effective Project Managers, besides being well educated in the Project management basic concepts and tools, must develop three types of skills: technical, conceptual and human relations. The companies in high-tech sector reveals different organizational models to describe the roles of their Project Manager. Many parameters influence a company's approaches such as: Company budget, Team size, Capability maturity model level, Soft skills, Organization's depth, Technical expertise, Firm's project management expertise, Client relationship, etc. The balance needed to mitigate the risk introduced when splitting task responsibilities should be found by the company's management. The clear definition of roles and commitment to those roles reduce the chances that jobs will be overlooked. However, the dual role of technical expert and project manager enables the project manager to be acutely aware of risks and provides an understanding of company's commercial model, customers and competitors. In accordance with such recommendations we include basic training for potential project managers enabling them a short stay in industry to be in the center of the events and to understand project manager's duties and responsibilities.

III. Remarks on the Today's Engineering Workforce and Needs

However, today's work climate has changed. The number of engineering jobs has increased and engineering positions are more varied, while greater job mobility has reduced the opportunity for engineers to take advantage of longer on-the-job training periods [6]. As a consequence, engineering educators are been challenged to look at their curricula and retool coursework to incorporate non-traditional information and subject matter. Students' need for theory understanding to enable engineering problem solving remains a top priority, but today's graduates could not enter the workforce only with technical skills. Engineering specialties have become varied and diverse, and therefore industry is demanding new "rounded" engineers whose initial skills stretches beyond technical competency having supporting work skills. The Needs of the parties are linked and identified and could be listed as follow:

Student Needs: Strong theory foundation; engaging; real-world application work; creativity/problem solving skills; critical work skills; risk analysis.

Industry Needs: Technical competency; communication skills (written, verbal, presentation); leadership and teamwork skills; enthusiasm and personal drive (a sense of a mission); intuitiveness; integrity; other supporting work skills.

Curriculum Needs: Theory; hands-on trough lab work and practical examples; more intense hands-on applications work using industry-based scenarios and problems; opportunities to develop basic yet critical supporting work skills.

IV. Curriculum Needs and Project Management Trends

In Graduates may initially be hired primarily for their technical skills while long-term career success is more dependent on non-technical skills. And these skills should be integrated throughout the curriculum, rather than be taught in isolation. That is a way the students will receive a full-picture view of real world of engineering. The professors in engineering and especially in emerging and fast changing technology, should seeking ways to introduce more workplace related experience earlier in the curriculum and to incorporate "supporting work skills" such as: communication skills, project planning and execution, project manager leadership, risk assessment, time management, decision making and ethics. However, to teach the undergraduate students project management theory and practice is not easy task, since there is a lot necessary and preparatory business things to be learnt. In that sense, curriculum has to follow the underway trends in project management and deliver the knowledge and skills that will prepare the engineers for project management profession. There are discussions [7] and studies [8] that reflect this issue and that could be easy checked-out among companies and young professionals.

The selected modern trends in project management are listed as follow:

1. Agile will gain more popularity in IT projects and continue to be accepted in wide range of industry projects. The Agile approach [8] is cost-effective, relies on self-organizing teams, and appeals to human needs like autonomy, mastery, and purpose. However, to be successful, it relies on timely frequent feedback, that could be needed in complex projects. Up to now there are discrete examples and sampling of the very few industries that are even familiar with project management to begin with. And that's about to change. Recently, agile project management has visibly

crept into all industries. IT no longer has a monopoly on the project management technique. Several industries, including marketing, finance, and construction, were beginning to adopt the Agile development framework. Various research has shown that the Agile methodology improves communication, makes teams more adaptive to change, and has an overall higher return on investment, especially for small to midsize teams. There is the rise of business Agile. Moreover, DevOps (development and Operations) will be considered a part of Agile. Atlassian [9] explains this point well: "DevOps seeks to bring that Agile attitude toward change to a new audience: IT operations." If a project management community, can accept marketing Agile, construction Agile, and business Agile, then IT operations Agile should be no stretch.

2. Risk management will be an incredibly popular topic for project managers to be learnt and applied. In addition, with the growth of Agile, risk management will become a necessity for project professionals seeking new opportunities. While there is software available to help identify risks, the project managers have to be creative enough to identify potential budget or scope overreach and maintain organization over your projects.
3. PMP certification will become more popular, but changes are needed. Since there are graduate engineers struggle to find jobs, many will turn to a PMP certification to try to get a leg up in the project management field. Getting a PMP carries quite a bit of weight—it allows young workers to learn the knowledge of the field, attain networking opportunities, and distinguish themselves from the competition. The Program Management Improvement and Accountability Act (PMIAA) signed in 2016, increase the importance of certification and formal project management job titles in the government. There is an increasing need for professionals to understand the foundation of project management through formal training and develop their skill sets in communication, leadership, and stakeholder engagement. However, recently, since the companies just want someone who can manage projects and do it well, there is also a trend loosing the interest in PMPs. The PMP exam and PRINCE2 have been around since 1984 and 1989, respectively. There are a lot of certified project managers and many of them have begun calling into question the value of their certification, concluding that PMP certifications just aren't worth it. The reason for that, they are arguing, is in fact that exam content isn't updated often enough to keep pace with today's small business needs.
4. The Need and project requirements – not organizational chart will create project team. Celoxis [10] smartly predicts that project requirements will chip away at the organizational chart. Project complexities will decide how the teams will be structured, what tools they will use and how the execution will happen. It's time for enterprises to take a cue from the manufacturing sector and bring that level of rigor to their project management methodology. Another study [11] suggests, that managers should be as aware as possible of how interdependent relationships are distributed across a cohort before they do any reorganization.
5. Remote teams will become normal. According to Intuit [26], 40% of the American workforce will be freelancers, contractors, and/or temporary employees by 2020. These contractors are already working remotely using sites like Upwork, Freelancer.com, and Demand Media. Also, full-time employees are increasingly telecommuting. A recent Gallup study [12] found that 37% of Americans have telecommuted and it's even more likely that those working in white-collar professions have telecommuted (44% vs. 16% in blue collar jobs). That means that the technology can help with the growing presence of remote teams. It is well known that distributed teams allow for companies to break down geographical borders in their quest to find top talent in non-traditional businesses.
6. The rise of BYOD will cause project management software to have more ticketing options. BYOD ("Bring Your Own Device") has been making waves in all industries throughout last years. The concept is simple: companies are allowing employees to use their own smart phones, computers, tablets, and other electronics for work purposes. This trend is contributing to growth in mobile project management software. In the business environment it is pointed out that BYOD will also pose a number of challenges to the project management field such as, dramatically increase of the complexity of technical support provided by the IT Help Desk and the development of software for multi-platform use. In other words [13], BYOD means that not all technologies are streamlined in the office. While there is specific help desk software for IT (like Samanage IT Help Desk & Asset Management or C2 Atom), project managers will likely prefer help desk ticketing included in their multi-platform project management software. Examples include Workfront and Clarizen.
7. Emotional intelligence will be most desired skill for new project managers. It is (by Psychology Today) the ability to identify and manage your own emotions and the emotions of others. In projects it is project management personality assessments of the right candidate that "feel," has charisma, or has developed "soft skills." Whatever it means, emotional intelligence (EQ) will get a lot more important to project managers. In general EQ include three skills:
   - Emotional awareness, including the ability to identify your own emotions and those of others;
   - The ability to harness emotions and apply them to tasks like thinking and problems solving; and
   - The ability to manage emotions, including the ability to regulate your own emotions, and the ability to cheer up or calm down another person.

With all value statements aside, this is a trend that is trickling in from the business community and the political climate as a whole. It is for sure that the Internet of Things (IoT) and Artificial Intelligence (AI) are going to affect most small businesses in 2018. The growth of IoT and AI in union are going to change how project management is effectuated. In

some industries, that change will be dramatic. The most notable effect from these concurrent trends in 2018 will be in project management role and project management software itself.

Artificial intelligence is slowly making its way into project management tools, for example an options such as *Aurora, Clarizen, ClickUp, Forecast, and Rescoper*. These tools are using AI to automate many existing project management tasks, including matching talent to tasks, reducing calculations for level for effort, providing a hub for knowledge management, and creating reports with untiring objectivity [13].

However, it is not easy all above mentioned trends to include in the courses and to teach the students, but as professors we are creating the environment and apply methodology such trends to be more familiar with their project tasks and seminar works. Their success depends on our efforts to be open to new knowledge and practice and creative approach making them ambitious to be well prepared for the professional work [14, 15, 16].

V. CONCLUSION

Since the education is a life-long process, introducing project management to meet the needs of the students and industry (and not to forget the entrepreneurs) is a challenging mission. It is no doubt that our industry, companies and organizations will embrace, value and utilize project management and attribute their success to it. However, the efforts that are making both University and industry working together, a common engineering education in project management will find the way to ensure that our graduates and master degree students have all of the tools they need to succeed and continue to contribute in economic and technological growth. The Engineering universities in their curriculum have to be up-front in the following, understanding and applying new trends in project management. It is a very complex task the professors are faced with and their mission is challenging that ever before.

REFERENCES

[1] US CIO Magazine, “The State of the CIO 2006” survey, 2007.
[2] Computer Weekly/SSL Europe, “Survey of Appointment, Data and Trends”, 2007.
[3] L. Martinich, Training for New Engineering Managers, IEEE EMS Newsletter, Vol 56, No.4, 2006.
[4] R. Foti, Mission Control, PM Network, July 2003.
[5] Project Management Institute, www.pmi.org.
[6] “The Changing Nature of Engineering”, 1998, ASEE PRISM, ASEE.
[7] Published December 29th, 2014 by Rachel Burger in Project Management, The 5 Biggest Project Management Trends for 2015.
[8] V. Fustik, The Advantages of Agile Methodologies Applied in the ICT Development Projects, International Journal on Inform. Technologies & Security, № 4 (vol. 9), 2017.
[9] Atlassian, Agile Coach.
[10] Celloxis, All-in-one Project Management Software.
[11] Intuit 2020 Report, Oct. 2010.
[12] Jeffrey M. Jones, Telecommuting for Work Climbs.
[13] Rachel Burger, Capterra, The 6 Biggest Project Management Trends in Project Management, 2017.
[14] V. Fustik, Project and Risk management, lecture notes, FEIT, Skopje, 2014.
[15] A. Iliev, Project Analysis, lecture notes, FEIT, 2015.
[16] N. Kiteva Rogleva, lecture notes, FEIT, 2015.

Possibilities of Synthesized WDFs Models of Lowpass Filters Exhibiting Equiripple Characteristics

Biljana P. Stošić and Nebojša S. Dončov

Abstract – **This paper introduces possibilities of synthesized lowpass wave digital filter models. The design process starts by calculations of element values of the reference lowpass structure for given specifications. Equations yield the network elements of the reference structure in order to simplify design procedure. Then, the design is solved by direct transformation of reference structure, i.e. its one-port elements and interconnections, into structure of wave digital filter (WDF). Different case studies of synthesized WDFs are presented and compared here.**

Keywords - **Digital signal processing; Wave digital filters; Lowpass ladder structures; MATLAB/Simulink; Chebyshev approximation, Direct transformation method.**

I. Introduction

Digital filters are essential part of most digital communication systems. They have important role in digital signal processing (DSP) systems. The most attractive properties of the wave digital filters, short WDFs, include an excellent sensitivity and consequently the low accuracy requirements for the register wordlength, higher dynamic range, and automatic insurance of stability. The basic concept of WDFs has their origins in the field of filter design. It is developed by A. Fettweis [1-3]. The theory, principles and properties of wave digital filters are well explained in variety of publications delivered by several authors [1–10].

In the literature, variety of design methods and algorithms are available for WDFs. The most valuable papers are [6-8]. In the chapter [6], "Design of the wave digital filters", authors propose a very simple procedure for design, analysis and realization of low-pass, high-pass, band-pass and band-stop wave digital filters from reference LC filters given in the ladder configuration. The input data for the LC filters are obtained from introduced tables for different attenuations. MATLAB program for calculation of coefficients of the wave digital filters is obtained directly from the structure. The paper [7] presents a direct design method for highpass ladder wave digital filters.

WDFs are made up of individual modules that represent components of a system. In the case of model synthesis, this means that components of the system, such as resistors, capacitors, inductors and voltage sources are replaced with theirs wave digital counterpart. Each element contains a wave input, a wave output, and a value which determines port resistance. These basic building blocks are known as one-port elements. An adaptor is a memoryless interface which interconnects one-port wave digital elements.

This paper is subjected to design of lowpass WDFs. A detailed description of different synthesis processes is given here. The contributions of this paper include an introduction to wave digital structures, as well as new aspects that relate to design procedure and WDF model possibilities, and finally use of MATLAB/Simulink to simulate filter characteristics. One can follow the described procedure and develop wave digital filter for any possible required filter specifications.

Different case studies of WDF design are presented and can be applied to any filter structure with even or odd order [5]. The presented case studies of this example of design, modeling and simulation of WDFs is very good for engineering education. A computer-aided design (CAD) software-based engineering drawing course had a positive effects on developing engineering students' spatial visualisation skills.

The multiport parallel and series adaptors and their WD models are described in [1-4,9]. A brief review of WDF design procedure is described in Section 2. Then, different forms of lowpass WDFs are synthesized and simulated here. They can be implemented directly in the Simulink toolbox of MATLAB environment. In this section, block-based wave digital models of these filters are described. Response in WDFs can be found by use of formed block-diagram networks and some basic MATLAB functions. Possibilities of the synthesized models in different case studies are described in Section 3. Conclusion of the paper is given in Section 4.

II. Design Procedure of Wave Digital Filters

Special emphasis is put here on the synthesis procedure and simulation of lowpass filter. This section presents shortly efficient algorithms for the automatic generation of lumped elements of reference filter structure and further corresponding blocks of WDF structure.

At the beginning of the WDF design and synthesis processes one should specify some filter requirements.

Here, a design of a lowpass wave digital filter that has a passband attenuation upper edge frequency of $f_p = 1.5\ KHz$, and stopband attenuation lower edge frequency of $f_s = 2.5\ KHz$ is described. Suppose a Chebyshev prototype with passband ripple (maximum passband attenuation) of $0.5\ dB$ and minimum stopband attenuation of $30\ dB$ is to be used in the design.

A. Synthesis of WD Model

WDF is designed by transforming the network of reference low-pass filter structure. Let's calculate firstly the reference filter elements. The entered filter specifications include passband upper edge and stopband lower edge frequency. The first step is calculation of sampling frequency, and choosing

The authors are with University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mails: biljana.stosic@elfak.ni.ac.rs and nebojsa.doncov@elfak.ni.ac.rs

the most suitable electrical length $\theta_p = \pi/2$. The sampling frequency is calculated as $F_o = 4 \cdot f_p = 6\,KHz$. With determined sampling frequency, the normalized stopband edge frequency for the reference structure and the selective factor can be found as follows $\Omega_s = \tan(\pi \cdot f_s / F_o) = 3.73$, and $\gamma = \Omega_s / \Omega_p = 3.73$, where the normalized passband frequency is $\Omega_p = 1$. In order to simplify design of the WDFs, simple equations yield the network elements of the reference structure. They are given in [5]. The designed values of the reference structure elements are $g_1 = g_3 = 1.6$, and $g_2 = 1.1$.

The designed filter is of order 3, and in case of filter Type B that means that it has 2 series inductors separated by one shunt capacitor, Fig. 1 [5].

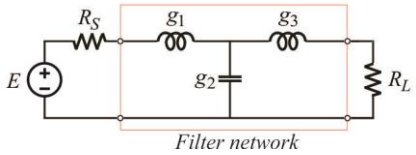


Fig. 1. *The structure of reference filter - Type B.*

The WDF structure is reached by applying direct transformation of its elements and their interconnections, i.e. by connections of one-port elements and two- and three-port parallel and series adaptors. Direct transformation of the refrence structure into WD model happens within several steps, wherein some intermediate model is constructed. Starting point in the synthesis procedure is a structure model in the form of an electrical circuit (reference filter ladder structure [5]). First of all, identification of one-port elements and adaptor types in a structure model is required. After that, characterization of adaptors should be done (determination of port resistances and coefficient values). The network structure of a 3rd order Chebyshev lowpass filter and its adaptor connections and port resistances for the digital form transformation are depicted in Fig. 2.

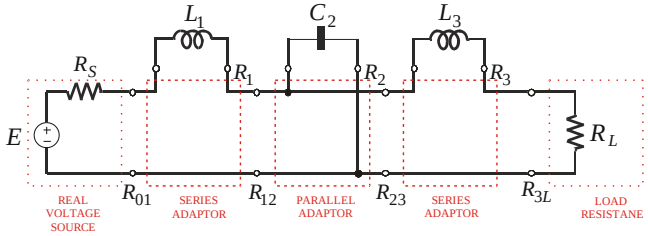


Fig. 2. *A structure model of reference filter.*

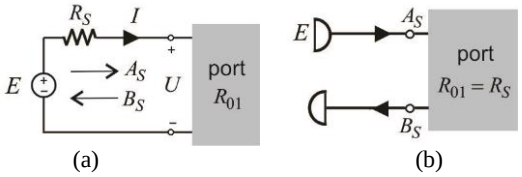
Starting with the K-domain representation of one-port element it is very simple to derive WD equivalent of the observed element. The one-port elements and their WD models are already presented and used in practice [1-5]. Some of them are shortly reported here once again because it could be useful for those readers start to deal with WDF theory. One-port elements identified in the reference structure are discretized by use of bilinear transform. WD representation of capacitor is simply a delay element, i.e. the reflected wave is simply a unit-delayed version of the incident wave ($B = z^{-1} \cdot A$). In case of inductors, the reflected wave is a unit-delayed incident wave with a change in phase ($B = -z^{-1} \cdot A$). These models are shown later in synthesized WDF models.

B. Different Case Studies of WDF Model Design

The building blocks of a structure model shown in Fig. 2 are resistive voltage source, inductors, capacitor and resistive load. A filter network driven by resistive voltage source indicates two case studies of filter design. That allows one to design wave digital filter in two possible ways: CASE A - Synthesis procedure starts from the driven side or CASE B - Synthesis starts from the load side.

CASE A: The first case study is the digital network synthesis **starting from the driven side.** In this case the first wave port resistance is selected as $R_{01} = R_S$. Port resistances corresponding to one-port inductors are set to $R_1 = L_1$ and $R_3 = L_3$. Port resistance corresponding to one-port capacitor is set to $R_2 = 1/C_2$.

Resistive voltage source with independent variables U and I in K-domain, and A_S ans B_S, and port resistance R_{01} in WD-domain is shown in Fig. 3*a*. Its wave digital representation for case $R_{01} = R_S$ is depicted in Fig. 3*b*.



(a) (b)

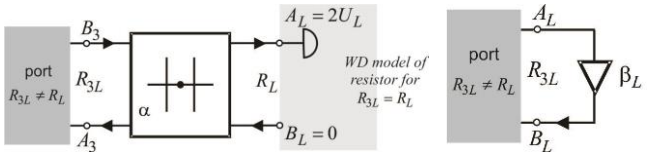
Fig. 3. *Resistive voltage source: (a) one-port analog network, and (b) wave digital model for port resistance chosen to be* $R_{01} = R_S$.

For network given in Fig. 3*a*, Kirchoff's voltage law gives $E = U + R_S \cdot I$, and using the relations for wave variables in the frequency domain $A_S = U + R_{01} \cdot I$ and $B_S = U - R_{01} \cdot I$, one can obtained relation $E = \frac{A_S + B_S}{2} + R_S \cdot \frac{A_S - B_S}{2R_{01}}$. In case of $R_{01} = R_S$, the relation is $A_S = E$, i.e. the incident wave is equal to voltage of the resistive voltage source. WD model of this resistive voltage source is shown in Fig. 3*b*.

One-port network of the resistive load R_L, and its WD representations for different cases $R_{3L} = R_L$ and $R_{3L} \neq R_L$ are described further. The following part of paper presents two different forms of WD-representation of the resistive load (Form I and Form II).

Form I: The first WD model of resistive load for port resistance chosen to be $R_{3L} \neq R_L$ is represented as cascade of two-port adaptor and wave digital model of resistor for case $R_{3L} = R_L$, Fig. 4*a*. Two-port adaptor coefficient is $\alpha = \frac{R_{3L} - R_L}{R_{3L} + R_L}$. In the case of one-port resistor with port resistance of $R_{3L} = R_L$, the reflection coefficient is equal to zero, $\frac{B_L}{A_L} = \frac{Z_L - R_{3L}}{Z_L + R_{3L}} = 0$, because of $U_L = R_L \cdot I_L$ and $R_{3L} = R_L$. This results in $B_L = 0$, i.e. the reflected wave is

zero. The incident wave is $A_L = 2 \cdot U_L$. This WD-model of resistive load for case of $R_{3L} = R_L$ is shaded part in Fig. 4*a*.



(a) (b)

Fig. 4. *Symbolic models of wave digital network for resistance in case of* $R_{3L} \neq R_L$.

Symbolic representation of the synthesized WDF structure for a given electrical circuit is shown in Fig. 5 and obtained by direct transformation of network from Fig. 2 by use of wave digital models of real voltage source and load resistance.Wave digital formulation uses matched (reflection-free) ports to eliminate reflections and avoid delay-free loops.

Response in the formed model can be easily found by use of block-diagram network drawn in Simulink toolbox, Fig. 6, and some basic MATLAB functions allowing for accurate and fast modeling and analyzing of circuits. The frequency responses in dBs (transmission and reflection parameters) obtained from this model are given in Fig. 7. The magnitude function S_{21} has equal ripple in the pass-band and decreases monotonically in the stop-band.

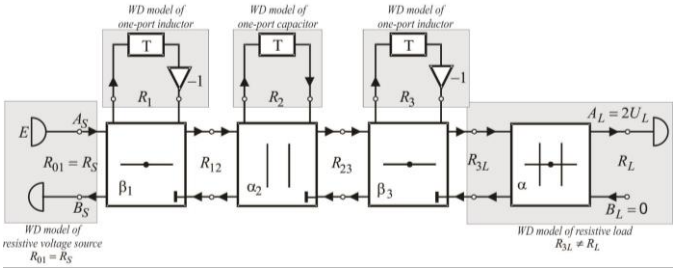


Fig. 5. *Symbolic represenation of WDF realized starting from the excitation side - Form I.*

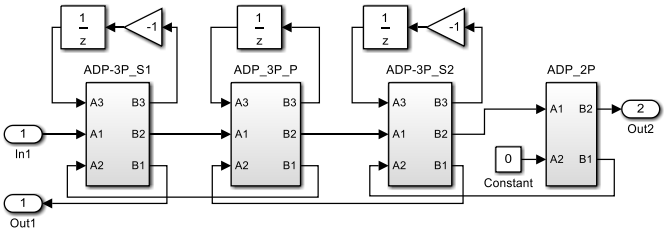


Fig. 6. *MATLAB/Simulink model of WDF structure - Form I.*

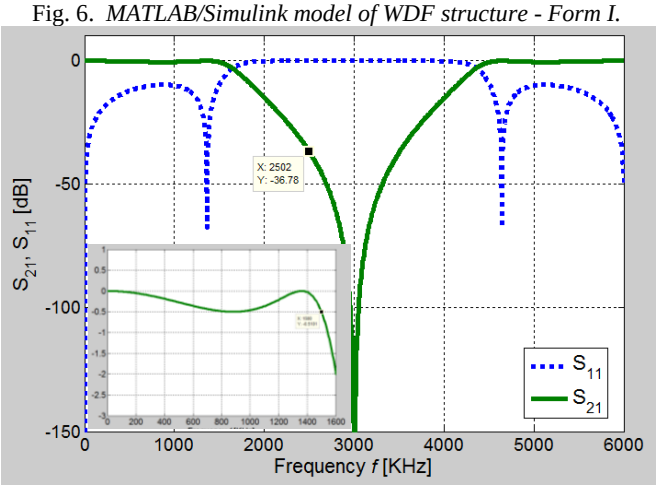


Fig. 7. *Frequency responses in dBs.*

Form II: WD model of resistive load is one multiplier shown in Fig. 4*b*. The reflection coefficient of one-port resistance network for case of $R_{3L} \neq R_L$ is $\frac{B_L}{A_L} = \frac{R_L - R_{3L}}{R_L + R_{3L}} = \beta_L$, and this results in $B_L = \beta_L \cdot A_L$ .The output is $U_L = \frac{(1+\beta_L) \cdot A_L}{2}$.

Symbolic representation of the synthesized wave digital structure for a given electrical circuit is shown in Fig. 8 and obtained by direct transformation of network from Fig. 2 by use of wave digital models described and Figs. 3*b* and 4*b*. Model has one input and one output port.

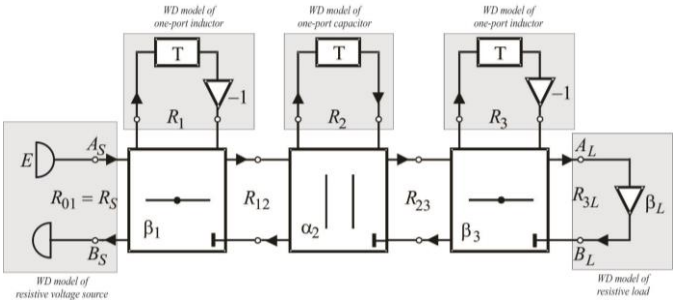


Fig. 8. *Symbolic represenation of WDF realized starting from the excitation side - Form II.*

The model for the analysis of the wave digital filter drawn in MATLAB/Simulink follows here, Fig. 9. Sub-block models of three-port series and parallel adaptors are given in [1-4]. The frequency response in dBs obtained from this model is just a reflection coefficient (S_{11} parameter), the same as the one given in Fig. 7.

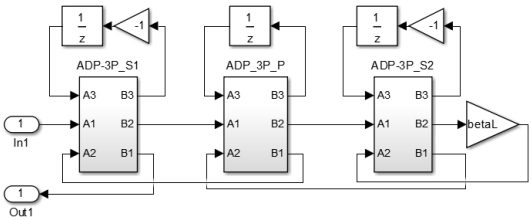


Fig. 9. *MATLAB/Simulink model for WDF structure - Form II.*

The wave digital network model is completely characterized by its parameters: port resistances and adaptor coefficients. These coefficients are explicitly found from the defined port resistances. As can be seen from [1-4], in case of series adaptors one have to sum up the port resistances, and for parallel adaptors port conductances. The calculation begins with the first unknown port resistance at the beginning of the structure, R_{12}, and goes further to the port resistance at the structure end, R_{3L} .

CASE B: In the second case study the digital network is formed **starting from the load side** and the last wave port resistance is selected as $R_{3L} = R_L$. Port resistances corresponding to inductors are set to $R_1 = L_1$ and $R_3 = L_3$. Port resistance corresponding to capacitor is set to $R_2 = 1/C_2$. For the network of real voltage source, in the case of $R_{01} \neq R_S$, and coefficient elected as $\beta_S = \frac{R_S - R_{01}}{R_S + R_{01}}$, it is obtained the relation for reflected wave as

$A_S = \beta_S \cdot (B_S - E) + E$. Model of one-port resistive load shown in Fig. 10 for case of $R_{3L} = R_L$ is the part of the network model shown in Fig. 4*a* (shaded part with excluded two-port adaptor).

Symbolic representation of the synthesized wave digital structure for a given electrical circuit is shown in Fig. 10. It is obtained by direct transformation of network from Fig. 2 by use of wave digital models previously described here.

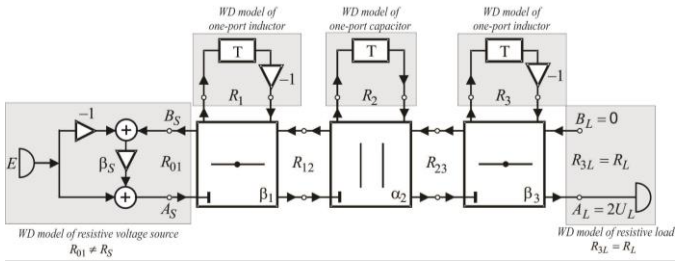


Fig. 10. *Symbolic represenation of WDF designed from the load side.*

The model for the analysis of the wave digital filter drawn in MATLAB/Simulink, Fig. 11, can be loaded into MATLAB script in order to calculate the frequeny reponse in dBs. The frequency response in dBs obtained from this model is just a transmission coefficient (S_{21} parameter), the same as the one given in Fig. 7. The port resistance calculation begins with the unkown port resistance at the structure end, R_{23}, and goes further to the port resistance at the structure beginning, R_{01}.

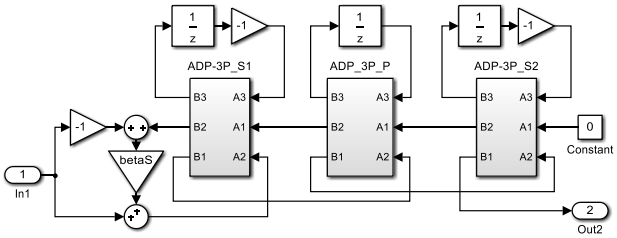


Fig. 11. *MATLAB/Simulink model of WDF designed from load side.*

III. Possibilities of Synthesized Models

This paper presents an approach for the systematic synthesis of lowpass wave digital filters for given specifications. The synthesis of WDF is realized from the designed reference circuit with elements calculated with explicit formulas for filters exhibiting equiripple characteristics. The concept presented here is examplified for one lowpass filter and different case studies of its synthesis.

Models for different case studies can be synthesized for any required specifications. Explicit formulas allows simple and easy calculation of filter order and structure elements. Synthesized models can be also used for other values of structure elements, but for the same filter order as here. Synthesis of wave digital model of structure with more or less elements is easy task with simple addition or substraction of existing blocks. The calculated frequency response are the same and do not depend of the chosen case study. Chosen case study just allows calculation of different responses.

IV. Conclusion

In this paper an essential example is presented. In this way, students and engineers can understand basic principles of wave digital filter design and its software implementation before moving to complicated hardware-based design projects and its related issues. Once students have a good foundation in the theory, they are then ready to move on and begin design projects using hardware. Examples on how to implement developed WD equivalent structures are given in MATLAB which not require any specific hardware. The reason of choosing it is its simplicity which allows one to concentrate on the WD concept rather than programming itself. Explicit formulas fot the calculation of the reference structure elements are used which makes design for any possible required specifications very simple and not time consuming. This is also important because it allows each user to select his own filter for design and analysis with desired specifications. In that manner, a lot of variances among the resulting filters depending of the specifications can be seen.

In general, the influence of CAD software-based instruction on the spatial visualisation skills of freshman engineering students in computer-aided engineering drawing course is very positive.

Acknowledgement

This paper has been supported by the Ministry for Education, Science and Technological Development of Serbia under Grant No. TR32052.

References

[1] A. Fettweis, "Digital Filter Structures Related to Classical Filter Networks", *Archiv für Elektronik und Übertragungstechnik*, vol. 25, no. 2, pp. 79-89, 1971.

[2] A. Fettweis, "Digital Circuits and Systems", *IEEE Transactions on Circuits and Systems*, vol. CAS-31, no. 1, pp. 31-48, 1984.

[3] A. Fettweis, Wave Digital Filters: Theory and Practice, *Proc. IEEE*, vol. 74, pp. 270-327, 1986.

[4] W.K. Chen, *The Circuits and Filters Handbook*, (Wave Digital Filters, 2634-2661), CRC Press, 1995.

[5] M.V. Gmitrović, *Microwave and Wave Digital Filters*, Faculty of Electronic Engineering, 2007.

[6] B. Psenicka, F.G. Ugalde and A. Romero M., "Design of the Wave Digital Filters", Chapter 9 in book: *Automation Control - Theory and Practice*, InTech, 2009.

[7] M. Yaseen, "Robust and Direct Design for Highpass Ladder Wave Digital Filters Exhibiting Equiripple Characteristics", *Digital Signal Processing*, vol. 23, pp. 1059-1064, 2013.

[8] K. Ochs and B. Stein, *On the Design and Use of Wave Digital Structures*, Paderborn University, 2001.

[9] B.P. Stošić and N.S. Dončov, "Combined Wave Digital/Full-wave Approach in Modeling and Analysis of Microstrip Structures with Examples in MATLAB/Simulink", Chapter 4 in book: *Advances in Engineering Research*, vol. 12, pp. 75-140, Nova Science Publishers, Inc., 2016.

[10] B.P. Stošić, D.I. Krstić and J.J. Joković, "Matlab/Simulink Implementation of Wave-based Models for Microstrip Structures Utilizing Short-circuited and Opened Stubs", *Electronics*, vol. 15, no. 2, pp. 31-38, 2011.

Modelling Framework for Performance Analysis of Circular Runway Airport Traffic: A Petri Net Approach

Ilija Hristoski[1] and Tome Dimovski[2]

Abstract – **In this paper, a modelling framework for performance analysis of a circular airport runway, proposed within the Dutch project 'the Endless Runway', is given. The proposed solution takes into account the basic project's parameters and air traffic policies and is based on the usage of the class of Generalized Stochastic Petri Nets (GSPNs).**

Keywords – **Modelling, Generalized Stochastic Petri Nets, Performance analysis, Aircraft traffic, Circular runway**.

I. INTRODUCTION

As the global economy is becoming more and more connected, passenger air travel is expected to maintain positive growth rates up to 2030, despite a number of challenges faced by the aviation industry, like the sluggish economic growth worldwide and the high jet fuel prices. It is believed that between 2017 and 2036, the number of airline passengers is expected to grow at a compound annual growth rate (CAGR) of 4.7%. More precisely, in 2017, global air traffic passenger demand increased by 7.5% on the year before, and by the end of 2018, traffic is projected to grow with another 6% [1].

Such intense dynamics in the demand poses quite new challenges vis-à-vis the operational performances of all affected players in the aviation sector, including aircraft manufacturers, aircraft operators, travel agencies, and especially the existing and future airports, which have to cope successfully with the ever-increasing number of flights.

II. THE CONCEPT OF A CIRCULAR RUNWAY

The concept of the circular or endless runway is not an entirely new idea. It has been explored since the early days of aviation. In France, the first circular take-off took place at the end of the 19th century. US Navy has launched such project back in the 1960s, and consequently, many successful landings and take-offs with propeller and jet planes were made between 1964 and 1965 [2]. However, commercial circular runways have never been built, nor tests with passenger aircraft have been conducted so far.

Recently, Dutch researchers have revived the concept of a circular runway, this time for civil airports, based on the idea of Henk Hesselink, senior R&D engineer in the Netherlands Aerospace Centre (NLR). Researchers at "the Endless Runway", a project funded by European Commission, believe circular runways could have several benefits, including being more environmentally friendly and less noisy [3]. The group proposes constructing a 3.5 km wide circular runway with banked sides divided into 18 runway segments, with an airport hub situated in the center of the circle, along with four terminals (A, B, C, and D) with total capacities of 81, 81, 66, 66 planes, respectively. The length of the circular runway would be equal to three straight runways while being able to handle the air traffic of four. Circular runways will allow planes to land and take off at any point in the circle. Pilots will be able to land in directions with the most favorable weather conditions while avoiding difficult maneuvers and situations such as strong crosswinds. The circular design will also mean aircraft coming into land circle above the airport fewer times, thus using less fuel. The design allows for three planes to land and take-off at the same time. Because of the centripetal forces, aircraft will automatically go slower and move towards the center of the runway. Circular runways could also limit noise pollution by spreading it more evenly around the airport. Besides, they take up a third of the space of typical airports, making them better for both the environment and travelers. It is also noteworthy to mention that the Netherlands Aerospace Centre aims to make it a reality by 2050, which could revolutionize air travel.

III. GENERALIZED STOCHASTIC PETRI NETS

Generalized Stochastic Petri Nets (GSPNs) [4-5] are recognized as a widely-known tool for performance analysis of distributed systems, which utilizes the graphical notation introduced by ordinary Petri Nets (PNs). In GSPNs some transitions are timed, whilst others are immediate. Random, exponentially distributed firing delays are associated with timed transitions, whereas the firing of immediate transitions takes place in zero time, with priority over timed transitions. In addition, the selection among several possibly conflicting enabled immediate transitions is made by utilizing their corresponding firing probabilities. In general, immediate transitions are used for modelling instantaneous actions or logical actions (typically choices), whilst timed transitions with an exponentially distributed delays are used for modelling the duration of activities (events) within the GSPN model.

The analysis of a GSPN model can be two-fold: (1) qualitative: performed by studying the structural characteristics of the underlying Petri Net; (2) quantitative: performed by computing the steady-state (stationary) and/or the transient (time-dependent) probability distributions of the associated stochastic model (process), equivalent to a GSPN

[1]Ilija Hristoski is with the Faculty of Economics at "St. Kliment Ohridski" University in Bitola, 133 Marksova St, Prilep 7500, Republic of Macedonia, E-mail: ilija.hristoski@uklo.edu.mk.

[2]Tome Dimovski is with the Faculty of Information and Communication Technologies at "St. Kliment Ohridski" University in Bitola, Partizanska St, Bitola 7000, Republic of Macedonia.

model. GSPNs are isomorphic to semi-Markov processes, i.e. their quantitative analysis can be performed on a reduced Embedded CTMC (Embedded Markov Chain, EMC), defined solely on a set of tangible states, or by reducing the GSPN to an equivalent Stochastic Petri Net (SPN) [5]. The stationary distribution of the underlying stochastic process is usually a basis for obtaining a plethora of performance metrics, like calculating the probabilities of specific state conditions, resource utilization, expected throughputs, expected number of clients (active resources), expected waiting times, etc. On the other hand, transient analysis is a basis for investigating the system behaviour over time, i.e. it describes the evolution of the observed system at a given time and thus it can be used for obtaining specific performance metrics such as probabilities of reaching particular states and probabilities of satisfying assigned deadlines [5].

IV. The Proposed GSPN Simulation Model

The concept of the endless runway has been already evaluated using simulations. Three different areas have been identified to be evaluated, including the runway itself, the surrounding terminal maneuvering area (TMA), and the ground movement area (GMA). For the runway, a special simulation tool was set up by Office National d'Études et de Recherches Aérospatiales (ONERA) to optimize the usage. The TMA was simulated with TrafficSim, a Deutsches Zentrum für Luft- und Raumfahrt (DLR) in-house solution, whilst the GMA was implemented and evaluated in Simmod Pro!. All three areas have been evaluated separately by the used simulation tools. Different parameters like delays or capacities have been used to get the first glimpse into the feasibility of the concept [6].

The proposed GSPN-based simulation model is not aimed to replace the existing specific in-depth simulation models that deal strictly with the technical aspects of the implementation and the feasibility of the concept. It is rather intended to serve as a framework for carrying out performance analysis of the aircraft traffic dynamics, taking into account the known input parameters, including aircraft arrival rate (λ), aircraft departure rate (μ), total capacities of the particular terminals (C1, …, C4), aircraft mean landing delays ($1/\rho$), aircraft taxiing delays from the eighteen runway entry points towards particular terminals ($1/\varphi$) and vice-versa ($1/\beta$), aircraft mean waiting times at particular terminals ($1/\varepsilon$), aircraft mean waiting times to take-off ($1/\omega$), aircraft mean take-off duration ($1/\theta$), mean duration of aircraft movements at ground level ($1/\delta$), etc. Our approach is based on the assumption that the circular runway airport can be viewed as a complex Discrete-Event Dynamic System (DEDS), where activities belong to five phases: (1) Aircraft arrivals and landings; (2) Aircraft taxiing from landing points to airport terminals; (3) Aircraft operations at airport terminals; (4) Aircraft taxiing from airport terminals to departure points; (5) Aircraft take-offs and departures. In addition, since such DEDS is characterized by discrete (countable) state space and events in the presence of concurrency, cooperation, synchronization, blocking etc. vis-à-vis queuing, servicing, and routing of aircraft, a convenient formalism for their representation and performance evaluation is the class of GSPNs. Due to the complexity of the proposed solution, the GSPN model is divided into five sub-models, one per phase.

Fig. 1 depicts the activities related to phase #1. Aircraft arrivals follow the Poisson distribution with an arrival rate of λ (transition *T_arrival*). When ready to land (a token in the place *P_ready_to_land*), the arriving plane has to choose a single entry point *Ri* out of 18 possible ones (transitions *T_choose_entry_Ri*, $i = 1, \ldots, 18$). However, since maximum three planes are allowed to take-off or land simultaneously (place *P_max_planes*), a plane will be allowed to start landing (place *P_start_landing_Ri*) only if the chosen entry point *Ri* is clear (a token in the place *P_Ri_clear*) and there are no more than three planes in the air (the place *P_max_planes* is non-empty). Otherwise, the transition *T_stay_in_air* will be enabled, and the plane will have to stay in the air (transition *T_circling*) until the landing conditions are met. After the landing (firing of a transition *T_landing_Ri*), a single token is put back into the places *P_max_planes* and *P_Ri_clear*, for the chosen *i*, i.e. *Ri*. The durations of landings and the circling of planes in the air are assumed to be exponentially distributed with means of $1/\rho$ and $1/\nu$, respectively.

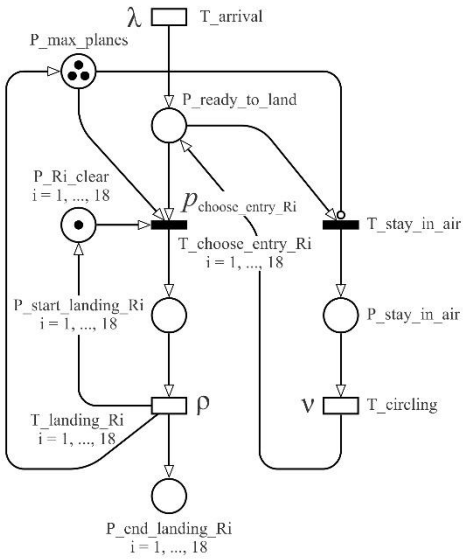


Fig. 1. GSPN sub-model of aircraft arrivals and landings (phase #1)

After landing at a chosen entry point *Ri* ($i = 1, \ldots, 18$), denoted by a token in the place *P_end_landing_Ri*, a plane starts taxiing to a chosen airport terminal *Tj* ($j = 1, \ldots, 4$), as depicted in Fig. 2 (phase #2). This would be possible only if the capacity of the chosen terminal, *Cj* ($j = 1, \ldots, 4$), is not exhausted (i.e. the place *P_Tj_capacity* is not empty). The taxiing to the chosen terminal *Tj* is assumed to take an exponentially distributed time with a mean of $1/\varphi$. The firing of the transition *T_taxiing_to_Tj* for the chosen *j* puts a single token into places *P_start_waiting_Tj* and *P_move_from_Tj*: a token in the first one denotes that the plain will reside at the chosen terminal *Tj* for an exponentially distributed time with a mean of $1/\varepsilon$ (transition *T_waiting_Tj*), whilst a token in the second one will allow a plane to move from the current terminal to another one during the next phase.

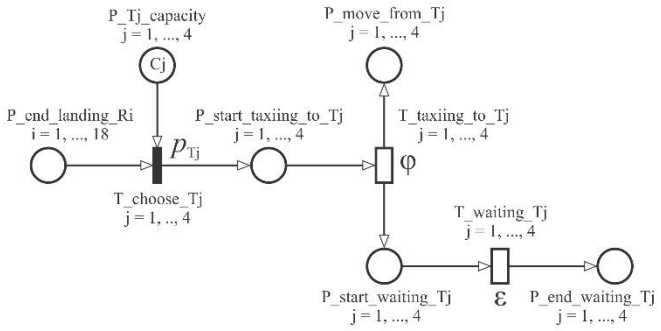


Fig. 2. GSPN sub-model of aircraft taxiing to airport terminals (phase #2)

Phase #3 describes aircraft activities at ground level. Fig. 3 describes the GSPN segment explicitly related to the terminal T4. Corresponding logic has to be applied for terminals T1, T2, and T3. After residing at the terminal T4 for a certain amount of time (a token in the place *P_end_waiting_T4*), each plane can move either to another terminal (transitions *T_move_T4_T1*, *T_move_T4_T2*, and *T_move_T4_T3*), or stay at the current one (transition *T_move_T4_T4*). Each moving from T4 to other terminals takes away a single token from places *P_move_from_T4* and *P_Tk_capacity* (*k* = 1, 2, 3). Choosing to stay at T4 takes away a single token solely from the place *P_move_from_T4*. Moving from a given terminal to another one takes some time, exponentially distributed with a mean of 1/δ (transitions *T_move_T4_to_Tk*, *k* = 1, 2, 3). The firing of any of these transitions puts a single token back into the place *P_T4_capacity*, as well as into the place *P_start_departure_Tk*, for a chosen *k* (*k* = 1, 2, 3).

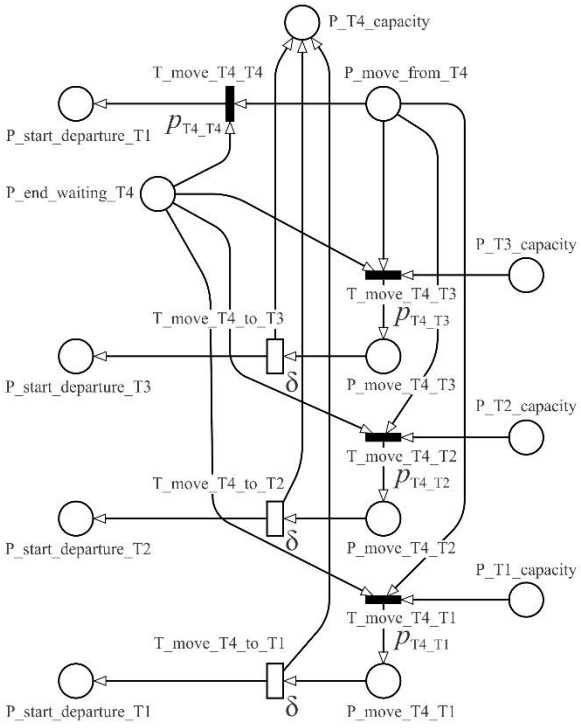


Fig. 3. GSPN sub-model of aircraft operations at airport terminals, for the airport terminal T4 (phase #3)

During phase #4, planes departure from any of the four terminals (a token in the places *P_start_departure_Tj*, *j* = 1, …, 4). The exponential transition *T_departure_Tj* (*j* = 1, …, 4) fires with a departure rate of μ (Fig. 4). First, for any given terminal *Tj* (*j* = 1, …, 4), departing planes choose an exit point *Ri* (*i* = 1, …, 18), by firing the corresponding immediate transition *T_choose_exit_Ri* with a probability of p_{choose_Ri}. Next, the plane starts taxiing to the chosen exit point (place *P_start_taxiing_to_Ri*, *i* = 1, …, 18), an activity that lasts, on average, 1/β. Because more than one plain can consequently choose a particular exit point *Ri*, they will be represented by an equivalent number of tokens in the place *P_wait_for_take-off_Ri*. Besides, each firing of the exponential transition *T_taxiing_to_Ri* for a chosen exit point *Ri* puts a token back into the place *P_Tj_capacity* for a given terminal *Tj* (*j* = 1, …, 4), meaning that a plane has left the terminal and waits in a queue to take-off at the circular runway.

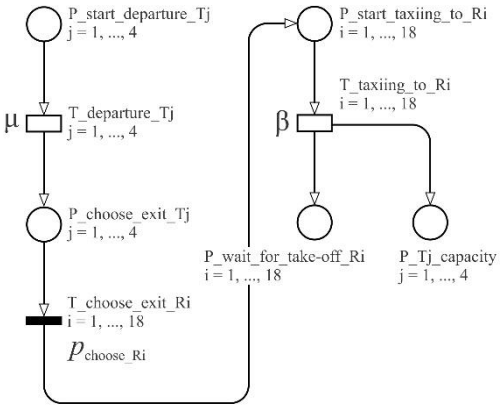


Fig. 4. GSPN sub-model of aircraft taxiing from airport terminals (phase #4)

Planes wait for a take-off at an exit point *Ri* (a token in the place *P_wait_for_take-off_Ri*) an arbitrary time that is exponentially distributed with a mean of 1/ω, until the chosen exit point is clear (a token in the place *P_Ri_clear*), as shown in Fig. 5 (phase #5). As soon as the last condition is met, the take-off from the exit point *Ri* starts (a token in the place *P_start_take-off_Ri*, *i* = 1, …, 18), but only if the number of tokens in the place *P_max_planes* is non-zero (i.e. if there are at most three planes landing or taking-off on the circular runway at the moment). The duration of the take-offs is exponentially distributed with a mean of 1/θ (transition *T_take-off_Ri*, *i* = 1, …, 18). The firing of this transition puts a single token back to places *P_max_planes* (to denote that the plane has flown away) and *P_Ri_clear* (to denote that the entry/exit point *Ri* is clear for future landings/take-offs), for a particular terminal *Tj* (*j* = 1, …, 4) and chosen exit point *Ri* (*i* = 1, …, 18). It also puts a single token into the place *P_end_departure_Ri*, for the chosen *Ri*, to denote that a departure of a plane has occurred at the exit point *Ri*. In order to avoid cumulating tokens in the place *P_end_departure_Ri*, it is connected to an immediate transition *T_departure* with an arc having a multiplicity of #*P_end_departure_Ri*. In such a way, tokens are immediately removed from the places *P_end_departure_Ri*, *i* = 1, …, 18.

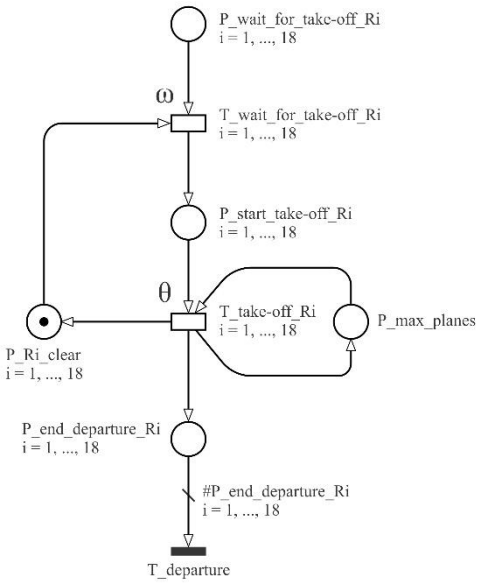


Fig. 5. GSPN sub-model of aircraft take-offs and departures (phase #5)

V. CONCLUSION

The operation of commercial airports with circular runways in terms of aircraft traffic dynamics is highly complex. The existence of multiple stochastic processes justifies its treatment as a DEDS, characterized by discrete (countable) state-space and a number of events, each lasting for a random time. The complexity found among various DEDS components suggests considering the evolution of such system as a stochastic process that can be used to assess its performance. Since stochastic processes can be successfully captured and described by the wide gamut of stochastic Petri nets, the class of GSPNs has been chosen as a modelling formalism, mainly for several reasons: (1) Arrivals and departures of aircraft follow the Poisson distribution, where inter-arrival and inter-departure times are exponentially distributed with parameters λ and μ, respectively. As such, the underlying stochastic process is a Continuous Time Markov Chain (CTMC), which is also an underlying stochastic process of GSPNs; (2) GSPNs are often used for modelling and evaluation of transport and traffic systems; (3) In order to keep the model structure as simple as possible, the durations of all events are supposed to be random times, exponentially distributed, i.e. the times between events conform the Poisson process where events occur continuously and independently at constant average rates; (4) The methodology for an analytical solution of GSPNs is well known and documented; (5) There are multiple dedicated software packages today, like TimeNET, GreatSPN, or WebSPN, that offer both modelling and numeric simulation/evaluation of GSPNs.

The proposed GSPN-based modelling framework is quite complex, and that was the reason for its partitioning into sub-models by particular phases. Each GSPN sub-model, corresponding to a particular phase, can be analyzed either as a stand-alone part or in conjunction with other sub-models. The proposed solution can be successfully utilized for obtaining numerous performance measures vis-à-vis the circular runway airport traffic, including the average number of planes waiting at the terminals, the average number of planes at the airport, the average sojourn time of planes at terminals, the average sojourn time of planes waiting for take-off, the utilization of the airport, etc. All of these can be evaluated against different values of the arrival (λ) and departure (μ) rates. Besides the performance evaluation, it can be successfully utilized for addressing additional critical issues related to the circular runway airport, such as correctness analysis, reliability evaluation, design optimization, scheduling (performance control), monitoring & supervision, traffic efficiency, implementation, system tuning, bottleneck identification, workload characterization, capacity planning, forecasting the performance at future loads, evaluation of airport design alternatives, etc.

Validation, as a process of checking whether the specification of the proposed solution captures the actual customer's needs, is an extremely subjective process. As per verification, all the activities vis-à-vis the production of a high quality performance evaluation model (testing, inspection, analysis etc.) can be carried out by using dedicated software.

The limitation of the proposed modelling framework is that the GSPN-based model does not make any difference between two major classes of aircraft regarding their size, as being originally anticipated within the project. This could be accomplished by utilizing the class of Colored Petri Nets (CPNs). Yet another limitation of the proposed solution is that, structurally, it does not take into account particular stands at any of the four terminals (294 in total). The inclusion of such information will significantly improve the accuracy of the simulation model, but it will also make its structure extremely complex and, possibly, computationally intractable due to the state-space explosion.

REFERENCES

[1] Statista.com, "Annual growth in global air traffic passenger demand from 2005 to 2018", URL: https://www.statista.com/statistics/193533/growth-of-global-air-traffic-passenger-demand/ (Accessed March 5th, 2018)

[2] E. Grey, "Endless Runway concept could pave the way for future airports", URL: https://www.airport-technology.com/features/featureendless-runway-concept-could-pave-the-way-for-future-airports-5866253/ (Accessed March 7th, 2018)

[3] H. Hesselink, "the Endless Runway: D5.4 Final Report", 2014, URL: https://cordis.europa.eu/docs/results/308292/final1-d5-4-the-endless-runway-final-report.pdf (Access. March 14th, 2018)

[4] M. Ajmone Marsan, G. Balbo, G. Conte, S. Donatelli and G. Franceschinis, *Modelling with Generalized Stochastic Petri Nets*, Wiley, 1995.

[5] G. Balbo, "Introduction to Generalized Stochastic Petri Nets", In: M. Bernardo and J. Hillston (Eds.), *Formal Methods for Performance Evaluation*, SFM 2007, Lecture Notes in Computer Science, vol. 4486, Springer, Berlin, Heidelberg, pp. 83-131, 2007.

[6] S. Loth, H. Hesselink, R. Verbeek, M. Dupeyrat and S. Aubry, "the Endless Runway: D4.3 Simulation - Modelling and Analysis", 2014, URL: http://www.endlessrunway-project.eu/downloads/d4.3-simulation-modelling-and-analyses-v2.pdf (Accessed March 14th, 2018)

Zero-Suppressed BDD and collaborative filtering

Katarina Jovanović[1] and Milena Stanković[2]

Abstract – **Techniques that enable efficient data sets manipulation are very important in recommendation systems. In this paper, we present a method for creating recommendation using collaborative filtering and Zero-suppressed binary decision diagram (Zero-suppressed BDD). The binary decision diagrams (BDDs) are used to analyze problems that occur with large data sets. One of BDD is ZBDD which is suitable for presenting user-product matrix that preserves the history of user's behavior in the collaborative filtering system. Using diagram traversal, we allocate similar users and based on the product which users rated, we create recommendations for a certain user. This way of creating recommendations can be effective with systems that work with large number of users and products.**

Keywords – **Collaborative filtering, Zero-suppressed BDD, Tuple histogram, Pattern histogram.**

I. Introduction

Collaborative filtering system (CF) is one of the ways for the recommendation of products that is based on users' interests and preferences, and selects automatically the products that might be of interest to people. The goal of these systems is to assist in the search and selection of products. The systems for recommendation are based on the assumption that users who have agreed in the choice of products in the past, they will agree in choice of products in the future and they will like a similar kind of products. These users represent a group of similar users.

CF collects feedback from users on the basis of their rates of the products and determines that the users are interested in some products based on their behavior (review, purchase, commenting on the product). History of users' behavior represents the matrix that is called user - product matrix.

Collaborative filtering is widely applied in electronic commerce, where customers can rate and buy different products. On the basis of users' activities, CF generates recommendations that include products of interest. Application of this technique is found in the creation of social networks for recommendation of new friends, groups and pages.

As already known, ZBDD is used to analyze and solve problems that arise in large databases. This data structure manipulates datasets simpler and more efficient than the original BDD (Binary Decision Diagram) and also provides a unique and compact set representation. This paper will present the process of creating recommendations using ZBDD.

[1]Katarina Jovanović is with the Faculty of Electronic Engineering at University of Niš, Aleksandra Medvedeva 14, Niš 1800, Serbia, E-mail: katarina_jovanovic@outlook.com.

[2]Milena Stanković is with the Faculty of Electronic Engineering at University of Niš, Aleksandra Medvedeva 14, Niš 1800, Serbia.

II. Zero-Suppressed BDD

BDDs have been developed to handle Boolean functions, however, they can also be used to represent sets of combinations. The term "sets of combinations" represents a set of elements, where the element is a combination of n items. Examples of such data model are combinations of switching devices (ON/OFF), fault combinations, and sets of paths in the networks [7].

A combination of n items can be represented by an n-bit binary vector, $(x_1 x_2 x_3 \dots x_n)$ where each bit, $x_k \epsilon \{0,1\}$ expresses whether or not the item is included in the combination. ZBDD is a special type of BDDs which manipulates efficiently with combination sets and it is based on the special reduction rules.

- Delete all nodes whose 1-edge directly points to the 0-terminal node, and jump through to the 0-edge's destination, as shown in Fig 1.
- Share equivalent nodes as well as ordinary BDDs.
- Not to delete the nodes whose two edges point to the same node, which used to be deleted by the original rule [7].

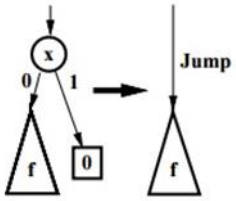


Fig. 1. ZBDD reduction rule

The features of ZBDD.

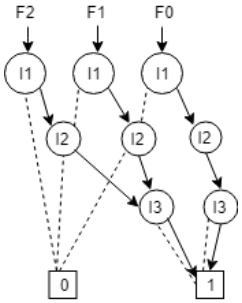
- In ZBDD, the nodes of irrelevant items (those that are not part of any combination) are automatically deleted by ZBDD reduction rules.
- ZBDD is especially effective for representing sparse combinations
- Each path from the root node to 1-terminal node corresponds to each combination in the set. In fact, the number of such paths in ZBDD corresponds to the number of combinations in the set.
- ZBDD structure explicitly stores all items in all combinations, as well as using an explicit linear linked list data structure. [7].

Tuple or transaction is a record that contains a combination of items. The table that contains the number of appearance of each tuple in the given database is tuple histogram.

In the example of Fig. 2 we show a method of representing tuple histogram by using ZBDD. The numbers of tuple's

appearances we decompose into n-digits of ZBDD vector $\{F_0, F_1, \ldots, F_{n-1}\}$, actually we encode the appearance numbers into binary digital code as shown in Fig. 2. F_0 represents a set of tuples appearing odd times, least significant bit, while F_1 represents set of tuples whose second lowest bit of number of appearances is 1. In similar way we define the set of each digit up to F_{n-1}. In the example the tuple frequencies are decomposed as: $F_0 = \{I_1I_2, I_1I_2I_3, I_2I_3\}$, $F_1 = \{I_1I_2, I_2\}$, $F_2 = \{I_1I_2I_3\}$ and each digit can be represented by ZBDD vector. ZBDD vector in Fig. 2 is constructed base on presented tuple-histogram. Also, ZBDDs can share their sub-graphs with each other.

tuple	frequency	F_2	F_1	F_0
I_1I_2	3 (011)	0	1	1
$I_1I_2I_3$	5 (101)	1	0	1
I_2	2 (010)	0	1	0
I_2I_3	1 (001)	0	0	1



$F_0 = \{I_1I_2, I_1I_2I_3, I_2I_3\}$

$F_1 = \{I_1I_2, I_2\}$

$F_2 = \{I_1I_2I_3\}$

Fig. 2. ZBDD vector for tuple histogram

The procedure of generating tuple-histogram F_T for a given database D is shown with pseudocode [7].

$$
\begin{aligned}
&F_T = 0 \\
&forall\ T\ \in D\ do \\
&\quad F_T = F_T.add(T) \\
&return\ F_T
\end{aligned}
\qquad (1)
$$

F_T is tuple-histogram for given database D, $F.add(T)$ adds combination (tuple) T in the *ZBDD* vector F.
After creating a histogram, it is easy to single out the tuples that appear more than α times. This procedure will be used for creating ZBDD vector. The number of ZBDD nodes in each digit is bounded by total appearance of items in all tuples.

A subset of items included in the tuple represents a pattern. A pattern-histogram is a table that contains the number of appearance of each pattern in any tuple in the given database. In fact, a tuple of k items includes 2^k patterns.
The procedure of generating pattern-histogram based on ZBDD is given by pseudocode [7].

$$
\begin{aligned}
&F_p = 0 \\
&forall\ T \in D\ do \\
&\quad P = T \\
&\quad forall\ v \in T\ do \\
&\quad\quad P = P\ \cup P.onset(v) \\
&\quad F_p = F_p.add(P) \\
&return\ F_p
\end{aligned}
\qquad (2)
$$

Procedures of generating tuple- histogram and pattern-histogram have been developed by Shin-ichi Minato and his team.

III. ZBDD FOR COLLABORATIVE FILTERING

Recommender systems, within the web sites, keep the history (behavior) for each user in a user-product matrix. User's history can be written as a combination of products that are rated by the user. How the binary decision diagrams are used to represent the combination, the combination of n products can be represented by n-bit binary vector $(x_1x_2x_3 \ldots x_n)$, where each bit, $x_k \epsilon \{0,1\}$, indicates whether the product is the part of the combination or not.

Table I represents the history of the fifteen users and their interest in the products (user-product matrix). Numbers greater than zero indicates that users rated the products while zero indicates that the products are not rated by users.

TABLE I
USER – PRODUCT MATRIX

Product / User	I_1	I_2	I_3	I_4	I_5	I_6	I_7
U_1	5	0	4	0	3	0	2
U_2	0	2	0	3	0	5	0
U_3	4	5	5	0	0	2	3
U_4	0	2	2	0	2	2	0
U_5	5	0	4	0	0	3	0
U_6	2	0	1	0	0	0	0
U_7	1	2	0	0	0	0	0
U_8	0	0	3	3	3	4	5
U_9	5	5	0	0	4	3	2
U_{10}	0	2	2	0	0	3	3
U_{11}	4	0	4	0	5	0	0
U_{12}	0	0	1	5	0	0	0
U_{13}	0	5	0	0	0	4	1
U_{14}	0	3	3	0	0	4	4
U_{15}	5	0	5	0	0	0	0

If we look at the user U_2, from Table I, it can be seen that the user rated the products I_2, I_4 and I_6, while the user U_5 rated I_1, I_3 and I_6 products.

The user for whom system creates a recommendation is an active user and in our case it is U_{13} user. Before creating recommendations, it is necessary to form a group of similar users that are similar to the active user in terms of interests in products. To find similar users, we take advantage of tuple-histogram and operations that can be performed on them. ZBDD vector for tuple histogram is represented in Table II. Each combination represents one path in the ZBDD diagram. Therefore, finding similar users is based on the tour of diagram and extracts paths that match with the paths that are appropriate to the active user in k or more segments.

TABLE II
ZBDD VECTOR FOR TUPLE HISTOGRAM

Tuple	Frequency	F_1	F_0
$I_1I_3I_5I_7$	1(01)	0	1
$I_2I_4I_6$	1(01)	0	1
$I_1I_2I_3I_6I_7$	1(01)	0	1
$I_2I_3I_5I_6$	1(01)	0	1
$I_1I_3I_6$	1(01)	0	1
I_1I_3	2(10)	1	0
I_1I_2	1(01)	0	1
$I_3I_4I_5I_6I_7$	1(01)	0	1
$I_1I_2I_5I_6I_7$	1(01)	0	1
$I_2I_3I_6I_7$	2(01)	1	0
$I_1I_3I_5$	1(01)	0	1
I_3I_4	1(01)	0	1
$I_2I_6I_7$	1(01)	0	1

$$F_0 = \begin{Bmatrix} I_1I_3I_5I_7, I_2I_4I_6, I_1I_2I_3I_6I_7, I_2I_3I_5I_6, I_1I_3I_6, I_1I_2, \\ I_3I_4I_5I_6I_7, I_1I_2I_5I_6I_7, I_1I_3I_5, I_3I_4, I_2I_6I_7 \end{Bmatrix}$$
$$F_1 = \{I_1I_3, I_2I_3I_6I_7\}$$

Operation, which will be used in determining similar users, is finding all combinations that contain given pattern.

$$\begin{aligned} &S = \cup F_k \\ &forall\ v \in P\ do: \\ &\ S = S.onset(v).change(v) \\ &return\ S \end{aligned} \quad (3)$$

S is union of tuples which contain pattern *P*, *S.onset(v)* selects the subset of combinations including item *v* and *S.change(v)* inverts existence of *v* (add / delete) on each combination (see table with primitive ZBDD operations [7]).

Patterns with one element that match the active user's combination is $P \rightarrow \{I_2, I_6, I_7\}$. *ZBDDsimilarUsers* algoritham finds all combinations which contain patterns from the set *P*. Combinations that contain α or more patterns from set *P* belong to group of similar users. The *ZBDDsimilarUsers* algorithm is shown in Fig 3.

```
ZBDDsimilarUsers(F)
{
    S = ∪ F_k
    forall v ∈ P do:
      S = S.onset(v).change(v)
    forall T ∈ S
      countT = 0
      for i = 1 to S.count
        if(T = S[i])
          countT = countT + 1
      if(countT ≥ α)
        S_u = S_u U T
    return S_u
}
```

Fig. 3. Find similar users algorithm

All combinations are stored in a collection *S*. The parameter α depends on the recommendation system and represents number that indicates that two users are similar if they rated α or more the same products.

If the combination *T* from set *S* appears α or more times, it means that the active user and the user to whom refers this combinations are similar. All combinations that satisfy the requirement are placed in a collection S_u which contains all similar users.

In order to determine the product group for the recommendation, it is necessary for each combination from the set S_u to extract patterns that contain one element. If the pattern is part of active user's combinations, it will not be included in the group for recommendation while other patterns will be included in recommendation. A set of products for a recommendation is obtained by separating nodes that are not the part of active user's path in ZBDD. If *Q* represents patterns of similar users, *P* represents patterns of the active users and *R* represents products for recommendation, procedure to extract the products is:

$$R = Q - Q \cap P \quad (4)$$

After determining group of products for recommendation it is needed to define top *N* products for recommendation and order of recommendation. By computing the average rate of products from the group and sorting them by the average rates, we obtain the order of the products for the recommendation. The average product's rate represents weight of the branch in ZBDD vector.

IV. CONCLUSION

In this paper we presented a method for creating recommendations using collaborative filtering and ZBDD. This way of creating recommendation provides simpler manipulation with data sets. Here are given procedures for finding similar users and creating group of products for recommendation. Finding similar users is based on a touring of ZBDD diagram and extracting sub-diagram which represents similar users. The nodes in the ZBDD sub-diagram, which are not part of the active user's path, are products for recommendation. Sorting the products by rates we determine the order in which products will be recommended to active user.

REFERENCES

[1] D. Asanov, "Algorithms and Methods in Recommender System", Berlin Institute of Technology, Berlin, Germany.
[2] M. D. Ekstran, J. T. Riedl. and J. A. Konstan, "Collaborative Filtering Recommender Systems", Foundations and Trends® in Human–Computer Interaction: vol. 4, no. 2, pp. 81–173, 2011.

[3] M. Jones, "*Recommender systems"*, http://www.ibm.com/developerworks/library/os-recommender1/, accessed on 01.05.2017.

[4] J. Lee., M. Sun, G. Lebanon, "A Comparative Study of Collaborative Filtering Algorithm", Workshops, Demos, and ArXiv Preprints, 2012

[5] P. Melville and V. Sindhwani, "Recommender system", In: C. Sammut and G. Webb, Eds., Encyclopedia of Machine Learning, Springer, Berlin, 2010, pp. 829-838.

[6] S. Minato and H. Arimura , "Generating Frequent Closed Item Sets Based on Zero-suppressed BDDs", Hokkaido University, Division of Computer Science, TCS Technical Reports, TCS-TR-A-06-17, Jul. 2006

[7] S. Minato and H. Arimura, "Combinatorial Item Set Analysis Based on Zero-Suppressed BDDs", Hokkaido University, Division of Computer Science, TCS Technical Reports, TCS-TR-A-04-1, Dec, 2004.

[8] X. Su and T. M. Khoshgoftaar, "A Survey of Collaborative Filtering Techniques", *Advances in Artificial Intelligence*, vol. 2009, Article ID 421425, 19 pages, 2009.

[9] B. Sarwar, G. Karypis, J. Konstan, and J. Reidl, "Item-based collaborative filtering recommendation algorithm", In Proceedings of the 10th International Conference on World Wide Web (pp. 285-295). (WWW '01). New York, NY, USA: ACM.

ENERGY SYSTEMS AND EFFICIENCY, AND MEASUREMENT SCIENCE AND TECHNOLOGY – ORAL SESSION

Standardized design solutions for typical and adapted distribution substations 10(20)/0.4 kV

Mile Spirovski[1], Nikolce Acevski[1], Blagoja Arapinoski[1] and Igor Sterjovski[1]

Abstract – **In this paper will be shown the whole process of designing type and adapted distribution substations 10(20)0.4 kV using the standardized solutions. In addition, will be given calculations required for the dimensioning of the medium voltage and low voltage equipment, grounding, such as the calculations of the short-circuit currents on the MV and LV side, depended the current load on the MV and LV side, the generated heat at the maximum load of the transformer and the value of earth resistance. All calculations will be performed for standardized types of equipment, depending on the power of the transformer.**

Keywords – **distribution substations, designing, standardized solutions.**

I. INTRODUCTION

The subject of this paper is part of the technical solution for power supply to an appropriate facility from an adapted substation. For connection to the new substation of the power distribution network is planned new dual cable which is included input-output connection for up to two injection cells in the MV switchgear of the substation, which is not the subject of this paper.

According to the approved peak power of 656 kW, given in the Decision for the consent for connection (issued by the Distribution System Operator - ODS in the Republic of Macedonia - EVN Macedonia), a transformer with a power of 800 kVA is chosen, and the measurement of the electricity should be placed on medium voltage side.

Technical data for substation: Rated operating voltage 10 (20) kV, Maximum operating voltage 24 kV, Operating voltage (Ue) on the low side 0,4 kV, Rated frequency 50Hz, Level of insulation on the medium side 50 kV, Impulse voltage that can be maintained (U 1.2/50 μs) on medium side 125 kV, Impulse voltage that can be maintained (U 1.2/50 μs) on low side 8 kV, Rated current of circuit breaker (line feeder) 630 A, Rated current of circuit breaker (transformer feeder) 200 A, Rated current of circuit breaker between MV switchgear and LV switchboard 1250 A, Thermal current which can be maintained for a short time on LV side 25 kA, Dynamic current which can be maintained for a short time on LV side 40 kA, Power of transformer 800 kVA.

[1]Mile Spirovski, [1]Nikolce Acevski, [1]Blagoja Arapinoski and [1]Igor Sterjovski are with the Faculty of Technical sciences at University of Bitola, st. Makedonska Falanga 33, Bitola 7000, Macedonia.

II. DISPOSITION OF EQUIPMENT

The facility consists of two sectors: a medium voltage switchgear, a low voltage switchboard and transformer box.

The transformer box is placed in the right part of the facility and in this part an oil transformer with porcelain insulators 10.5 (21) /0.42 kV, 800 kVA will be installed. The substation in the transformer box has an oil pit to collect the oil that may expire in the event of a major accident.

The middle voltage unit will be placed in the sector for the MV and LV plants at the far left side of the building.

The LV switchboard will be placed on the barrier wall bordering the transformer box.

III. MEDIUM - VOLTAGE SWITCHGEAR

The middle voltage facility will be equipped with a type 20 kV switchgear consisting of three supply input-output terminals, one coupling, one measuring and one transformer cell.

In the metallic cabinets of the injection cells, a circuit breaker with a grounding device is installed, the measuring current transformers and measuring voltage transformers will be placed in the measuring cell, a circuit breaker is placed in the coupling cell, in the transformer cell the circuit braker is combined with fuses, and the switchgear is filled with gas SF6 under pressure.

IV. POWER TRANSFORMER

In the substation, a transformer will be built with the following characteristics:

TI-ERM 630 kVA: power 800 kVA, transformation ratio 3 x 10,5(21)/0,42 kV, connection Dyn5, frequency 50 Hz, transformer type oil, no conservation, cooling mode ONAN, no load losses 1400 W, load losses at 75º 8700 kW, impedance voltage at 75º 6%, dimensions 1735x830x1900 mm, oil weight 460 kg, total weight 2300 kilograms.

The connection between the low voltage connections of the transformer and the input switches of the low voltage switchboard, as a type solution, is with a 3 x (4 x NYY-0 1 x 2 40 mm^2 Cu RM) + 2 x NYY-0 1 x 240 mm^2 Cu RM.

V. LOW-VOLTAGE SWITCHBOARD

The low voltage installation is performed as a type switchboard on a stand-alone frame. The plant consists of a supply, a protective and an exhaust part.
The supply part is in the middle upper part of the board and the cables that are driven from the transformer's second side are connected here. It is equipped with a low-voltage three-phase circuit breaker type NS 1250N 3p 1250A, Micrologic 2.0, Schneider Electric.

The product consists of ten power supply terminals to equipped with a vertical three-phase circuit breaker with fuses 630 A.

VI. MEAUSURING

Measuring the electricity is an obligation for the electrical distribution system's operator, with single-pole isolated measuring voltage transformers along their transmission ratio 20000 / 1.73 (10000 / 1.73) /100/1.73/100/1.73 V / V / V and electrical voltage transformers with transformation ratio 2x25/5/5 A/A/A, 24 kV.

VII. COMMANDS

The circuit breaker can be manually operated onsite, using manipulating handles. Moreover, turning the LV switch on and off, can be manually controlled onsite.

VIII. BLOCKADES

In order to prevent making mistakes during usage and handling the substation's equipment, the following blockades will be provided: the separate switch can be turned on along with a special handle, it will be located in the middle voltage block, next to the entrance door and access to the MV fuses because of their replacement, it is accessible only when the separated switch is turned off, as an alternative like the only case when the door is opened and the fuses can be replaced.

IX. LIGHTING

ATS is illuminated with two 1 × 24 W lamps, IP65 housing, mounted in the sector of MV and LV installation and transformer box.

X. PROTECTION

For transformer's protection, from short-circuits currents on the medium voltage side and short-circuits currents from LV busbars, we will provide and highly powered fuses such as MV VM50 A, with a striking needle that automatically activates the circuit breaker deactivation inside the transformer filed. The transformer is protected from overloading by a secondary thermal replay(4-8)A, which automatically turns the switch off inside the transformer field on the high voltage side, in conditions of extremely increased temperature above the prescribed one.

XI. EARTHING

In order to protect our employees from highly unaccepted voltage, we will provide working and protective grounding. Besides protecting our stuff, grounding will be used in KBTS for transformer's proper and normal work, where we predict to perform grounding on the main substation.

To accomplish our working grounding goal, we are willing to perform a set of three galvanized probes, with length of 1.5 meters and a cross section of 63 millimeters. Probes will be placed as vertices of a equilateral triangle with side's length of 15 meters. Each other, probes are connected with two rows of galvanized tape FeZn 40 x 4 millimeters square, where they will be set 25 meters from substations. Whole working grounding is connected with cable H07V-K-1 x 50 millimeters square and it is directly connected with transformer's zero. (Due to the spatial availability, the inverstor during the performance will decide the exact location of the grounding and consequently its proper way of realization).

Protective grounding will be performed as a set of two rectangular shapes, placed on galvanized tape FeZn 40 x 4 millimeters square, set on appropriate position around substations. Their interconnection must be correct, as well as connection to the potential equalization rail. First shape is placed on 1 meter distance and 0.5 meters depth of substations's shapes. On the other hand, the second one is placed on 1 meter distance from the first one and 1 meter depth. Additionally to our protective grounding will be connected a galvanized steel tape, which is parallel to the middle voltage cables.

On the equalization potential will be linked: the grounding of the low voltage board, transformer's core, surge arresters, metal shears from medium voltage cables and all metal parts of substation's equipment, which are not under voltage in normal operation.

All electrical installation work will be performed in accordance of valid norms and regulations. In the end of our work, we will execute all the necessary testing, for which will be determinated and specified suitable protocols.

XII. COMPENSATION OF REACTIVE POWER OF THE TRANSFORMER

The reactive power compensation of the transformer with a power of 800 kVA is resolved with a capacitor battery alongside the LV switchboard, with a power of 50 kVAr. The capacitor battery will be connected to a single drain line on the LV switchboard via the NAY2Y-J 4x50 mm2 cable. The lead-out cable will be protected by fuses LV 125 A. This battery will only be used to compensate for the reactive power required to magnetize the core of the transformer, while for full compensation of the engaged reactive power by consumers, it

is necessary to perform appropriate analyzes of the different load modes in operation, and to select appropriate automatic compensation.

XIII. FIREFIGHTING AND PROTECTION AT WORK

In the substation, it is planned to install a rubber insulating carpet 24 kV, in the sector of the MV swichgear and LV switchboard. Also, firefighting appliances are provided in both premises. In the sector of the MV switchgear, a box with protective tools and insulation equipment should be placed. On the wall beside the MV block to place plates with a single-pole pattern, the five golden rules, a manual for indicating first aid and a warning for the danger of high voltage.

XIV. CALCULATIONS OF THE SHORT-CIRCUIT AT THE 10(20) KV BUSBAR

The choice of equipment and dimensioning of the busbar is based on the power of a three-phase short circuit of the 20 kV busbar, which equals Sk3 "= 500 MVA.
- Initial three-phase short circuit is:

$$I_{k3}'' = \frac{S_{k3}''}{\sqrt{3}\cdot U_n} = \frac{500\cdot 10^3}{\sqrt{3}\cdot 20} = 14.43 \text{ kA}$$

Peak short curcuit current is:

$$I_u = k_u \cdot \sqrt{2}\cdot I_{k3}'' = 1.75\cdot\sqrt{2}\cdot 14.43 = 35.71 \text{ kA}$$

Where:
k_u -peak coefficient that depends on the ratio R / X on the network and for R / X = 0,1 ku = 1,75;

Effective short-circuit current value is calculated bellow:

$$I_{ks} = I_{k3}''\sqrt{m+n} = 14.43\sqrt{0.2+0,8} = 14.43 \text{ kA}$$

Where:

m and n are coefficients that depend on the direct and alternating short-circuit current component. Their values are obtained from the diagram in dependence of the peak coefficient, and for the minimum switch-off time of the switch = 0,25 s.

Breaking short-circuit current is given bellow:

$$I_r = I_{k3}'' = 14.43 \text{ kA}$$

XV. CALCULATIONS OF THE SHORT-CIRCUIT AT THE 0.4 KV BUSBAR

To calculate the short-circuit currents and power of the 0.4 kV side, an equivalent impedance to the 0.4 kV side should be determined.

The calculations will use the selected transformer data. The impedance of the network, reduced to a voltage of 0.4 kV, is calculated as follows:

$$Z_Q = \frac{c\cdot U_Q}{\sqrt{3}\cdot I_{kQ}''}\left(\frac{U_{rTLV}}{U_{rMLV}}\right)^2 = \frac{1,1\cdot 20}{\sqrt{3}\cdot 14,43}\left(\frac{0.42}{21}\right)^2 = 0.352 m\Omega$$

where:
c - a voltage factor that depends on the voltage of the system;

U_Q - nominal voltage on the MV network;
U_{rTLV} - nominal voltage on the LV side of the transformer;
U_{rHLV} - nominal voltage on the MV side of the transformer;
I''_{kQ} - initial short-circuit current of MV network.
It follows that the inductive and ohmic resistance will be:

$$X_Q = 0,995\cdot Z_Q = 0.35 \text{ m}\Omega;\ R_Q = 0,1\cdot X_Q = 0.035 \text{ m}\Omega$$

The direct short-circuit impedance of the two-way transformer, as well as the ohmic resistance, are calculated according to the transformer data:

$$Z_{TLV} = \frac{u_{krT}(\%)}{100}\cdot\frac{U_{rTLV}^2}{S_{rT}} = \frac{6}{100}\cdot\frac{0.42^2}{0.8} = 0.0132\,\Omega$$

$$R_{TLV} = \frac{P_{krT}}{3\cdot I_{rTLV}^2} = \frac{8700}{3\cdot 1156^2} = 0.0022\ \Omega$$

$u_{krT(\%)}$ - short-circuit voltage of the transformer;
S_{rT} - nominal power of the transformer;
P_{krT} - losses in the windings of the transformer at a nominal current;
I_{rTLV} - nominal current on the LV side of the transformer.
As it follows, the inductive resistance of the transformer will be:

$$X_{TLV} = \sqrt{Z_{TLV}^2 - R_{TLV}^2} = 0.013\,\Omega$$

Equivalent impedance of 0.4 kV busbar:

$$R_e = R_Q + R_{TLV} = 2.235 \text{ m}\Omega;$$

$$X_e = X_Q + X_{TLV} = 13.35 \text{ m}\Omega;$$

$$Z_e = \sqrt{R_e^2 + Z_e^2} = 13.53 \text{ m}\Omega$$

The calculated value of the maximum current of a symmetrical three-phase short-circuit is:

$$I_{k3}'' = \frac{cU_n}{\sqrt{3}Z_e} = \frac{1,1\cdot 0,4\cdot 10^3}{\sqrt{3}\cdot 13.53\cdot 10^{-3}} = 18.84 \text{ kA}$$

For ratio R / X = 0.17 it is read ku = 1.7 and calculated the value of the peak current of a three-phase short circuit, the Breaking and durable short-circuit current is adopted to be:

$$I_r = I_t = I_{k3}'' = 5.928 \text{ kA}.$$

XVI. 20 KV CABLE DIMENSIONING

Nominal current of 20 kV busbar is:

$I_n = \frac{S_n}{\sqrt{3} \cdot U_n} = \frac{800 \cdot 10^3}{\sqrt{3} \cdot 10 \cdot 10^3} = 46.24$ A - when working at 10 kV level;

For the 20 kV connection power transformer-20 kV switchgear the selected cable type NA2XS (F) 2Y 1 × 50 mm2, with crosslinked polyethylene insulation, can be charged with a current of 200 A.

Thermal control of the cable (short circuit control)

The minimum allowable section of the conductors will be:

$$A_{\min} = C \cdot I''_{k3} \cdot \sqrt{t} = 10.9 \cdot 14,43 \cdot \sqrt{0.004} = 9.95 < 50 \text{ mm}^2$$

The conclusion is that the cable is thermically satisfactory.

XVII. 0.4 KV CABLE DIMENSIONING

$$I_n = \frac{S_n}{\sqrt{3} \cdot U_n} = \frac{800 \cdot 10^3}{\sqrt{3} \cdot 0.4 \cdot 10^3} = 1156 \text{ A}$$

For 0,4 kV connection energy transformer - 0,4 kV switchgear, the selected cable type NYY 4 × (1 × 240) mm2, with PVC insulation, can be charged with a current of 4 x 520 A = 2040 A, to accept the coefficient of load reduction due to poor heat dissipation in parallel cable guidance, which in the case of a bundle of four cables in the air is 0,65:

$$I_n = 1156 < 0.65 \cdot 2040 = 1326A$$

Thermal control of the cable (short circuit control)
Accordingly, the minimum allowable section of the conductors will be:

$$A_{\min} = C \cdot I''_{k3} \cdot \sqrt{t} = 8.9 \cdot 18.84 \cdot \sqrt{1} = 167 mm^2 < 4 \times 240 = 960 \text{ mm}^2$$

The conclusion is that the cable is thermically satisfactory.

VI. CONCLUSION

This paper presents the use of standardized solutions currently used in the Republic of Macedonia in the design of typical and adapted substations, respecting all positive norms, laws, technical recommendations, as well as the recommendations of EVN Macedonia. Here is shown how one adapted substation with an installed power of 800 kVA is designed, how the elements are dispositioned, calculations of short circuits and at the same time the dimensioning of the equipment.

REFERENCES

[1] EVN Macedonia Handbook - Transformer Tech
[2] EVN Macedonia Handbook - Cable Tech
[3] Mile Spirovski, Igor Sterjovski, Projects from Mehanotehnika DOOEL, Bitola

Correlation between Power System Load and Air Temperature in Republic of Macedonia

Metodija Atanasovski[1], Mitko Kostov[1], Blagoja Arapinoski[1], Igor Andreevski[1]

Abstract – **The goal of this paper is to investigate the correlation between power system load and air temperature in Republic of Macedonia. Power system load hourly values (8760 per year) for year 2014 and 2015 are used as input data. In absence of hourly values for air temperature in Republic of Macedonia, data from internet is used for air temperature in city of Skopje for each day of years 2014 and 2015. Linear and nonlinear correlation coefficients are calculated between minimum, average and maximum load as dependent variables and respective air temperatures as independent variables. Results show very strong dependence between power system load and air temperature. Therefore, in order to determine a relation between three typical loads and appropriate air temperatures, a procedure of linear and nonlinear regression is performed.**

Keywords – **Power system load, Air temperature, Regression**.

I. INTRODUCTION

Power system load curve shows load variation as a function of time and its analysis is very important for determination of a set of factors that have impact on consumption of electricity in power system. Those factors are: increase and structure of gross domestic product, demographic changes, people's standard, energy efficiency, climate changes, people's mobility, habits, customs etc. Air temperature is one of the factors which can have significant impact on electricity consumption and power system load. This fact is evident in power system of Republic of Macedonia due to high variations of consumption and load in year seasons.

Correlation between power system load and air temperature is deeply investigated in the literature. There are number of techniques which have been used for load forecast, such as: single or multiple linear or nonlinear regressions, stochastic time series, exponential smoothing, state space and Kalman filter, knowledge based approach, neural networks, wavelet transformations, semi parametric additive model, etc.

The goal of this paper is to investigate the correlation between power system load and air temperature in Republic of Macedonia. Power system load hourly values (8760 per year) for years 2014 and 2015 are used as input data. In absence of hourly values for air temperature in Republic of Macedonia, data from Internet is used for minimum, average and maximum air temperature in city of Skopje for each day of years 2014 and 2015 [1]. Complete analysis in the paper is performed with EXCEL and SPSS (Statistical Package for the Social Sciences) software package specialized for statistical analysis (demo version for educational and research purposes) [2].

The paper is consisted of five sections. Input data is elaborated in section II and basic statistical analysis is done. Coefficients of correlation for linear and nonlinear dependence between minimum, average and maximum load as dependent variables and minimum, average and maximum air temperatures as independent variables are calculated in section III. Results show very strong dependence between power system load and air temperature. Section IV is the main contribution of this paper. Namely, procedure of linear and nonlinear regression is performed for determining the functions of variation between three typical loads and average air temperature. Several conclusions and further work possibilities about this matter are presented in section V.

II. INPUT DATA AND BASIC STATISTICAL ANALYSIS

Power system load hourly values (8760 per year) for year 2014 and 2015 are used as input data (Fig. 1) [3]. As it was emphasized in the introduction, in absence of hourly values for air temperature in Republic of Macedonia, data from Internet is used for minimum, average and maximum air temperature in city of Skopje for each day of year 2014 and 2015. According to this 3x730 temperatures are available (3 for each day in 2014 and 2015) and in order to have equal number of data for both stochastic variables (load and air temperature), minimum, average and maximum load is determined for each day of year 2014 and 2015. This approach gives opportunity to investigate the correlation between 3 daily loads in power system maximum (Pmax), average (Pavg) (average of 24 hourly load values for each day), minimum (Pmin) and respective temperatures for each day of year 2014 and 2015. Figure 2 depicts one daily diagram of power system of Republic of Macedonia for the day 08.01.2014. Three typical points are marked on the diagram and they are used in the analysis from each daily diagram. Average load is average of all 24 hourly loads in one daily diagram.

Maximum load of 1507 MW was obtained in year 2014. It is registered on December 31 at 6.00 PM. Minimum temperature of - 10^0C in year 2014 was also registered on December 31. Maximum load of 1439 MW was obtained in year 2015 on January 8 at 11.00 PM. Minimum daily temperature of -16 ^{0}C in year 2015 is registered on January 2. Minimum daily temperature on January 8 was -14 ^{0}C.

Maximum summer load of 1017 MW in year 2014 was registered on July 27 at 10.00 PM (maximum daily temperature on that day was 31 ^{0}C, what is 5^0C less than

[1]Metodija Atanasovski, Mitko Kostov, Blagoja Arapinoski and Igor Andreevski are with the Faculty of Technical Sciences-Bitola, St. Kliment Ohridski University, Republic of Macedonia, E-mail: metodija.atanasovski@tfb.uklo.edu.mk.

maximum daily temperature of 36 ^{0}C registered on August 13 in year 2014).

Maximum summer load of 976 MW in year 2015 was registered on July 21 at 2.00 PM (maximum daily temperature on that day was 37^{0}C, what is 1^{0}C less than maximum daily temperature of 38 ^{0}C registered on July 29 in year 2015).

Minimum system load of 555 MW is registered on September 14 at 5 AM in year 2014 (temperature at that day varied in range from 15^{0}C to 22^{0}C). Minimum system load of 530 MW is registered on June 2 at 6.00 AM in year 2015 (temperature at that day varied in range from 12 ^{0}C to 29^{0}C). Minimum - maximum load ratio in power system of Republic of Macedonia in both years was 0.3683. This ratio implies high variation between hourly loads in the power system on year basis.

Hourly loads at days when maximum load (31. 12. 2014; 8. 1. 2015) and minimum load (14. 9. 2014; 2. 6. 2015) is registered in the system, have shown daily variation of several hundred of MWs (from 370 MW to 470 MW).

Basic statistical parameters for minimum (Pmin), average (Pavg) and maximum (Pmax) load of power system and minimum (Tmin), average (Tavg) and maximum (Tmax) air temperature are given in Table I for year 2014 and 2015.

TABLE I
BASIC STATISTICAL ANALYSIS OF INPUT DATA FOR DAILY LOADS AND AIR TEMPERATURES FOR YEAR 2014 AND 2015

Statistical parameter	Pmax (MW)	Pavg (MW)	Pmin (MW)	Tmax (^{0}C)	Tavg (^{0}C)	Tmin (^{0}C)
Mean	1073	896,9	698,2	18,9	13	7
Stdev	164,6	141,2	118	9,96	8,3	7,3
Median	1039,5	850	658,5	19	13.5	7
Range	684	593	513	42	39	38
Minimum	823	696	530	-4	-9	-16
Maximum	1507	1289	1043	38	30	22

Maximum daily load during years 2014 and 2015 varied in range of 684 MW (823-1507) MW, with mean value of 1073±164.6 MW. Average daily load in year 2014 and 2015 varied in range of 593 MW (696-1289) MW, with mean value of 896.6±141.2 MW. Minimum daily load during year 2014 and 2015 varied in range of 513 MW (530-1043) MW with mean value of 698.2±118 MW.

Maximum daily temperature during year 2014 and 2015 varied in range of 42 ^{0}C (-4 - 38) ^{0}C, with mean value of 18.9±9.96 ^{0}C. Average daily temperature during year 2014 and 2015 varied in range of 39 ^{0}C (-9 - 30) ^{0}C, with mean value of 13±8.28 ^{0}C. Minimum daily temperature during year 2014 and 2015 varied in range of 38 ^{0}C (-16 - 22) ^{0}C, with mean value of 7±7.3 ^{0}C.

III. COEFFICIENTS OF CORRELATION

Coefficients of correlation (Pierson coefficients) are given in Table II. These coefficients are measure for linear relation between daily maximum, average, minimum power system load and daily maximum, average and minimum air temperature. Coefficients of correlation can have values in range from -1 to 1 (negative to positive linear relation). Empiric rules for correlation coefficients are [4]: Correlation coefficient less or equal of 0.3 shows not significant linear relation; Correlation coefficient between 0.5 and 0.7 shows linear correlation with practical importance; Correlation coefficient between 0.7 and 0.9 shows close linear correlation; Correlation coefficient greater than 0.9 shows very close linear correlation.

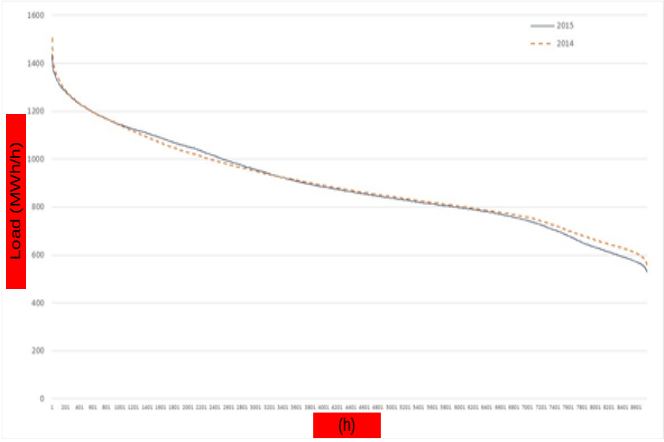


Fig. 1. Load duration curve for year 2014 and 2015 for power system of Republic of Macedonia

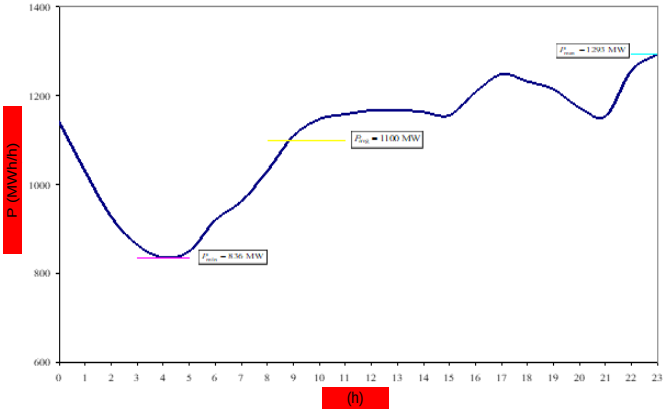


Fig. 2. Daily load diagram for 08.01.2014 and typical load points

According to results presented in Table II, the dependence between daily power system typical loads and air temperature can be defined as close to very close linear relation. Correlation coefficients have values in range from -0.8 to -0.9 what implies close to very close negative linear relation between all combinations of typical daily loads and air temperatures.

TABLE II
PIERSON COEFFICIENTS OF CORRELATION

P/T	Tmax	Tavg	Tmin
Pmax	-0.873	-0.909	-0.875
Pavg	-0.868	-0.895	-0.850
Pmin	-0.839	-0.864	-0.825

Index of nonlinear correlation is used as a measure for nonlinear relation between two variables (has values between 0 and 1). Table III summarizes indexes of nonlinear correlation for the case when relation between power system load and air temperature is treated as quadratic and as cubic function. Indexes of correlation are in range from 0.85 to 0.95. These values have shown very close nonlinear correlation between all combinations of daily loads and air temperatures. Coefficients are calculated with EXCEL and SPSS.

TABLE III
INDEXES FOR NONLINEAR CORRELATION FOR QUADRATIC AND CUBIC FUNCTION

	Quadratic		
P/T	Tmax	Tavg	Tmin
Pmax	0.892	0.923	0.876
Pavg	0.901	0.921	0.856
Pmin	0.851	0.889	0.848
	Cubic		
P/T	Tmax	Tavg	Tmin
Pmax	0.902	0.939	0.887
Pavg	0.917	0.943	0.869
Pmin	0.860	0.903	0.858

IV. REGRESSION ANALYSIS

Linear and nonlinear regression analysis is performed using SPSS software package. The goal of this analysis is to estimate dependence curve of the three typical loads (Pmin, Pavg, Pmax) from independent variables presented with three typical temperatures (Tmin, Tavg, Tmax). The curve estimation of linear, quadratic and cubic function of typical daily loads from average temperature is only presented in the paper. Fig. 4 depicts estimated linear, quadratic and cubic functions between minimum (Pmin), average (Pavg) and maximum (Pmax) daily load and average daily temperature (Tavg).

Table IV summarizes function equations presenting polynomial coefficients and determination coefficients (R^2). Determination coefficient shows how many percent of dependent variable are predictable with independent variable using regression analysis (it is in range from 0 to 1) [4].

For example determination coefficient of maximum daily load due to average daily temperature is in range from 88.2% to 82.7%, depending on used nonlinear and linear regression. Determination coefficient of average daily load due to average daily temperature is between 89% and 80.1%, depending on used nonlinear and linear regression.

Minimum daily load due to average daily temperature is predictable 81.5% with estimated cubic function, 79% with quadratic function and 74.6% with linear estimated function.

Regression analysis shows high prediction degree of daily typical loads from air temperature. Cubic estimated function of daily loads from average daily temperature shows highest degree of prediction.

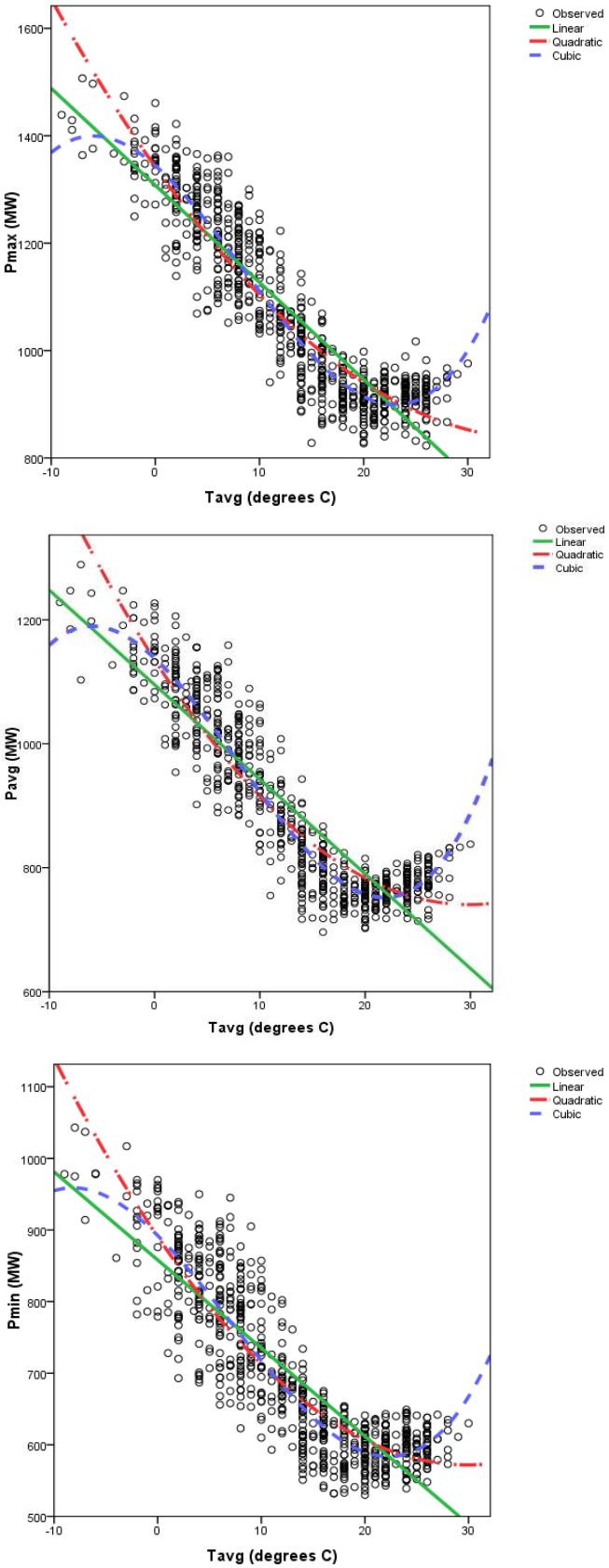


Fig. 3. Estimated linear quadratic and cubic functions of maximum (Pmax), average (Pavg) and minimum (Pmin) daily load from average daily temperature (Tavg)

TABLE IV
ESTIMATED FUNCTIONS COEFFICIENTS AND DETERMINATION INDEXES

Function	R^2	Constant	b1*T	$b2*T^2$	$b3*T^3$
Pmax=F(Tavg)					
Linear	0.827	1307.618	-18.073		
Quadratic	0.852	1341.734	-17.54	0.375	
Cubic	0.882	1344.212	-16.986	-1.045	0.041
Pavg=F(Tavg)					
Linear	0.801	1095.251	-15.262		
Quadratic	0.849	1135.576	-26.452	0.443	
Cubic	0.890	1138.047	-15.921	-0.974	0.041
Pmin=F(Tavg)					
Linear	0.746	858.073	-12.304		
Quadratic	0.790	890.537	-21.313	0.356	
Cubic	0.815	892.136	-14.504	-0.56	0.026

V. CONCLUSION

According to authors information the paper is the first one which has investigated correlation between power system load and air temperature in Republic of Macedonia. In absence of hourly values for air temperature in Republic of Macedonia, correlation between daily minimum, average and maximum load and air temperature for all days in year 2014 and 2015 is investigated. On the basis of statistical analysis it can be noticed that there is close time matching in appearance of power system maximum load and minimum air temperature. The same time matching is noticed between power system summer maximum load and registered maximum temperature in analyzed years.

The usual approach in investigation of load correlation with air temperature is to establish the multi regression model, because the load depends on temperature, wind, season, hour, day in the week and other factors (holidays for instance). However, in absence of relevant and available data about temperature, wind, solar radiation and other factors, the analysis in the paper was performed on the basis of single linear and nonlinear regression approach. Temperature is only used as climate parameter that has significant impact on load variation in power system of Republic of Macedonia. Also it has to be emphasized that in Macedonia electricity is widely used for heating of residential and commercial buildings and houses. This fact explains the very close linear and nonlinear relation between power system load and air temperature in Macedonian power system.

Calculation of linear and nonlinear correlation coefficients has shown close to very close linear and nonlinear relation between all combinations of daily typical loads and air temperatures.

Regression analysis and obtained estimated functions have shown high degree of predictability of power system minimum, average and maximum load and average air temperature. High values of determination coefficients between 70% and 89% are the best prove for this fact. Cubic estimated function has shown highest degree of predictability.

Authors' opinion is that achievement of higher exactness in regression analysis for obtaining estimated functions between power system load and air temperature can be done with use of hourly values for air temperature for total territory of Republic of Macedonia. In that case presented analysis can be done for months of the year, using hourly allocation of load to substations and air temperature in the town where substation is located. However, having on mind consumers load temperature inertia, the hourly temperature may not be a good solution as a predictor variable, and the usage of minimum, average and maximum temperature is possible to be a better option. These facts have to be verified by intensive analysis using multiple years in the past.

Also sophisticated mathematical algorithms can be implemented for correlation investigation, such as wavelet transformations which are adequate for correlation analysis in time-frequency space.

REFERENCES

[1] https://www.wunderground.com/history
[2] SPSS, software product of IBM, https://www.ibm.com/analytics/us/en/technology/spss/
[3] Davor Bajs, Metodija Atanasovski, *Longterm Forecast Study of Electrical Energy and Power Balance and Adequacy Analysis of Transmission Network of Republic of Macedonia*, Zagreb/Skopje EIHP, 2016.
[4] Svetozar Vukadinovic., *Elements of Probability Theory and Mathematical Statistics*, Belgrade, 1973.

Effects of Measurement System on Transfer Function of Gas Mask Speech Membrane

Dejan Ćirić[1], Marko Ličanin[2], Ana Đorđević[3] and Miljan Miletić[4]

Abstract – **Speech membrane is one of important elements of some masks such as gas or oxygen mask. This membrane can be characterized from acoustical point of view by certain parameters and quantities such as cut-off frequencies, distortions and transfer function. Regarding the measurement of the membrane transfer function, the results should be free of effects of the measurement system itself, which might be a challenging task. These effects are topic of the present research. Focus here is on influence of a device coupling the sound source and tested speech membrane as well as of position of microphone in reference to the sound source. For that purpose, five different couplers and four microphone distances are applied.**

Keywords – **Gas mask, Speech membrane, Transfer function, Measurement system.**

I. Introduction

There are situations where wearing a mask of specific type is of crucial importance. These include flying in a high-performance aircraft, military actions or activities of firefighters. By placing a mask on a person's face, the conditions for speech generation and emission are changed in comparison with a situation without the mask (no mask situation) [1,2]. The impedance loading the vocal tract is changed since the cavity between the mask and person's face is formed, and there is a specific force imposed on jaw by the mask affecting the speech generation [3-5]. When the vocal track of a speaker is closed with a mask chamber, it seems feasible to introduce the mask as a part of the speech chain.

The mentioned mask effects influence the procedure of speech generation, but also properties of the produced speech, which differentiates from the no mask case. This is why it is reasonable to asses the impact of the mask, and in some cases even to compensate for some adverse effects, as done in some studies, see, for example [1].

Regarding transfer of speech from a mask interior to surrounding environment, some masks have more or less complex device with a main task to enable as efficient as possible speech transfer, and to provide enough efficient protection from fluids (gases) existing in a mask surrounding. This device can be based on a special membrane or diaphragm called speech or speaking membrane. Acoustic behaviour of such a membrane can be described using the quantities such as frequency bandwidth (cut-off frequencies), distortion characteristics and transfer function.

In measurement of the speech membrane transfer function, it is of importance how to couple the membrane to the sound source and microphone, and minimize as much as possible the effects of the measurement system. This is analyzed here using five different couplers placed on an artificial mouth used as a sound source. Coupling of the membrane to the microphone is realized through a free space, where microphone is placed at different distances from orifice of the artificial mouth.

[1]Dejan Ćirić is with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: dejan.ciric@elfak.ni.ac.rs.

[2]Marko Ličanin is with the Faculty of Occupational Safety, University of Niš, Čarnojevića 10a, 18000 Niš, Serbia.

[3]Ana Đorđević is with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia.

[4]Miljan Miletić is with the College of Applied Technical and Technological Sciences, Kosančićeva 36, 37000 Kruševac, Serbia.

II. Speech membrane effects

In literature, there are studies dealing with change of the speech spectrum, formant frequencies and other parameters of speech caused by different masks, see, for example [4,6,7]. However, literature presenting and analyzing speech membrane is rather rare. There some patents dealing with this topic, e.g., [6,8,9], and the transfer function from inside to outside of the mask is given in [1].

A cavity enclosed by a mask placed on a talker's face has resonances in the speech pass-band. The first resonant frequencies occur in the frequency range where the enclosed interior dimensions have the same order of magnitude with the sound wavelength [4]. The second important effect of a mask comes from a force due to a tight fit of the mask preventing normal jaw motion. These effects result in changes of long-time spectrum of continual speech mostly in frequency range over 800 Hz [4]. There are also some other factors affecting the speech spectrum in the mask situation.

In order to improve the speech transfer through a mask, different systems or elements (sometimes called voice emitting port, see Fig. 1) can be used. One of possible solutions is to use the speech membrane/diaphragm typically made of thin sheet material (or multiple sheets), such as cellulose film or plastic (polyvinylidene chloride) [8]. Protective filters made of metal or plastic (peripherally perforated) are commonly used to protect the speech membrane against mechanical damage. When the speech is transmitted through such a membrane (port), the transmission is called direct one, while transmission through an electro-acoustic system using a microphone is called indirect one [6].

It is questionable how to test and assess the speech membrane from an acoustical point of view. If a parallel between the speech membrane and any other membrane is drawn, then the speech membrane could be tested in the same way as other membranes. In most of such cases, this would

mean that the speech membrane is measured in an impedance or transmission loss tube, see for example [10]. Another possibility is to use a specific measurement system containing an artificial mouth as a sound source to which the tested membrane is coupled. Such a system can be used for measurements of frequency bandwidth, distortions and transfer function.

Fig. 1. Gas mask with indicated voice emitting port

III. Measurement System and Procedure

Transfer functions between the sound source (the artificial mouth B&K 4216) and measurement microphone (B&K type 4144) located at certain distances (heights) above the source were measured. The microphone surface was positioned exactly 5 cm, 8 cm, 9 cm and 12 cm above the orifice of the artificial mouth. The gas mask speech membrane was placed in between the source and microphone, and it was coupled to the artificial mouth through a device - element here called coupler. The measurement system was installed in a small anechoic chamber of approximate dimensions 1.2×0.4×0.8 m, see Fig. 2. The excitation signal was an exponential swept sine of length of 5 s. Every measurement was repeated five times under the same conditions.

Five different couplers were used, see Fig. 3(a). Three of them (couplers 1, 2 and 3) are made of clay with interior shape of a conical waveguide. The coupler entrance (lower) port is of approximately the same area as the opening of the artificial mouth. The outer (upper) port is of approximately the same area as those of the tested speech membrane. The bottom surface of these couplers fits closely the top inclined surface of the artificial mouth. The heights (waveguide lengths) of the clay couplers are 0.6 cm, 1 cm and 1.9 cm. The fourth coupler is a metal ring with circular opening of height of approximately 1.1 cm. The fifth coupler represents a ring of isolating tape with a holder. The height of this coupler is about 1.8 cm. Fitting of the speech membrane to the coupler is improved by using a rubber ring placed on the coupler. Transfer function of two speech membranes was measured, the membrane 1 and 2, shown in Fig. 3(b).

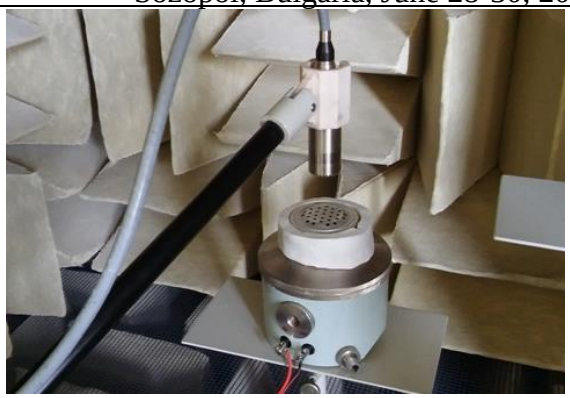


Fig. 2. Measurement of transfer function of speech membrane in a small anechoic chamber using the coupler 2

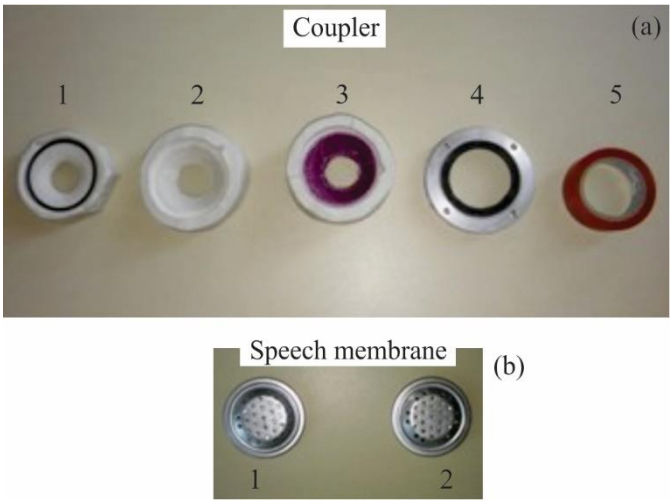


Fig. 3. (a) Five couplers for coupling the speech membrane to the artificial mouth; (b) two used speech membranes

IV. Results

The smallest value of signal to noise ratio is above 25 dB, while its largest value is close to 80 dB showing that the measurements are reliable from that point of view. The transfer functions of the artificial mouth measured without any coupler and speech membrane are shown in Fig. 4. The functions for different distances of the measurement microphone from the artificial mouth are almost parallel. Two prominent peaks at about 1.5 kHz and slightly above 5 kHz as well as a dip at about 4 kHz can be seen in the presented transfer functions. The mouth response decreases below 300 Hz.

By placing a coupler on the artificial mouth, the response of the mouth is changed, see Fig. 5. The couplers cause a certain increase of transfer function amplitude in a large part of the analyzed frequency range. At frequencies up to the first peak at about 1.5 kHz, the presented transfer functions are almost parallel. Above that frequency, there are some specific behaviors associated to every particular coupler. The two peaks seen in the functions measured without couplers are also prominent in the functions measured with the couplers.

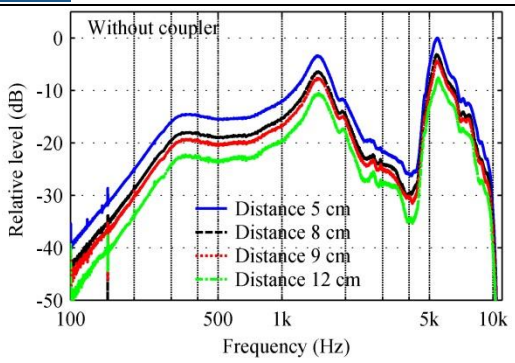


Fig. 4. Transfer functions of the artificial mouth without coupler and membrane measured at different microphone distances

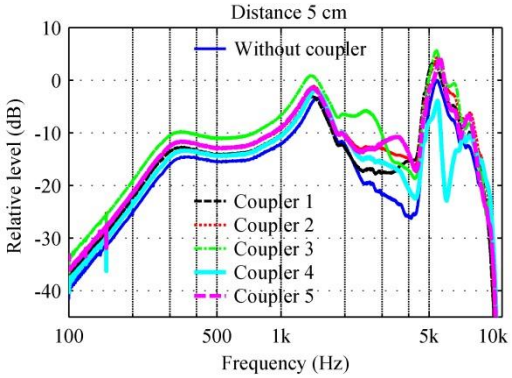


Fig. 5. Transfer functions measured without coupler and with all 5 couplers without membrane for microphone distance of 5 cm

Similar as in the case of transfer functions of the artificial mouth, change of the microphone distance in the measurements with the couplers placed on the mouth leads to almost parallel transfer functions, as shown in Fig. 6 for the coupler 1. Here, closer the microphone to the artificial mouth, higher the level of the transfer function. The situation is similar for other couplers.

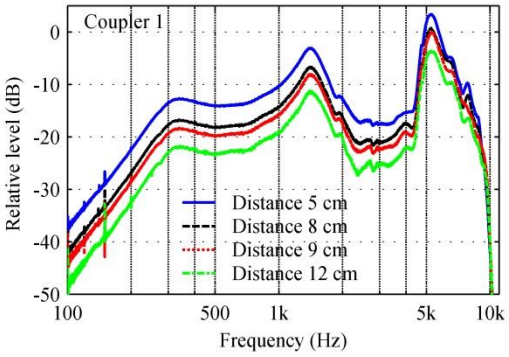


Fig. 6. Transfer functions measured with coupler 1 and without speech membrane for different microphone distances

The function obtained by subtracting the transfer function measured without a coupler from the transfer function measured with the coupler is here called transfer function of the coupler. Such transfer functions for the coupler 1 and different microphone distances are presented in Fig. 7. The functions are very similar to each other. The main trends are the same, and there are only slight deviations among the curves. The main contribution of this coupler is seen in the frequency range above 2 kHz. There is a prominent peak at about 4 kHz. The characteristics of other couplers are in accordance with the results given in Fig. 5.

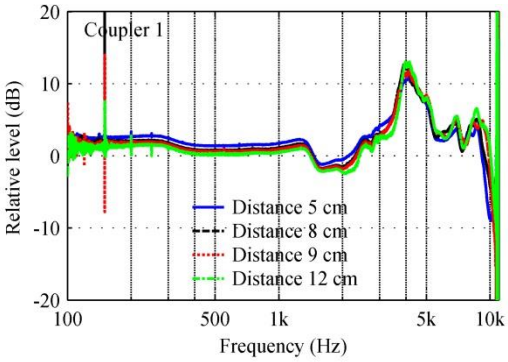


Fig. 7. Transfer functions of the coupler 1 obtained as a difference of the transfer functions measured with and without coupler for different microphone distances

When the speech membrane is placed on the coupler, it causes certain changes of the transfer functions (see Fig. 8) measured without the membrane, shown in Fig. 6. The peak at about 1.5 kHz is less prominent, and there is an additional peak between 3 kHz and 4 kHz. The transfer functions decreases with frequency decrease below 500 Hz, having a peak at about 200 Hz. These changes caused by the speech membrane are easier to be observed in the functions obtained as a difference between the transfer functions measured with and without the membrane, called here transfer functions of the membrane. They are shown in Fig. 9. It is interesting to note that the smallest influence of the membrane itself exists in the frequency range between 500 Hz and 1.5 kHz.

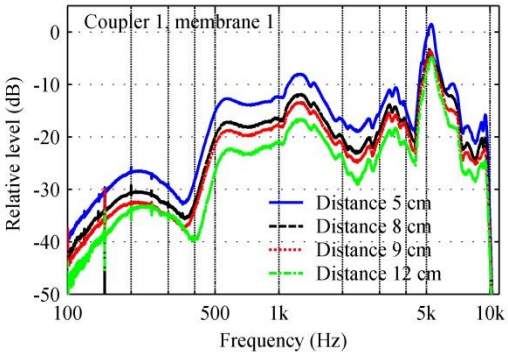


Fig. 8. Transfer functions measured with coupler 1 and membrane 1 for different microphone distances

For the used measurement setup, the effects of changing the microphone distance on the transfer functions of the membranes seem to be rather small up to about 10 kHz. Thus, the functions of the membrane 1 for different microphone distances coincide very well with each other, as shown in Fig.

10. There is a slight difference at lower frequencies (below several hundreds of Hz), and this difference might be a consequence of not completely stable leakage between the coupler and artificial mouth as well as between the speech membrane and the coupler. The situation is similar for another speech membrane used here – the membrane 2, see Fig. 11. This membrane has also some specific features, and the most important difference in comparison with the membrane 1 is ringing manifested by fluctuations of the transfer functions.

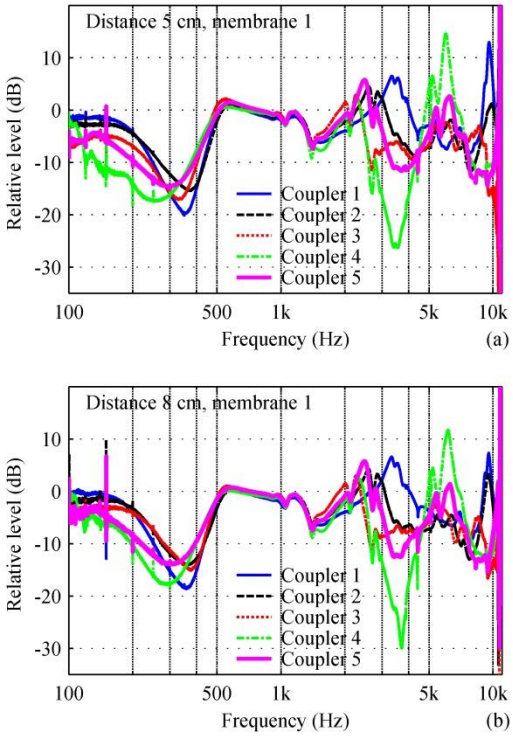


Fig. 9. Differences between transfer functions measured with particular coupler with membrane 1 and without the membrane for microphone distance of (a) 5 cm and (b) 8 cm

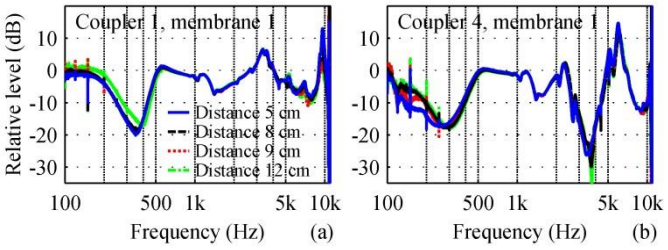


Fig. 10. Differences between transfer functions measured with coupler (a) 1 and (b) coupler 4 with membrane 1 and without the membrane for different microphone distances

V. Conclusion

A specific system for measurements of some acoustic properties of a gas mask speech membrane is presented here. One of these properties is the membrane transfer function. It is shown that in spite of normalization done through the subtraction of the response (transfer function) of the measurement system and coupler, the obtained transfer function of the membrane still depends on the coupler.

The smallest influence of the coupler is between about several hundreds of Hz (e.g., 500 Hz) to approximately 1.5 kHz. The deviations at lower frequencies might be attributed to the effects of leakage, while the deviations at higher frequencies can be attributed to different resonant properties of the couplers in two conditions – open (without the membrane) and closed (with the membrane). Taking into account the presented results, the preference can be given to the coupler 1 of height of 0.6 cm. For the setup applied here, the effects of distance of the microphone from the sound source are not considered to be of large importance.

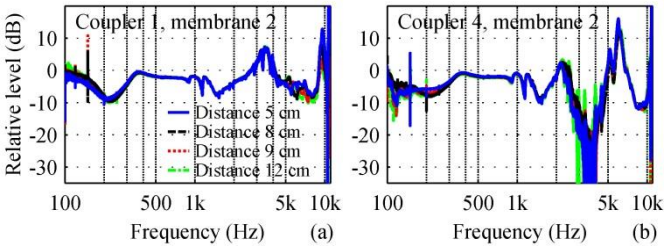


Fig. 11. Differences between transfer functions measured with coupler (a) 1 and (b) coupler 4 with membrane 2 and without the membrane for different microphone distances

References

[1] W. M. Kushner, S. M. Harton, R. J. Novorita, "The Distorting Effects of SCBA Equipment on Speech and Algorithms for Mitigation," 13th European Signal Processing Conference, Antalya, Turkey, 2005.

[2] H. Sari, *Underwater Acoustic Voice Communications Using Digital Techniques*," Ph.D. dissertation, Loughborough University, UK, 1997.

[3] M. Vojnović, M. Mijić, D. Šumarac Pavlović, "A simplified Model of Mouth Radiation Impedance Closed by Mask Cavity," Applied Acoustics, vol. 115, no. 1, pp. 3-5, 2017.

[4] M. Vojnović, M. Mijić, "The Influence of the Oxygen Mask on Long-Time Spectra of Continuos Speech," Journal of Acoustical Society of America, vol. 102, no. 4, pp. 2456-2458, 1997.

[5] C. T. Morrow, A. J. Brouns, "Speech Communication in Diving Masks. I. Acoustics of Microphones and Mask Cavities," Journal of Acoustical Society of America, vol. 50, no. 1A, pp. 1-9, 1971.

[6] H. Ryback, "Protective Breathing Mask Having a Speaking Diaphragm for Close Communication and an Electroacoustic Transducer System for Indirect Speech Transmission from Inside the Mask," US Patent no. 4,756,308, 1988.

[7] Z. S. Bond, T. J. Moore, B. Gable, "Acoustic-Phonetic Characteristics of Speech Produced in Noise and while Wearing an Oxygen Mask," Journal of Acoustical Society of America, vol. 85, no. 2, pp. 907-912, 1989.

[8] A. V. Motsinger, "Gas Mask Spech Transmission," US Patent no. 3,140,754, 1964.

[9] G. Vandeputte, "Gas Mask Coupled to Monolithic Member with Speech Membrane," US Patent no. 4,957,106, 1990.

[10] F. Langfeldt, J. Riecken, W. Gleine, O. von Estorff, "A Membrane-Type Acoustic Metamaterial with Adjustable Acoustic Properties," Journal of Sound and Vibration, vol. 373, no. 7, pp. 1-18, 2016.

POSTER SESSIONS

RADIO COMMUNICATIONS, MICROWAVES, ANTENNAS AND MEASUREMENTS SCIENCE AND TECHNOLOGY – POSTER SESSION

Minkowski Fractal Yagi Antenna

Boncho Bonev[1], Zornitsa Radkova[2], Luboslava Dimcheva[3] and Peter Petkov[4]

Abstract –**In this paper Yagi antenna with elements shaped as modified Minkowski first order curve is presented. The antenna is designed and optimized for 1800 MHz frequency band. Usage of modified fractal gives possibilities for achievement of wider bandwidth and almost the same gain than in case of classical Yagi antenna with small antenna size. The return losses (SWR), antenna gain, front-to-back ratio and radiation pattern are simulated and discussed.**

Keywords – **Fractal Antenna, GSM 1800, LTE, Minkowski curve, Yagi antenna.**

I. INTRODUCTION

The interest of researchers in fractal antennas experienced significant growth in the resent years [1-4], because these antennas allow achievement of multiband and wideband performance and relatively small size. That matches with the needs of modern communications.

Yagi arrays can be used to achieve high gain with a relatively simple antenna structure [1]. Despite these incontestable advantages, these antennas have limited frequency bandwidth. Usage of fractal antenna elements in Yagi antennas could lead to wideband and multiband performance. In other hand the analysis and optimization of fractal antennas are very complicated [2], because after every adjustment of the length or width of one of the elements or the distances between elements the optimization has to start almost from the beginning.

In our previous works [5-7] have been presented our studies on fractal antennas - studies on different types of fractal antennas and antennas designed with fractal modifications. Fractal modification allows achievement of wideband performance [7]. Despite the Koch curve is most used in wired antenna design in this paper are proposed and analyzed a three element fractal Yagi antenna based on Minkowski curve. This approach represents a wideband and multiband performance and relatively small size. The main antenna parameters - radiation pattern, return loss, antenna gain and front-to-back (F/B) ratio were simulated in comparison with these of conventional Yagi antenna optimized for the same frequency band.

[1]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.

[2]Zornitsa Radkova is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria. E-mail: zornitsaradkova@abv.bg

[3]Luboslava Dimcheva is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria. E-mail: lddimcheva@gmail.com

[4]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

II. MINKOWSKI CURVE

The antenna based on the Koch curves is very close in its properties to the antenna which is formed from Minkowski curve. As is shown in Fig.1 the shape of the Minkowski curve is formed by taking the initiator (a) and the generator structure (b) which is the first iteration of the recursive process.

In contrast to the Koch curve where triangles are used, the Minkowski curve uses squares (rectangles in some modifications) for its geometry. [8]

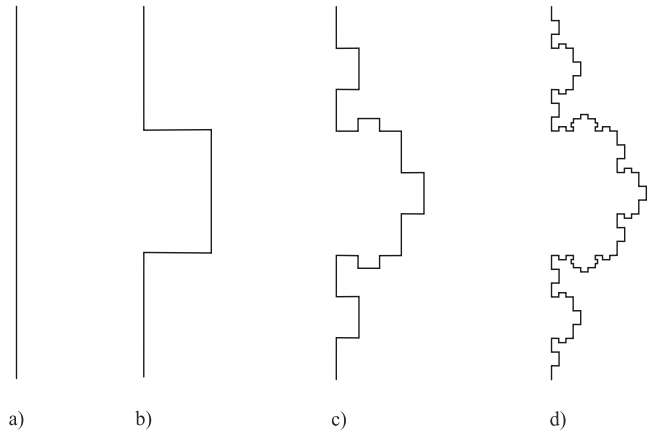


Fig. 1. Minkowski curve: a) initiator; b) first iteration; c) second iteration d) third iteration

The first iteration increases antenna gain but the next iterations do not influence the gain, they only broaden the frequency bandwidth and the antenna itself becomes more compact.

As is the case with the Koch curve, only the first 5 or 6 iterations are effective. Increasing the iterations leads to reducing the diameter of the conductors which in its own way increases the resistance, i.e. that results in gain loss. [9]

There is a modification of Minkowski fractal which is based on rectangular form of the curve rather than square. This modification is called Fractal Rectangular Curve – FRC and it is shown on Fig. 2. The length of the rectangle is $L/3$ and height is $L.r/3$, where L denotes length of original antenna and r denotes ratio coefficient. [8, 10]

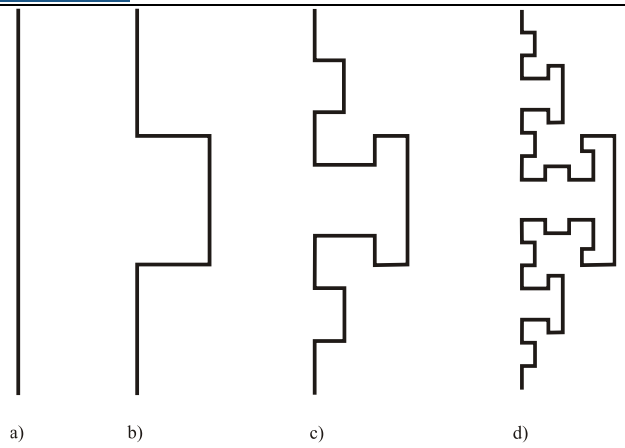


Fig. 2. Fractal Rectangular Curve: a) initiator;
b) first iteration; c) second iteration d) third iteration

III. The Minkowski Fractal Antenna Design

The proposed Yagi antenna (Fig. 3) is designed with a plane structure in order to save space and easiness of manufacturing. It is optimized for GSM 1800 MHz band which is also widely used in 4G technology. The physical dimensions of the antenna are given in Table 1, where the small letters from *a* to *d* designate the fractal segments, without digit for dipole element, with 0 for director and with 1 for reflector, the capital letters *D*0 and *D*1 designate the distance between driving element and reflector and driving element and director (Fig. 3). Wire diameter is 1.5 mm.

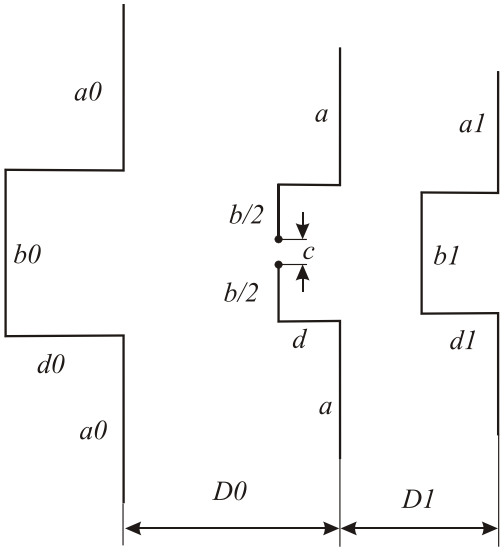


Fig. 3. Proposed antenna design

Table 1. Antenna dimensions

Design parameters, mm					
a	***b***	***c***	***d***	***a0***	***b0***
23.35	22.58	2	7.3	25.92	20.24
d0	***a1***	***b1***	***d1***	***D0***	***D1***
11.32	20.16	17.74	10.84	33.22	22

The first order of Minkowski fractal is used, since as was already written higher orders do not improve the antenna gain, but lead to decrease in the usable bandwidth. The new in the design are the each of three antenna elements is shaped as first iteration of Minkowski curve and feeding is in the middle of the central element of the driving element. Also each part of the fractal curve is with different length which allows better antenna performance. For optimization and simulation of proposed antenna is used 4nec2 software [11] which is based on Moment Method.

Return losses of the Modified Koch Fractal Yagi Antenna are displayed on Fig 4. It shows a distinctive minimum (i .e. the structure is tuned to) at 1818 MHz with a bandwidth of 12,2% (217 MHz) in compare with 11,1 % for conventional Yagi antenna. Proposed antenna has also second frequency band with return losses lower than -10 dB – from 5140 to 5230 MHz but it is not in interest for this design.

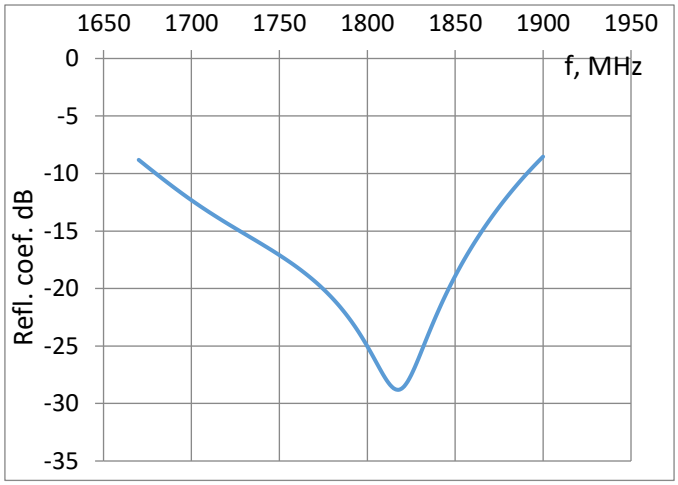


Fig. 4. Return losses

Antenna gain and F/B ratio for simulated model are given in Fig. 5 and Fig. 6 respectively. The antenna gain is higher than 6.48 dBi for the whole operating band which is on par with classical Yagi antenna with straight elements radiating in free space. The F/B ratio for the frequency band from 1710 to 1880 MHz (GSM 1800) is higher than 12.97 dB.

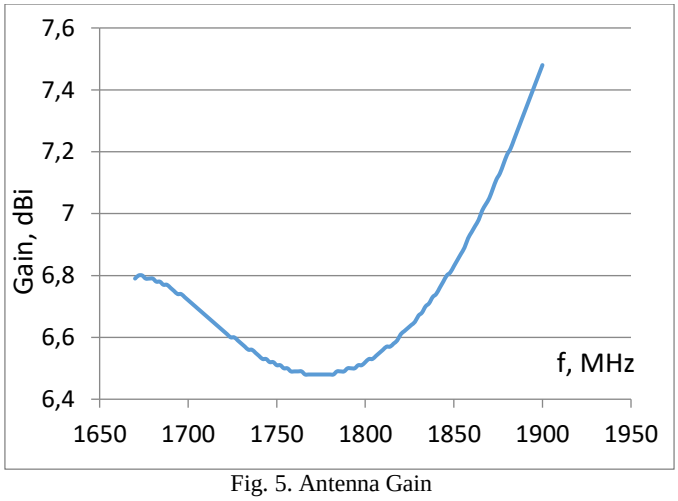


Fig. 5. Antenna Gain

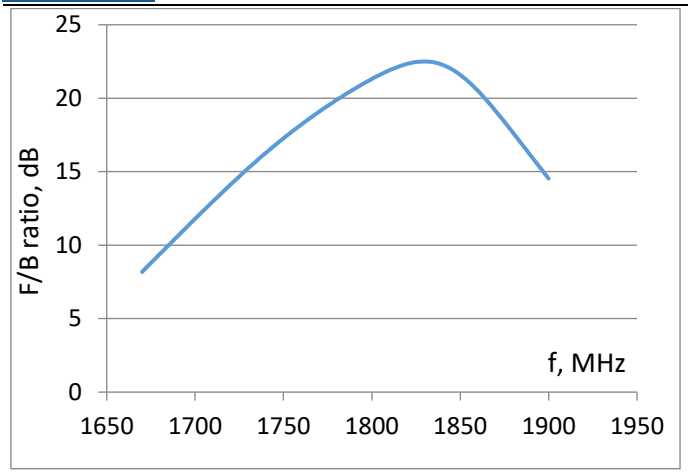


Fig. 6. Front-to-back ratio

The radiation patterns in H and V plane (antenna elements are oriented horizontaly) for frequencies 1710 MHz, 1800 MHz and 1880 MHz are shown on Fig.7, Fig. 8 and Fig. 9 respectively. For these frequencies antenna has a pattern similar to a classical Yagi antenna with gain of 6,68 dBi, 6,52 dBi and 7,19 dBi for three frequencies (6,9 dBi for conventional one).

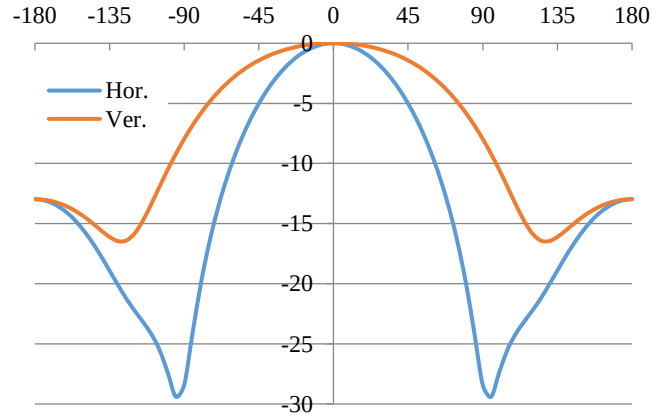


Fig. 7. Radiation pattern for 1710 MHz

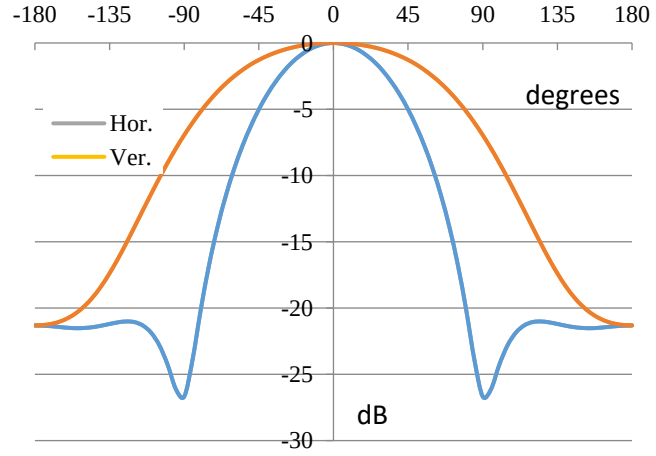


Fig. 8. Radiation pattern for 1800 MHz

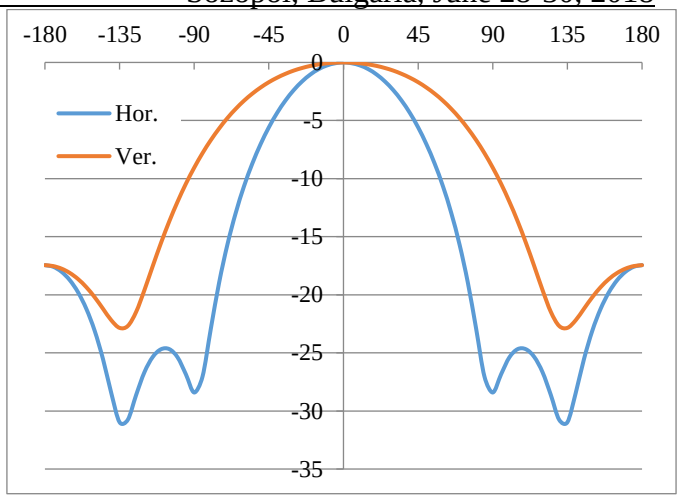


Fig. 9. Radiation pattern for 1880 MHz

The beamwidth at half power in horizontal plane is around ±35° and in vertical plane is around ±60° for these three frequencies. Front-to-Back ratio are 12.97 dB, 21.31 dB and 17,63 dB respectively.

IV. Conclusion

A modified fractal Yagi antenna with elements shaped as Minkowski fractal curve has been developed and examined in this study. It shows typical Yagi antenna and fractal antenna features with wideband performance enhanced by new element in the design – different fractal parts length. Modified Fractal Antenna has very good electrical parameters and compact size, which makes it also a candidate for practical applications. It can be used for point-to-point and point-to-multipoint applications as well (GSM 1800, LTE). The proposed antenna will be used as base for future development and design of antennas with wider bandwidth performance by using of additional fractal modifications.

Acknowledgement

This work was supported in part by the Grant DN07/19/15.12.2016 “Methods for Estimation and Optimization of Electromagnetic Radiation in Urban Areas” of the Bulgarian Science Fund.

References

[1] 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.

[2] J. P. Gianvittorio, Y. Rahmat-Samii, "Design, Simulation, Fabrication and Measurement of a Multiband Fractal Yagi Antenna", IEEE Topical Conference on Wireless Communication Technology 2003, pp. 265 - 266, October 2003.

[3] H. X. de Araujo, S. E. Barbin, L. C. Kretly, "Design of an UHF Quasi-Yagi Antenna with Metamaterial Structures for RFID Application", 2011 SBMO/IEEE MTT-S International Microwave & Optoelectronics Conference (IMOC), pp. 8-11, October-November 2011.

[4] P. R. Prajapati, G. G. K. Murthy, A. Patnaik, M. V. Kartikeyan, "Asymmetrical plus shaped fractal slotted multilayered Yagi-Uda circularly polarized microstrip antenna with DGS", 2013 IEEE Applied Electromagnetics Conference (AEMC), Bhubaneswar, India, pp. 1-2, December 2013.

[5] P. Petkov, B. Bonev, "Analysis of a Modified Sierpinski Gasket Antenna for Wi-Fi Aplications", Proc. of 24th International Conference RADIOELEKTRONIKA'2014, Bratislava, Slovak Republic, pp. 1-3, April 2014.

[6] K. Angelov, P. Petkov, B. Bonev, "Analysis of Fractal Designed Antenna Systems", Proc. of TELEKOM'2008, Varna, Bulgaria, pp. 175-180, October 2008.

[7] B. Bonev, P. Z. Petkov, "Fractal J-pole antenna," 2016 26th International Conference Radioelektronika (RADIOELEKTRONIKA), Kosice, 2016, pp. 423-426.

[8] N. Sharma, V. Sharma, "A Journey of Antenna from Dipole to Fractal: A Review", Journal of Engineering Technology, July 2017, Vol. 6, Issue 2, pp. 317-351.

[9] V. Slyusar, "Fractal Antennas", Radioamator Vol.9, 2002, pp. 54-56.

[10] V. Slyusar, "Fractal antennas: Principal new type of "broken" antennas", Elektronika: Nauka, Tehnologia, Biznes, Vol. 6, 2007, pp. 82-89. (in Russian)

[11] www.qsl.net/4nec2/

Linearization of Harmonic Radar for Detection and Classification of Traffic Participants

Aleksandra Đorić[1], Nataša Maleš-Ilić[2], Aleksandar Atanasković[2] and Predrag Eferica[2]

Abstract – **This paper represents a new automotive radar concept based on a nonlinear target detection and classification. The target carries the nonlinear tag that generates the nonlinear products of the transmitter signals - the second harmonics and intermodulation products of the third-order, which are detected by the harmonic radar receivers. The cars are identified by reflection of the second harmonics of the radar transmitter signals - two LFM modulated carriers, whereas the vulnerable targets, such as pedestrians, children and similar, are detected by reflection of the third-order intermodulation products. As the power amplifier in the radar transmitter, due to its nonlinear transfer characteristic, generates the intermodulation products it is linearized by the linearization technique that exploits the second harmonics of the transmitted signals.**

Keywords – **Automotive harmonic radar, Vulnerable targets, Nonlinear passive tag, LFM signal.**

I. INTRODUCTION

The modern automotive industry has a great need for development of the automotive applications such as pedestrian safety systems, collision warning, turning-off and lane changing assistance. The automotive radar system requires the simultaneous and precise targets localization with high accuracy and resolution and also their classification to distinguish between the different object classes such as the automobiles and vulnerable road users (VRU) for their better protection.

The design of the harmonic car radar includes the passive reflector (tag) mounted on the back of the vehicles preceding that returns the second harmonic of the frequency transmitted from the vehicle behind [1]. This concept aims to avoid rear/end collisions and it is immune to the clutter reflection from the different objects and obstacles, and also it protects from blinding by the cars that travel in the opposite direction.

Significant effort has been devoted to the development of the harmonic radars for various applications. The second harmonic radar that emits signal at one frequency and passive tags reflecting signal at the emitted signal second harmonics, have been used for insect and animal searching [2], remote sensing and detection of vital signs in medical application [3], as well as target detection in ultrawideband UWB systems [4]. The harmonic radars that transmit signals at two or more carrier frequencies have been applied for classification and localization of the VRUs [5], detection of RF electronic component and devices [6] and in RFID system for detection and accurate localization by using the harmonic tags [7].

The VRUs localization system consists of a harmonic car radar that transmits two distinct frequency carriers (f_1 and f_2) modulated by the LFM baseband signal. The targets are carrying the nonlinear tag that generates the nonlinear intermodulation products of the transmitted signals that may be transmitted back to the radar. The radar system processes the reflections and separates the conventional targets from the VRUs.

However, when the power amplifier in the radar transmitter amplifies two or more modulated carriers, due to its nonlinear characteristic it generates the intermodulation products at the amplifier output that cause spectral regrowth with the useful fundamental signals. The various linearization techniques for minimizing the third- and fifth-order distortions of the power amplifiers have been reported in the literature [8]: feedback, feed-forward, predistortion, etc.

In the previous work, authors of this paper have deployed the linearization technique that uses the second- and fourth-order nonlinear signals (IM2, IM4) of the fundamental signals [9-12] and validated its effects on the several single stage RF and Doherty PA configurations throughout the simulation process and experiments. Very good results are achieved in the reduction of the third- and fifth-order intermodulation products of the amplifiers for multitone signals as well as for digitally modulated QAM, WCDMA and OFDM signals.

In this paper, the reduction of the third-order intermodulation products in the harmonic radar transmitter is accomplished by application of the linearization method that uses the second harmonics (IM2 signals) of the fundamental LFM modulated carriers.

In Section II the new harmonic radar concept is described and the theoretical background of the applied linearization approach is explained. The Section III relates to the results of the IM3 and IM5 products suppression obtained by the proposed linearization approach. Then, in the Section IV, some conclusions are reported.

II. HARMONIC RADAR CONCEPT

Figure 1 shows the new harmonic radar concept that includes the transmitter at the car front end that emits two LFM modulated carriers at frequencies f_1 and f_2. A linear frequency modulated (LFM) waveform is the simplest form of frequency modulation, commonly used in the automotive industry, where the frequency is repeatedly swept linearly in

[1]Aleksandra Đorić is with the Innovation Centre of Advanced Technologies Ltd. Niš-Crveni krst, Serbia, Bulevar Nikole Tesle 61/5, 18000 Niš, Serbia, E-mail: alexdjoric@yahoo.com.

[2]Aleksandar Atanasković, Nataša Maleš-Ilić and Predrag Eferica, are with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mails: [aleksandar.atanaskovic; natasa.males.ilic; efa]@elfak.ni.ac.rs.

frequency range, f_{sw}, (each frequency sweep is commonly known as a chirp, T_{chirp}). The transmitted signals travel to the harmonic tags (transponders, sensors, reflectors) that consist of the nonlinear element, receiving and transmitting antennas. The nonlinear element generates the nonlinear signals such as harmonics and intermodulation products of the transmitted signals. The adequate design of the transmitting antenna on the nonlinear tag provides reflection of the second harmonic or intermodulation products of the receiving signals. The vulnerable targets may carry the intermodulation nonlinear tag whereas the second harmonic tag may be installed on the back side of the car.

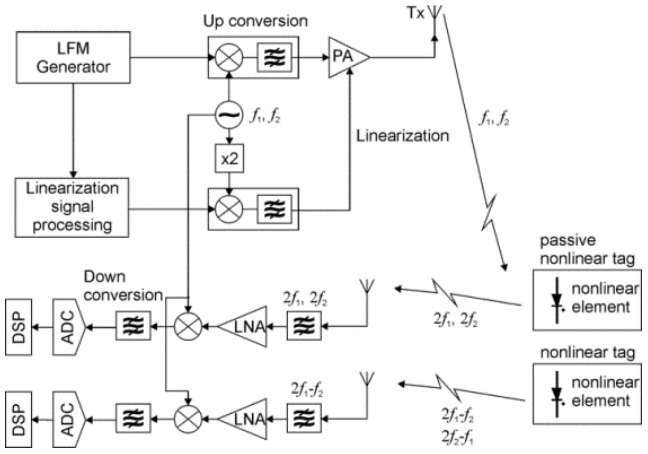


Fig. 1. Harmonic radar concept for target classification and range and velocity determination

The power amplifier in the radar transmitter is a nonlinear device and produces the nonlinear products - harmonics and intermodulation products at frequencies $2f_1 - f_2$ and $2f_2 - f_1$ that are close to the fundamental emitted frequencies f_1 and f_2. The signal for vulnerable target detection and identification appears at the intermodulation frequencies and can interfere with the intermodulation products generated at the transmitter due to the nonlinear transfer characteristic of the power amplifier. Accordingly, in this paper, the transmitter intermodulation products are suppressed by the linearization technique that utilizes the second harmonics of the transmitted LFM modulated fundamental carrier [9-12]. The adequate baseband signals for the linearization are up converted at the frequencies of the second harmonics of the fundamental signal carriers, adjusted in amplitude and phase and inserted at the input and output of the PA transistor. The frequencies of the local oscillators, necessary for the down conversion at the second harmonic receiver, $2f_1$ and $2f_2$, are generated at the transmitter for the linearization purposes.

Two identical LFM signals are upconverted to the second harmonics of the two fundamental carriers and sent to the two types of targets. One target type is normal (does not carry nonlinear tags) while the other type carries the harmonic tags. The nonlinear element in the tag produces the harmonic and intermodulation signals, whereas the reflecting tag antenna determines the frequency of the reflected signal: at the carriers' second harmonics or at the intermodulation products of the third-order. The reflected signals from the targets are detected at the radar receivers that select the certain channel by the appropriate bandpass filter. The target range and velocity may be determined according to the equations from the procedure of measuring beat frequencies and applying FFT according to the algorithm depicted in literature, [4], [13].

The operating principle of the proposed linearization approach could be described by a Taylor-series polynomial model, [14], in case when the memory effects are neglected. The linear frequency modulated signal can be represented in the form:

$$V_{LFM}(t) = A\cos\left(\omega_o t + \frac{f_{sw}}{2T_{chirp}}t^2\right) = A\cos(\phi(t)) \tag{1}$$

Complex version of the signals is:

$$I + jQ = A(\cos(\phi(t)) + j\sin(\phi(t))) \tag{2}$$

where $I = A_I\cos(\phi(t))$ and $Q = A_Q\sin(\phi(t))$ are the in-phase and quadrature-phase components of the baseband signal.
The complex envelope of the LFM triangular signal characterized with the $T_{chirp} = 1\,\mu s$ is created in ADS to modulate the fundamental carriers at the frequencies f_1 and f_2. The baseband signal required for the linearization is formed as squared of the complex envelope of the fundamental LFM signal:

$$v_{lin\_IM2} = (I + jQ)^2 = \left[\left(I^2 - Q^2\right) + j2IQ\right] \tag{3}$$

The baseband signal for the linearization modulates the second harmonics of the fundamental carriers at $2f_1$ and $2f_2$. After modulating, the signals are combined and adjusted on the appropriate amplitude and phase in the analogue RF domain throughout the two linearization branches. The signals for the linearization set on the appropriate values are inserted at the input and output of the power amplifier transistor over the bandpass filters centered at $f_c = \frac{f_1 + f_2}{2}$.

The signals injected at the amplifier transistor gate - the fundamental useful signals and the signals for linearization adjusted in amplitude and phase (a_{g2h}, θ_{g2h}) are given by:

$$\begin{aligned} v_{gs}(t) = \sum_{i=1}^{2}\{&v_{si}\left[I\cos(\omega_i t) - Q\sin(\omega_i t)\right] + \\ &a_{g2h}e^{-j\theta_{g2h}}\left[\left(I^2 - Q^2\right)\cos(2\omega_i t) - 2IQ\sin(2\omega_i t)\right]\} \end{aligned} \tag{4}$$

The signal at the drain of the transistor consists of the fundamental signal gained linearly (the first term in eq. (5)), and the signals for linearization tuned in amplitude and phase (a_{d2h}, θ_{d2h}) and injected at the amplifier transistor drain.

$$\begin{aligned} v_{ds}(t) = \sum_{i=1}^{2}\{&v_{oi}\left[I\cos(\omega_i t) - Q\sin(\omega_i t)\right] + \\ &-a_{d2h}e^{-j\theta_{d2h}}\left[\left(I^2 - Q^2\right)\cos(2\omega_i t) - 2IQ\sin(2\omega_i t)\right]\} \end{aligned} \tag{5}$$

where ($v_{oi}|I\cos(\omega_i t)-Q\sin(\omega_i t)|$, i=1,2 is the output signal at the fundamental frequencies.

The drain current at the frequencies of the third-order intermodulation products $2\omega_1-\omega_2$ and the fifth-order $3\omega_1-2\omega_2$, are represented by the equations (6) and (7), respectively. We supposed $v_{s1}=v_{s2}=v_s$ and $v_{o1}=v_{o2}=v_o$ to simplify the equations. Another the IM3 and IM5 products, at frequencies $2\omega_2-\omega_1$ and $3\omega_2-2\omega_1$, have the same form.

$$i_{ds}(t)\big|_{(2\omega_1-\omega_2)}=\left(\frac{3}{4}v_s^3 g_{m3}+a_{g2h}e^{-j\theta_{g2h}}v_s g_{m2}+\right.$$
$$-\frac{1}{2}a_{d2h}e^{-j\theta_{d2h}}v_s g_{m1d1}+\frac{1}{2}a_{g2h}e^{-j\theta_{g2h}}v_o g_{m1d1}$$
$$\left.+\frac{3}{2}v_s v_o^2 g_{m1d2}+\frac{3}{2}v_s^2 v_o g_{m2d1}\right)$$
$$\left(I^2+Q^2\right)\left(I\cos\left(\left(2\omega_1-\omega_2\right)t\right)-Q\sin\left(\left(2\omega_1-\omega_2\right)t\right)\right) \qquad (6)$$

$$i_{ds}(t)\big|_{(3\omega_1-2\omega_2)}=\left(\frac{5}{8}v_s^5 g_{m5}+\frac{3}{2}a_{g2h}^2 e^{-j2\theta_{g2h}}v_s g_{m3}+\right.$$
$$+\frac{1}{2}a_{d2h}^2 e^{-j2\theta_{d2h}}v_s g_{m1d2}-a_{g2h}a_{d2h}e^{-j\left(\theta_{g2h}+\theta_{d2h}\right)}v_o g_{m1d2}+$$
$$\left.+\frac{1}{2}a_{g2h}^2 e^{-j2\theta_{g2h}}v_o g_{m2d1}-a_{g2h}a_{d2h}e^{-j\left(\theta_{g2h}+\theta_{d2h}\right)}v_s g_{m2d1}\right)$$
$$\left(I^2+Q^2\right)^2\left(I\cos\left(\left(3\omega_1-2\omega_2\right)t\right)-Q\sin\left(\left(3\omega_1-2\omega_2\right)t\right)\right) \qquad (7)$$

According to the analysis carried out in the previous research [9-12], the third-order distortion of the fundamental signal is dominantly caused by the cubic transconductance (g_{m3}) term in the polynomial transistor model and can be reduced by selecting the adequate amplitude and phase of the second harmonic modulated by the appropriately modified baseband signal for the linearization, which are injected at the input and output of the amplifier transistor.

III. LINEARIZATION RESULTS

In the automotive industry different frequency bands have been used for applications whereas the 24 GHz and 77 GHz bands are dominant [13]. In this paper, we have analyzed the proposed harmonic radar concept operating at much lower frequency than standardized. The main reason for that is the intention to test the linearization effects of the technique that exploits the carrier second harmonics in case of the LFM continuous wave signal, which is utilized standardly in automotive radars for range and velocity determination. For the application and testing purposes of the proposed linearization technique [9-12], we have exploited previously designed power amplifier that operates in the frequency range 0.7 to 1.1 GHz [12]. The amplifier was designed in Agilent Advanced Design System-ADS based on the nonlinear MET model of the Freescale transistor MRF281S LDMOSFET. The maximum achieved gain is around 22 dB at 1 GHz, whereas the maximum power added efficiency (PAE) for the maximum output power of around 36 dBm is 50% at 1 GHz.

Two identical LFM signals characterized by the frequency bandwidth $f_{sw}=50\,\text{MHz}$ are upconverted by RF carriers $f_1=0.95\,\text{GHz}$ and $f_2=1.05\,\text{GHz}$ for sending to the targets that may be normal (does not carry nonlinear tags), and targets that carry harmonic tags. The nonlinear elements produce the harmonic and the intermodulation signals. The tag antenna operating frequency band enables that the reflected signals are at the carriers' second harmonics $2f_1=1.9\,\text{GHz}$ and $2f_2=2.1\,\text{GHz}$ or at the intermodulation products of the third-order $2f_1-f_2=0.85\,\text{GHz}$ and $2f_2-f_1=1.15\,\text{GHz}$.

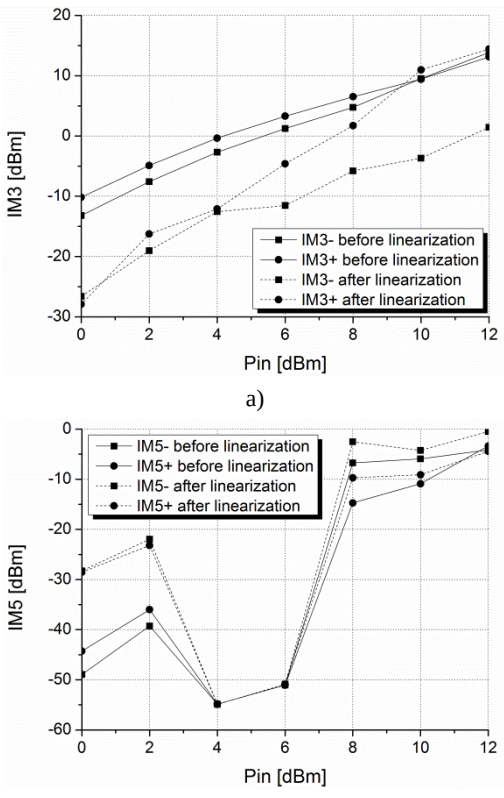


Fig. 2. The intermodulation products of the power amplifier at the transmitter of the automotive radar for the LFM signal in terms of input power levels before and after the linearization: a) third-order; b) fifth-order

The analysis has been carried out for different fundamental signal power levels at the amplifier input from 0 dBm to 12 dBm. The simulation results represented in Figure 2a show that the IM3 products are lessened more than 15 dB in the case of 0 dBm input power, whereas for the rise of the input power level the suppression rate is falling down to 10 dB up to approximately 6 dBm input power. In case of the higher power levels the linearization cannot be attained. It should point out that 6 dBm input power enables the output power of 26 dBm per carrier, which is around 3 dB below 1-dB compression point of the amplifier, [12]. Moreover, including the fact that the RF carriers are shifted by 100 MHz (25% of theamplifier bandwidth) the achieved linearization results are considered to be satisfied.

The simulated results of the performed linearization approach effect on the fifth-order intermodulation products, IM5, presented in Figure 2b, indicate that the IM5 products are slightly worsen in reference to the state before the linearization for almost all considered input power levels. The significant impairment is observed at lower power level from 0 dBm to 3 dBm. It should stress that the linearization has been performed, i.e. the optimization of the amplitude and phase of the signals for linearization has been carried out with the aim to suppress the third-order intermodulation products and to restrain the fifth-order intermodulation products at the levels below the reduced IM3 products, which is accomplished in entire power range.

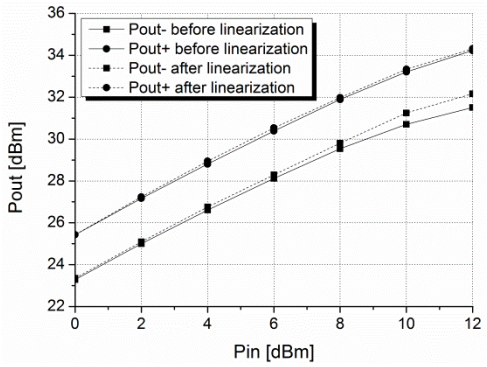


Fig. 3. Output power of the power amplifier in automotive radar transmitter for two RF carriers modulated by LFM signal

Additionally, the output power of the RF carriers is also enhanced when the proposed linearization approach is applied, which is illustrates in Figure 3.

IV. Conclusion

A new automotive radar concept for target detection and classification based on the idea of the harmonic radar is presented in this paper. The system generates the nonlinear products of the transmitted signals, such as the second harmonics and intermodulation products of the third-order, because the targets are carrying the nonlinear tag. The cars may be identified by reflection of the second harmonics of the two radar transmitter carriers modulated by the LFM baseband signal, whereas the vulnerable targets, such as pedestrians, children and similar will be identify by reflection of the third-order intermodulation products. The power amplifier in the radar transmitter amplifies the two carriers and, due to its nonlinear transfer characteristic, generates the intermodulation products that may interfere with the reflected products from the VRU targets. Accordingly, it is necessary to suppress the amplifier third-order intermodulation products and to retain the fifth-order distortions to the appropriate power level below the third-order. The reduction of the intermodulation products has been obtained by using the linearization technique that exploits the second harmonics of the fundamental signal carriers. The satisfactory reduction of the IM3 products is achieved and IM5 products are kept at the power level below the suppressed IM3 products even though two transmitter carriers are shifted in frequency 100 MHz (25% of the amplifier bandwidth). Also, the output power of the fundamental signals is increased for a certain percent of decibels.

Acknowledgement

This work was supported by the Ministry of Education, Science and Technological Development of Republic of Serbia, the project number TR-32052.

References

[1] J. Shefer, R. J. Klensch, G. Kaplan, H. C. Johnson, “Clutter-Free Radar for Cars”, Wireless World, pp. 117-122, May 1974.

[2] Z. Tsai, P. Jau, N. Kuo, J. Kao, K. Lin, F. Chang, E. Yang, H. Wang, “A High-Range-Accuracy and High-Sensitivity Harmonic Radar using Pulse Pseudorandom Code for Bee Searching”, IEEE Trans., Microwave Theory Tech., vol. 61, no. 1, pp. 666–675, 2013.

[3] L. Chioukh, H. Boutayeb, D. Deslandes, K. Wu, “Noise and Sensitivity of Harmonic Radar Architecture for Remote Sensing and Detection of Vital Signs”, IEEE Trans., Microwave Theory Tech, vol. 62, no. 9, pp. 1847-1855, September 2014.

[4] K. Gallagher, Harmonic Radar: Theory and Applications to Nonlinear Target Detection, Tracking, Imaging and Classification, Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy, The Pennsylvania State University, December 2015.

[5] J. Saebboe, V. Viikari, T. Varpula, H. Seppä, S. Cheng, M. Al-Nuaimi, P. Hallbjörner, A. Rydberg, “Harmonic Automotive Radar for VRU Classification”, Radar Conference Surveillance for a Safer World, Conference Proceedings, October 2009.

[6] G. Mazzaro, A. Martone, D. Mcnamara, “Detection of RF Electronics by Multitone Harmonic Radar”, IEEE Trans., Aerospace and Electronic Syst., vol. 50, no. 1 January 2014.

[7] D. Dardari, “Detection and Accurate Localization of Harmonic Chipless Tags”, EURASIP Journal on Advances in Signal Processing, 2015.

[8] S. Cripps, *RF Power Amplifiers for Wireless Communications*, Chapters 4-5, pp. 115–251, Artech House, 1999.

[9] N. Males-Ilić, B. Milovanović, Đ. Budimir, “Effective Linearization Technique for Amplifiers Operating Close to Saturation”, International Journal of RF and Microwave Computer-Aided Engineering, vol. 17, no. 2, pp. 169-78, 2007.

[10] A. Atanasković, N. Maleš-Ilić, B. Milovanović, “Linearization of Power Amplifiers by Second Harmonics and Fourth-order Nonlinear Signals”, Microwave and Optical Technology Letters, vol. 55, issue 2, pp. 425-430, February 2013.

[11] N. Maleš-Ilić, A. Atanasković, K. Blau, M. Hein, “Linearization of Asymmetrical Doherty Amplifier by the Even-Order Nonlinear Signals”, International Journal of Electronics, Taylor & Francis, vol. 103, issue 8, pp. 1318-1331, 2015.

[12] A. Đorić, N. Maleš-Ilić, A. Atanasković, “RF PA Linearization by Signals Modified in Baseband Digital Domen”, Facta Universitatis, Series: Electronics and Energetics, vol. 30, no. 2, pp. 209-221, 2017.

[13] J. Hasch, E. Topak, R. Schnabel, T. Zwick, R. Weigel, C. Waldschmidt, “Millimeter-Wave Technology for Automotive Radar Sensors in the 77 GHz Frequency Band”, IEEE Trans., Microwave Theory Tech., vol. 60, no. 3, pp. 845-860, March 2012.

[14] J. C. Pedro, 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.

Studying the short channel effect in MOSFET Transistor Amplifiers

Stoycho Manev

Abstract – **The presented paper is aimed toward educational-methodological purposes. The short channel effects in MOSFET Transistors are investigated. PSPICE simulation of circuits, containing MOSFET Transistors are performed.**

Keywords – **short channel effect, MOSFET Transistors, Analog circuits**

I. INTRODUCTION

The presented paper is aimed toward educational-methodological purposes. The contemporary high level of miniaturization requires that by the electronic design all possible factors should be considered. The short channel effects are very widespread in MOSFET devices. They can be summarized in the following directions[1]: channel length modulation; threshold voltage roll-off; narrow gate width effects; reverse short channel effects; punch through; mobility degradation; velocity saturation[1].

II. THEORETICAL BACKGROUND

A. Classification of the MOSFET Transistors

In Fig.1 is presented the perspective projection of MOSFET Transistor[2].

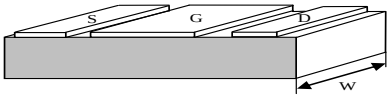


Fig. 1. MOSFET Transistor (perspective projection)

In Fig.2, is shown a long-channel MOSFET Transistor (cross section)[2].

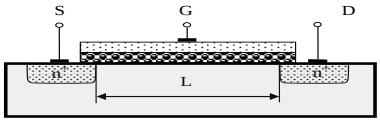


Fig. 2. Long-channel MOSFET Transistor(cross section)

If the channel length L >> sum of the depletion widths of the drain and the source, as shown on Fig.2, the MOSFET Transistor is classified as long-channel [1],[2].

Contrariwise, if L is comparable with the sum of the source and drain depletion widths, as presented on Fig.3 the MOSFET Transistor is classified as short-channel [1],[2].

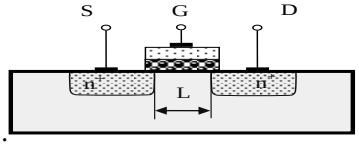


Fig. 3.Short-channel MOSFET Transistor(cross section)

B. Basic mathematical equations[3]

[1]Stoycho Manev is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria EMAIL: smanev@tu-sofia.bg

a) In the triode area the equation (1) is valid [3]:

$$I_{D0} = \mu_n . C_{ox} . \frac{W}{L} . \left[(V_{GS} - V_{th0}) . V_{DS} - \frac{V_{DS}^2}{2} \right], \quad (1)$$

where [3],[1]: I_{D0} designates the value of the drain current, ignoring the effect of the channel length modulation; C_{ox} is the gate oxide capacitance per unit area; V_{th0} is the threshold voltage; μ_n is the field effect mobility of electron[1]; V_{DS} is the voltage drain-source; V_{GS} is the voltage gate-source.

b) In the saturation area [3]:

$$I_{D0} = \frac{\mu_n . C_{ox}}{2} . \frac{W}{L} . (V_{GS} - V_{th0})^2 \quad (2)$$

The transconductance can be calculated by means of the relation (3)[3]:

$$g_m = \mu_n . C_{ox} . \frac{W}{L} . (V_{GS} - V_{th0}) = \sqrt{2 . \mu_n . C_{ox} . \frac{W}{L} . I_{D0}} \quad (3)$$

C. Short channel effects[3],[2]:

* Channel length modulation [3]

By channel length modulation the equations (1) and (2), concerning the drain current in triode area and in the saturation area, can be presented as follows[3]:

$$I_D = (1 + \lambda . V_{DS}) . I_{D0} \quad , \quad (4)$$

where λ is semi empirical constant.

* Threshold voltage roll off [2]

The threshold voltage of short-channel N - MOSFET Transistor is less than the corresponding threshold voltage of the long - channel MOSFET Transistor. It can be calculated by means of the relation [2]:

$$V_{th0\ short\ channel} = V_{th0\ \ long\ channel} - \Delta V_{th0}, \quad (5)$$

where:

$$\Delta V_{th0} = \frac{1}{C_{ox}} . \sqrt{2 . q . \varepsilon_{Si} . N_A |-2 . \phi_f|} . \frac{x_j}{2L} . \left[\left(\sqrt{1 + \frac{2 . x_{dS}}{x_j}} - 1 \right) + \left(\sqrt{1 + \frac{2 . x_{dD}}{x_j}} - 1 \right) \right] [2] \quad (6)$$

Here [2]: x_{dS} is the depth of the depletion region at source; x_{dD} is the depth of the depletion region at drain; N_A is the substrate doping density; ϕ_f is the substrate Fermi potential; ε_{Si} is the dielectric constant of silicon; q is the electron charge; C_{ox} is the gate oxide capacitance per unit area; x_j is the junction depth. The depletion regions are detailed presented on Fig.4[2]:

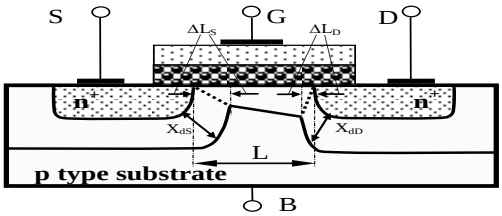


Fig. 4.MOSFET Transistor structure (cross section)

* Punch through[3]

When the depletion region at source X_{dS} and the depletion region at drain X_{dD} merge there is a punch through. The drain current cannot be driven by gate voltage.

* Velocity saturation[3]

In order to realize ever smaller in size devices modern technologies have been applied. In this case high-field effects must be taken into account. In this analysis essential is the effect of velocity saturation. In silicon the electron drift cease to depend from the applied electric field at value approximately 10^6 V/m and saturates at a value of approx. 10^5 m/s. The result, obtained from the expression: $\frac{V_{GS}-V_{th0}}{L}$ plays an important role by the short-channel effects. In case, that $\frac{V_{GS}-V_{th0}}{L}$ approaches E_{sat}, where E_{sat} is typical approximately 4.10^6V/m, I_D is not more dependent from L. It's value becomes:

$$I_D = \frac{\mu_n . C_{ox}}{2} . W . (V_{GS} - V_{th0}) . E_{sat} \text{ [3]} \qquad (7)$$

The velocity saturation leads to restriction of the value of the transconductance: $g_m = \mu_n . C_{ox} . \frac{W}{2} . E_{sat}$ [3] (8)

For its part the transit frequency ω_T can be expressed as follows: $\omega_T \approx \frac{3}{4} . \frac{\mu_n . E_{sat}}{L}$ [3] (9)

The reverse short-channel effects are related to the doping profile of the analyzed MOSFET device[3]

* Narrow channel width effects[2]

Considerations about the channel width: the analysis is similar to the analysis, concerning the channel length. If the channel width W is comparable with the maximum depletion regions thickness into the substrate (x_{dm}), the MOSFET Transistor can be classified as narrow-channel. The narrow channel effect leads to increase of the threshold voltage [2]:

$$V_{th0\ narrow\ \ channel} = V_{th0\ long\ channel} + \Delta V_{th0} \qquad (10)$$

In case that the shape of the depletion regions edges are modeled by means of quarter circular arcs the following expression is valid [2]:

$$\Delta V_{th0} = \frac{1}{C_{ox}} . \sqrt{2.q.\varepsilon_{Si}.N_A \left| -2.\phi_f \right|} . \frac{\pi . x_{dm}}{2.W} \qquad (11)$$

* Mobility degradation[3]

The increase of the gate potential leads to decrease of the mobility of the charge carriers. This effect can be quantitative evaluated by means of the normal field mobility degradation factor θ. It has a typical value in the range: $(0{,}1 \div 1)V^{-1}$ and is inversely proportional to the gate oxide thickness. In order to obtain the actual value of the drain current I_D the calculated value must be multiplied by[3]: $\frac{1}{1+\theta.(V_{GS}-V_{th0})}$ (12)

III. PSPICE MODELS OF THE ANALYZED CIRCUITS

In traditionally courses the focus of the analysis of the MOSFET devices is aimed toward the basic characteristics. In this paper special attention is turned on the short channel effects in MOSFET Transistors.

The circuit, which has been used in the presented paper for AC analysis, is shown on the Figure bellow.

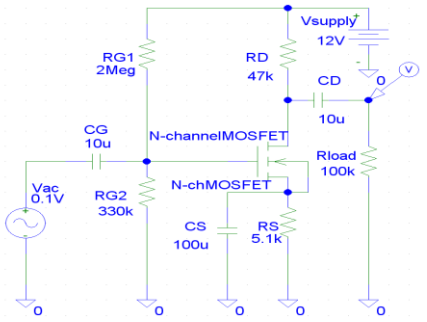


Fig. 5.PSPICE model for AC analysis

The circuit for analysis of mobility degradation and the dependence on the oxide thickness is displayed on Fig.6. In order to study the roll-off effect the circuit, presented on Fig.6, is used.

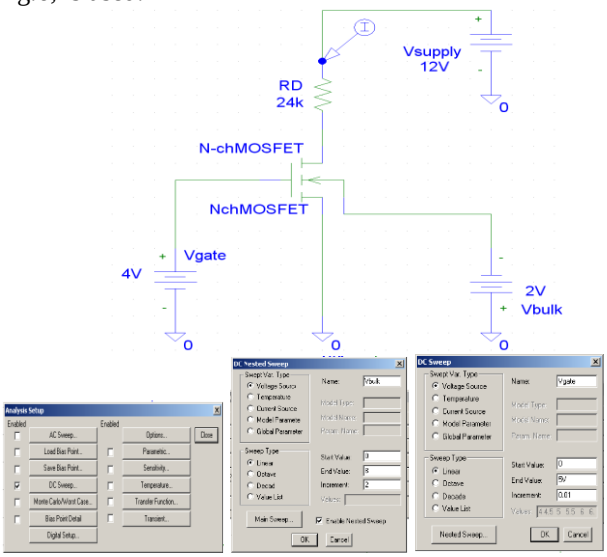


Fig. 6.PSPICE model for studying the "roll-off" effect

The PSPICE models of the transistors in these circuits are based on the Level 3 PSPICE parameters, given in [3].

The narrow gate width effects have been investigated by means of the both circuits (shown on Fig.5 and Fig.6).

The channel length modulation has been examined using the PSPICE MOSFET parameters Level 2, presented in [4].

As it can be seen from the analysis setup conditions concerning the circuit, presented on Fig.6, DC nested sweep has been activated. The gate voltage changes in the range: 0V ÷ 5V while the bulk voltage vary from 0V to 8V by step of increment 2V.

IV. IMULATION RESULTS AND ANALYZES

A Mobility degradation

The simulation experiments, concerning the AC analysis, have been done by mobility degradation factor θ (THETA) = $2{,}3.10^{-1}$, as given in [3]. The results are presented on Fig.7.

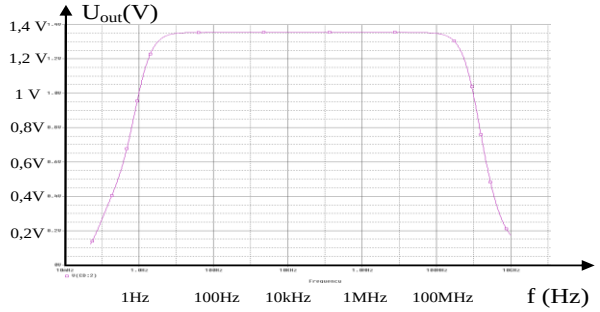


Fig.7.AC analysis by θ (THETA) = $2{,}3.10^{-1}$

The circuit, displayed on Fig.5, has been used by the simulation. Changing the value from θ = $2{,}3.10^{-1}$ to θ = 1 leads to consequent change in the AC characteristic. This is displayed on Fig.8.

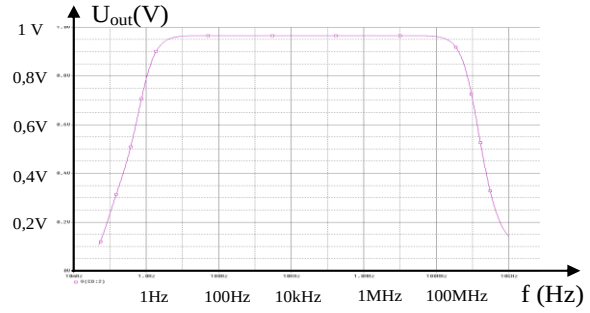


Fig.8. AC analysis by θ (THETA) = 1

As expected, taking into account the analytical expression (12), from the Figures above, becomes obvious that the output voltage is inverse proportional to the value of θ .

By the both simulations the value of the oxide thickness is $9{,}5.10^{-9}$.

B Influence of the oxide thickness

In order to establish qualitative relationship between this thickness and the AC parameters of the MOSFET amplifier another simulation has been performed.

Changing the oxide thickness from 9,5. 10^{-9} to 3.10^{-9} the following results, displayed on Fig.9, are obtained.

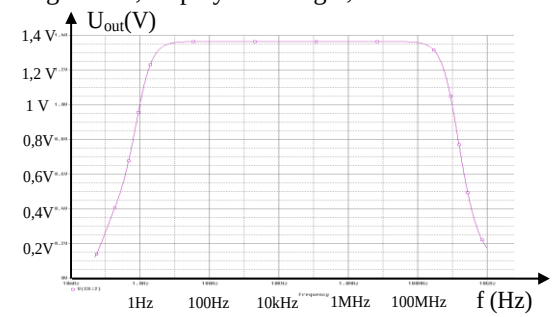


Fig.9. AC analysis by oxide thickness 3.10^{-9}, θ = $2{,}3.10^{-1}$

All simulations up to here have been done by: L = 2µm and W = 16 µm.

C Narrow channel width effects

In order to analyze the narrow gate width effects the value of W has been reduced to 8µm. The value of L remains 2 µm, θ = $2{,}3.10^{-1}$, the oxide thickness is $9{,}5.10^{-9}$. Taking into account the equations: (1), (2), (3) decrease of the output voltage is expected. This is presented on Fig.10.

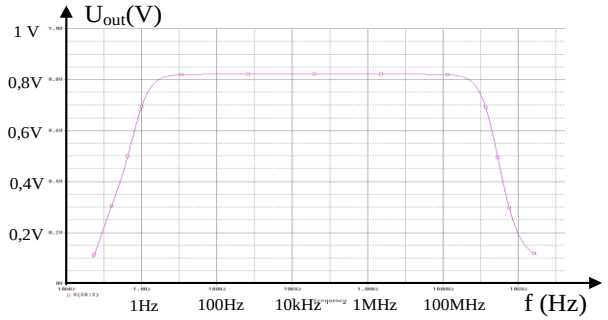


Fig.10. AC analysis by W = 8 µm

The graphical results, shown on Fig.10, confirm these expectations.

D Roll off effects

The investigations, concerning the roll-off effects, have been performed by means of the circuit, presented on Fig.6. The simulation results, done by L = 2µm, W = 16µm, are displayed on the Figure bellow.

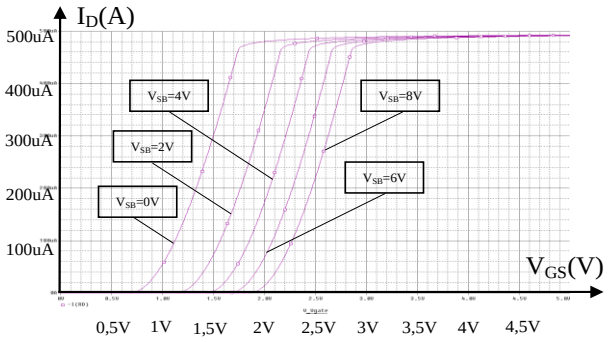


Fig.11.Characteristic I_D - V_{GATE} by L = 2µm, W = 16µm

The change of the bulk voltage toward more negative values leads to subsequent translation of the characteristic I_D - U_{GATE} to the right, as it becomes obvious from Fig.11.

a) Narrow channel width effects

According to expressions (11),(10) the reduction of width W to 8um leads analytically to increase of the threshold voltage. By initial conditions L= 2µm W = 8µm simulations have been done. The results are presented on Fig.12.

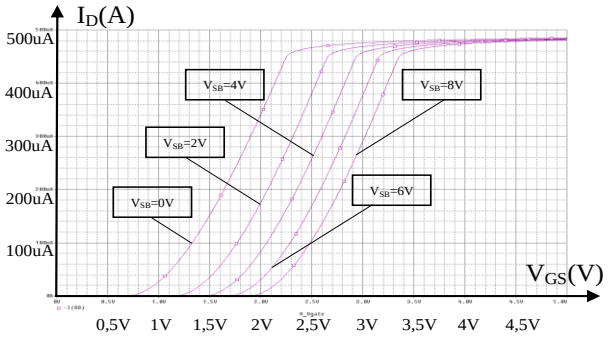


Fig.12. Characteristic I_D - V_{GS} by L = 2µm, W = 8µm

Comparing the characteristics from Fig.11 and Fig.12 the increase of the threshold voltage can distinctly be seen.

b) Short channel effects

While the width W reduction leads to increase of the threshold voltage, according to expressions (6) and (5), the supposedly decrease of the channel length L should cause increase of ΔV_{th0} and subsequent decrease of the threshold voltage $V_{th0\ short\ channel}$. Simulations using PSPICE model attributes L = 0,5µm, W = 8µm have been done. The obtained results are presented on Figure 13.

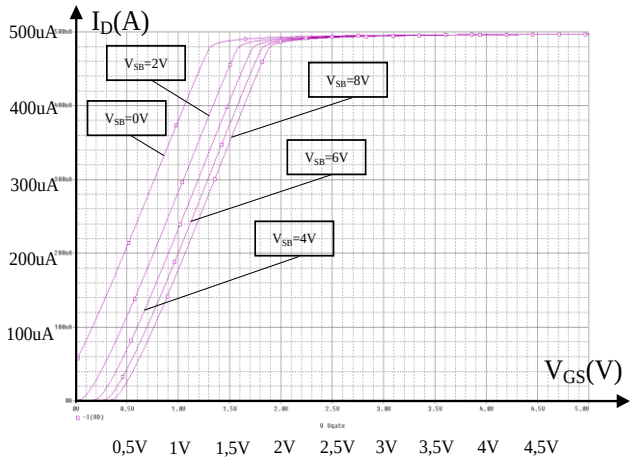


Fig.13. Characteristic I_D - V_{GS} by L = 0,5µm, W = 8µm

The translations of the characteristics to the left, compared to the characteristics, displayed on Fig.12, confirm the expected effect of decrease of the threshold voltage.

E Channel length modulation

By the simulations the PSPICE model, proposed in [4], has been used. The values of V_{SB} , L and W have been set to: V_{SB} = 2V, L = 2µm , W = 16µm .

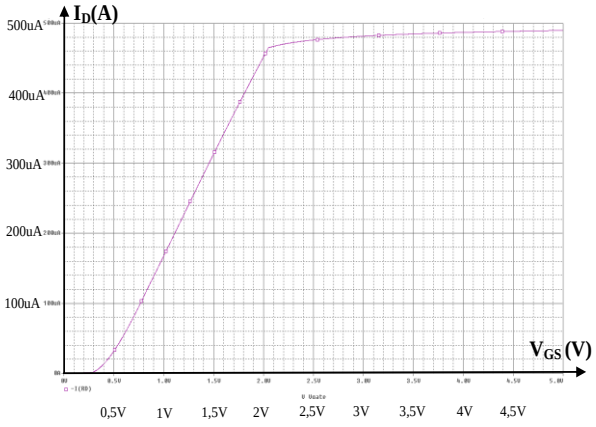


Fig.14. Characteristic I_D - V_{GS} by $\lambda = 3{,}903.10^{-2}$

The graphical results, shown above, concern the PSPICE model parameter values, given in [4]. The value of λ (LAMBDA), given there, is $3{,}903.10^{-2}$. In order to study the influence of the value of parameter λ upon the drain current, which analytically is described by means of expression (4), another value of λ has been inserted in the PSPICE model of the MOSFET Transistor. In this case λ has been set to value of 1.10^{-2}.

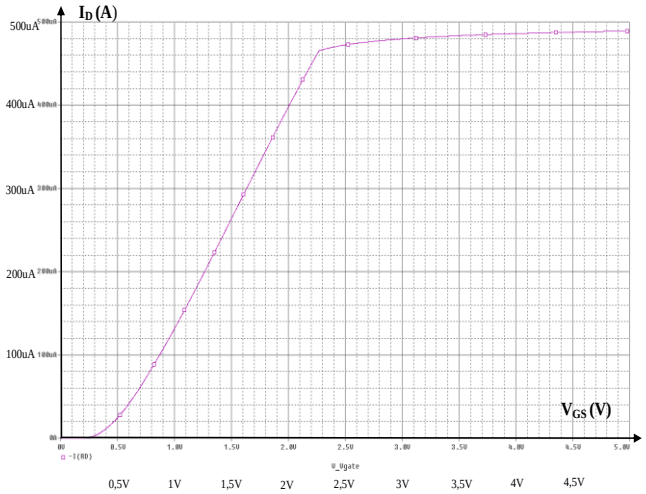


Fig.15. Characteristic I_D - V_{GS} by $\lambda = 1.10^{-2}$

An essential change in the slope of the characteristic I_D - V_{GS}, displayed on Fig.15, in comparison with the slope of the characteristic I_D - V_{GS}, presented on Fig.14, has been found out. This fact is in accordance with the analytical expression, concerning I_D, given in (4).

V. Conclusion

In this paper the short channel effects in MOSFET devices are discussed. The goal of the paper is to present a set of simulation experiments, in order a detailed examination of these effects to be performed.

References

[1] Dr. Lynn Fuller, *The Short Channel MOSFET*, Microelectronic Engineering Rochester Institute of Technology, https://people.rit.edu/lffeee/mosfet_s.pdf

[2] Milaim Zabeli, Nebi Caka, Myzafere Limani, Qamil Kabashi, *The impact of MOSFET's physical parameters on its threshold voltage*, Proceedings of the 6th WSEAS Internationalonference on Microelectronics, Nanoelectronics, Optoelectronics, Istanbul, Turkey, May 27-29, 2007

[3] Thomas H. Lee Handout #2: EE214 Fall 2002A Review of MOS Device Physics; https://web.stanford.edu/class/archive/ee/ee214/ee214.1032/Handouts/HO2.pdf

[4] Twesha Patel, *Comparison of Level 1, 2 and 3 MOSFET's* Course: Advanced Electronics,Semester: Fall 2014

[5] Prof. Wu,*Lecture17* EE105Spring 2008,UC Berkeley,http://inst.eecs.berkeley.edu/~ee105/p10/lectures/lecture17.pdf

[6] Prof.J.S.Smith,*MOS Transistor models:Body effects, SPICE models*, Lecture15, EECS105 Spring 2004; https://inst.eecs.berkeley.edu/~ee105/sp04/handouts/lectures/Lecture15.pdf

[7] SPICE MODEL PARAMETERS OF MOSFETS, https://www.seas .upenn.edu / ~jan/spice/spice.MOSparamlist.html

Studying the bulk effect in MOSFET Transistor Amplifiers

Stoycho Manev

Abstract – **In this paper the bulk effect in MOSFET Transistors is investigated. PSPICE results of simulation of the bulk effect in N-channel MOSFET Transistors are given.**

Keywords – **Bulk effect, Body effect, Back-gate effect, MOSFET Transistors**

I. INTRODUCTION

Devices with MOSFET Transistors are widely used in practice. By the design of such type of sets different effects must be taken into account. One of this effects is the Body effect. In a number of cases the source is connected to the bulk. If between the source and the bulk potential difference (V_{SB}) is available, another more complicated way of circuit design must be considered. The body effect is associated with the voltage, applied to the bulk. Due to this fact change in the parameters and the characteristics of the MOSFET transistors can be observed. The aim of the presented paper is to study the relationship between the basic transistor parameters and the value of the voltage V_{SB}.

II. THEORETICAL BACKGROUND

A Basic mathematical equations describing the Body effect in MOSFET Transistors

The structure of the N – channel MOSFET Transistor is shown on the Figure bellow [1]:

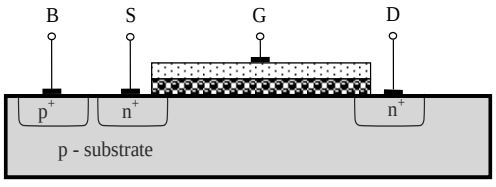


Fig. 1. N – channel MOSFET Transistor (cross section)

By MOSFET Transistors the characteristics $I_D = f(U_{GS})$ are from great importance. They are dependent from the voltage difference between bulk and source. The change of this value leads to consequent translation of the analyzed characteristics. This effect is presented on Fig.2[2].

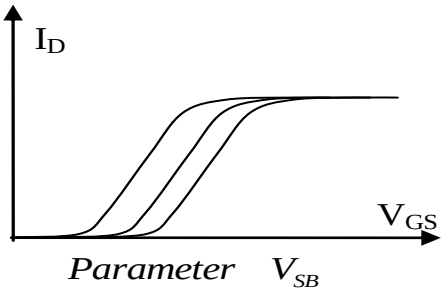


Fig. 2.Characteristic I_D -V_{GS}

It becomes obvious that the threshold voltage V_t depends on the value of V_{SB}.

a) Assuming the voltage $V_{SB} = 0$ the following expression is valid [3]:

$$V_{t0} = \Phi_{GC} - 2\varphi - \frac{Q_{B0}}{C_{ox}} - \frac{Q_{ox}}{C_{ox}} \quad (1)$$

Here : $\Phi_{GC} = \varphi_F(substrate) - \varphi_F(gate)$

φ_F is the substrate Fermi potential

$$Q_{B0} = \sqrt{-2.q.N_A.\varepsilon_{Si}|-2.\varphi_F|}$$

$$C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}}$$

C_{ox} is the gate oxide capacitance per unit area

q is the electron charge

N_A is the substrate doping concentration

ε_{Si} is the dielectric constant of silicon

b) If the voltage $V_{SB} > 0$[3]:

$$V_t = \Phi_{GC} - 2\phi - \frac{Q_B}{C_{ox}} - \frac{Q_{ox}}{C_{ox}} \quad (2)$$

$$V_t = V_{t0} + \gamma.\left(\sqrt{|-2.\phi_F + V_{SB}|} - \sqrt{|-2.\phi_F|}\right), \quad (3)$$

where: $\gamma = \frac{\sqrt{2.q.N_A.\varepsilon_{Si}}}{C_{ox}}$

Q_B is the depletion region charge density at surface inversion.

B Model of MOSFET Transistor taking into account the Body effect

The model of the analyzed MOSFET transistor is shown on the Figure bellow [4]:

[1]Stoycho Manev is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria EMAIL: smanev@tu-sofia.bg

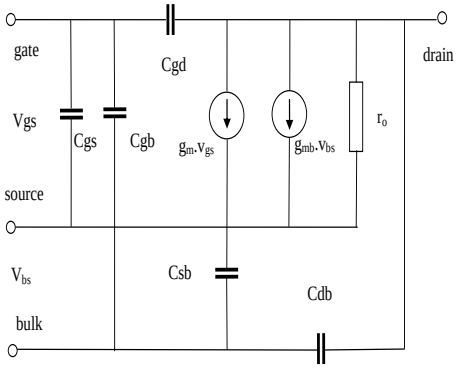


Fig. 3.Model of MOSFET transistor

The equation, describing the processes in this model is[4]:

$$i_{ds} = g_m.v_{gs} + g_{mb}.v_{bs} + \frac{1}{r_o}.v_{ds} \quad (4)$$

Here:

$g_{mb} = \frac{\partial i_D}{\partial v_{bs}}\Big|_{V_{GS},V_{DS},V_{BS}=const} = \frac{\gamma.g_m}{2.\sqrt{-V_{BS}-2.\phi_p}}$,where ϕ_p is the substrate Fermi potential (by NMOS transistors – p substrate)

The SPICE model of the MOSFET transistor depends on the nature and the complexity of the simulation. Accordingly the appropriate level of the SPICE transistor model should be chosen.

Using SPICE model with level 1 (Shichman - Hodges Model) devices with gate length greater than 10 um are considered. If the channel length is less than 4 mm this model is inappropriate.

Level 2 SPICE model is based on Myer's model. It is improved compared to the model with level 1[5].

Advantage of this model is the inclusion of the subthreshold conduction [6]. By means of this model the channel length modulation effect can be examined [5].

Level 2 SPICE model has problems mostly concerning the numerical convergence [6].

Level 3 SPICE model is an empirical model. It is more precise and effective in comparison to the SPICE models Level 1 and Level 2. By means of SPICE model level 3 devices with gate length approx. 2µm can be analyzed [5].

This model takes into account as the subthreshold conduction, as well the narrow width and short – channel effects[6].

The PSPICE Model of the MOSFET transistors, used in the presented paper, is as follows[6],[7]:

.model NMOS LEVEL=3; PHI=0.7; TOX=9.5E-09 XJ=0.2U TPG=1; VTO=0.7; DELTA=8.8E-1; LD=5E-8; KP=1.56E-4; UO=420 ;THETA=2.3E-01; RSH=2.0; GAMMA=0.62; NSUB=1.40E+17; NFS=7.2E+11; VMAX=1.8E+5; ETA=2.125E-2; KAPPA=1E-1; CGDO=3E-10; CGSO=3E-10; CGBO=4.5E-10; CJ=5.5E-4; MJ=0.6; CJSW=3E-10; MJSW=0.35; PB=1.1[6]

Here PHI is the surface potential; TOX is the oxide thickness; XJ is the Metallurgical junction depth; TPG is the Gate material type (in the case, considered in simulation it is opposite to the substrate); VTO is the zero-bias threshold voltage; KP is the transconductance; DELTA designates the width effect on threshold; LD is the lateral diffusion length; RSH is drain source diffusion sheet; GAMMA is bulk threshold parameter; NSUB is the Substrate doping density; NFS is the fast surface-state density; VMAX is the maximum drift velocity; ETA is the static feedback; KAPPA is the saturation field factor; CGDO is the Gate-Drain overlap; CGSO is the Gate - Source overlap; CGBO is the Gate-Bulk overlap; CJ is bulk p-n zero - bias bottom capacitance; MJ is bulk p-n bottom grading coefficient; CJSW is bulk p-n zero-bias perimeter capacitance; MJSW is bulk p-n sidewall grading coefficient; PB is the bilk p-n potential [7].

III. PSPICE MODELS OF THE ANALYZED CIRCUITS

For educational purposes, aimed toward examination of the Body effect in MOSFET transistor amplifiers in the presented paper several PSPICE circuit models are investigated.

A Body effect and $I_D - V_{GS}$ characteristics of the MOSFET transistors

The PSPICE MOSFET circuit model for study of the relationship between the Bulk effect and the $I_D - V_{GS}$ characteristics is shown on the Figure bellow.

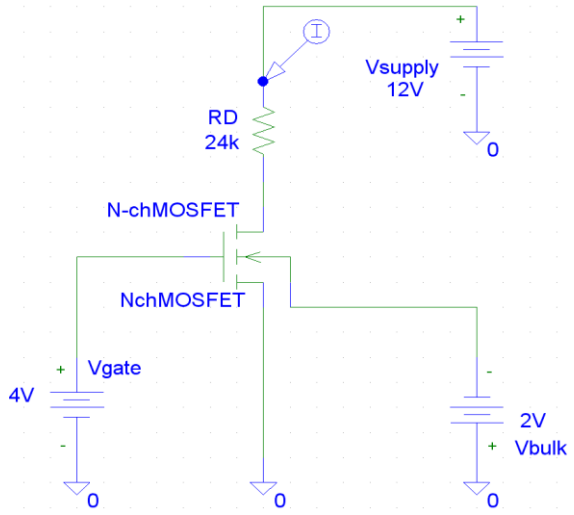


Fig. 4. PSPICE circuit model for study of the relationship between the Bulk effect and the $I_D - V_{GS}$ characteristics

DC sweep of the voltage, applied on the gate, is activated. With purpose to change the voltage on the bulk of the MOSFET transistor nested DC analysis is performed.

B Back gate effect and frequency response

In order to examine the frequency response of the MOSFET amplifier by different (initially set) values of V_{SB}, PSPICE circuit model, presented on Fig.6, is analyzed.

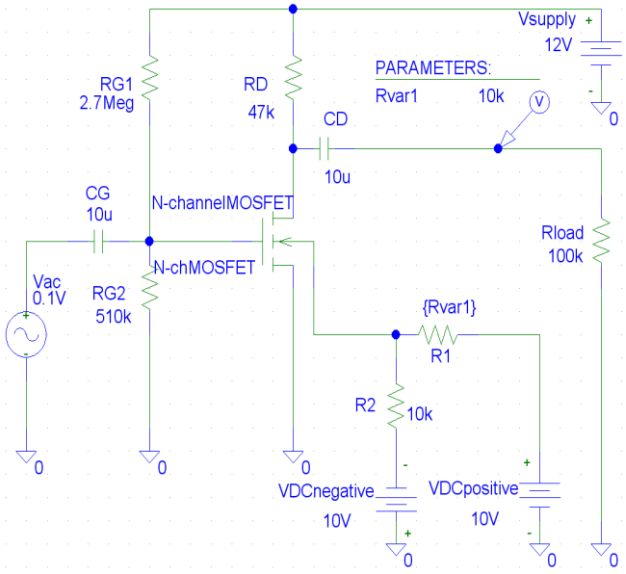


Fig. 5. PSPICE circuit model for AC analysis

The circuit, shown on Fig.5, is one stage common source MOSFET Transistor Amplifier. In order to perform AC analysis a sweep generator is applied on the input of the circuit.

IV. Simulation Results and Analyzes

A Results, concerning the $I_D - V_{GS}$ characteristics of the MOSFET Transistor in the amplifier

Results, obtained from PSPICE simulation of the circuit, shown on fig.4, are displayed bellow. A number of different values of the MOSFET Transistor Model parameter GAMMA have been used.

The graphical representation of the relationship: $I_D = f(V_{GS})$ is displayed on Fig.6.

The results are obtained by following conditions:

GAMMA = 0.2,

channel length L = 2µm,

width W = 16µm,

voltage Vsource-bulk (V_{SB}) changing in the range: $0V \div 8V$ with step = 2V

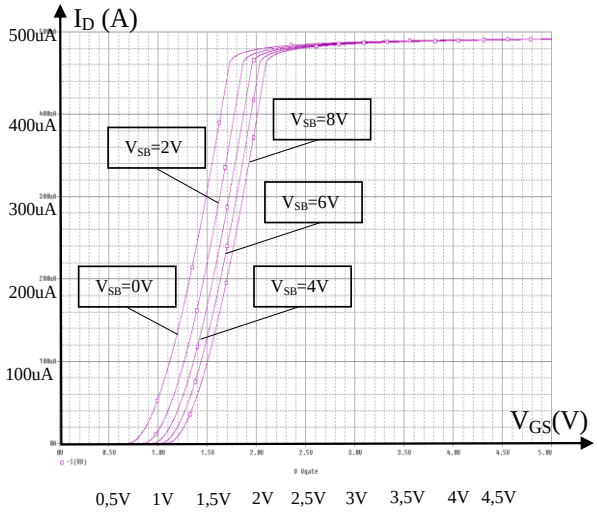


Fig. 6. Characteristics $I_D = f(V_{GS})$ by GAMMA = 0,2 and changing parameter V_{SB}

The increase of the Value of the parameter GAMMA leads to subsequent change in the characteristics: $I_D = f(V_{GS})$, as shown bellow.

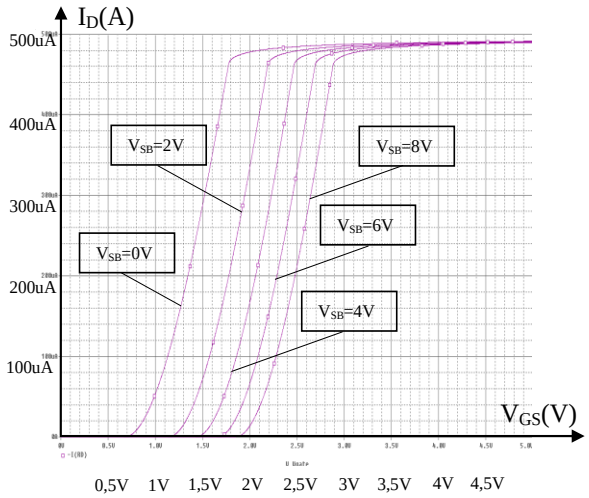


Fig. 7. Characteristics $I_D = f(V_{GS})$ by GAMMA = 0,62 and changing parameter V_{SB}

The results from Fig.7 concern GAMMA Value = 0.62, which is typical value for this PSPICE MOSFET Transistor model [6].

The results, obtained after an increase of the value of GAMMA to value = 1, are presented on Fig.8.

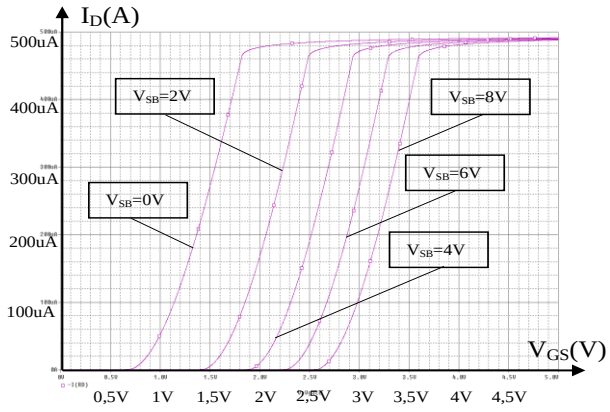


Fig. 8. Characteristics $I_D = f(V_{GS})$ by GAMMA = 1 and changing parameter V_{SB}

It becomes obvious the increase of the distance between the separate characteristics, obtained by defined value V_{BS}.

B Results, concerning the AC analysis of amplifier with MOSFET transistor.

As shown on Fig. 6, the examined single stage amplifier is common source. In order to obtain the bandwidth of this amplifier by different values of V_{SB} and fixed value of GAMMA, a number of experiments have been done.

a) The graphic, displayed on Figure 9, concern AC analysis by fixed value of GAMMA = 0.4 and V_{SB} set to 4,118V.

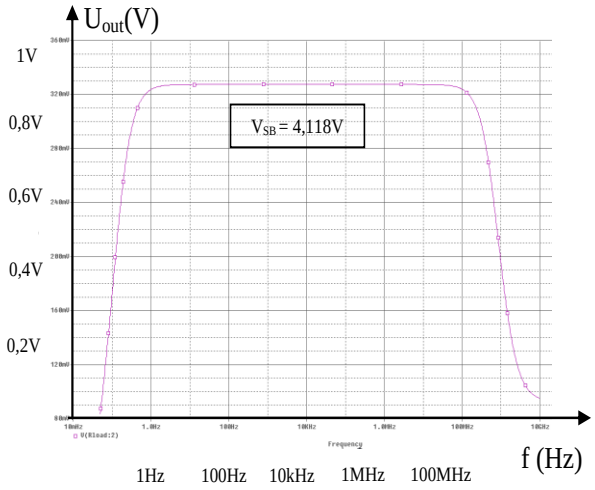


Fig. 9. AC analysis by GAMMA = 0,4 and V_{SB} = 4,118V

b) The increase of the value of parameter GAMMA to 0.62, which is typical value for this PSPICE MOSFET Transistor model [6], changes the location of the AC characteristic, as displayed on Figure 10.

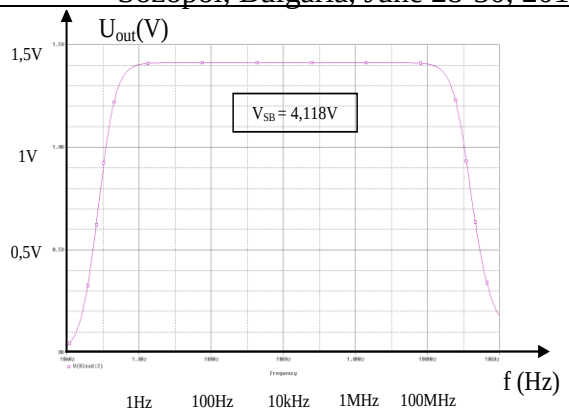


Fig. 10.AC analysis by GAMMA = 0,62 and V_{SB}=4,118V

It is obvious the great difference between the AC characteristics, presented on Fig.9 (where GAMMA = 0,4), and the graphical results, shown on Fig.10.

V. Conclusion

This paper is for educational purposes. It is aimed toward analysis of the Body effect in simple MOSFET circuits (one stage MOSFET transistor amplifiers). Instructions for simulation of the appropriate PSPICE models are given. Results, obtained from simulation are included.

References

[1] Prof. Wu , *Lecture 17* EE105 Spring 2008 ,UC Berkeley www-t.eecs.berkeley.edu/~ee105/sp10/lectures/lecture17.pdf

[2] Dr. Lynn Fuller, *SPICE Model Parameters for RIT MOSFET's*, November 15, 2013 Dr. Lynn Fuller Page 1 Rochester Institute of Technology *Microelectronic Engineeringhttps://people.rit.edu/lffeee/SPICE.pdf*

[3] Milaim Zabeli, Nebi Caka, Myzafere Limani, Qamil Kabashi, *The impact of MOSFET's physical parameters on its threshold voltage*, Proceedings of the 6th WSEAS International Conference on Microelectronics, Nanoelectronics, Optoelectronics, Istanbul, Turkey, May 27-29, 2007

[4] Prof. J. S. Smith, *MOS Transistor models: Body effects, SPICE models*, Lecture 15, EECS 105 Spring 2004,https://inst.eecs.berkeley.edu/~ee105/sp04/handouts/lectures/Lecture15.pdf

[5] Twesha Patel, *Comparison of Level 1, 2 and 3 MOSFET's* Course: Advanced Electronics,Semester: Fall 2014

[6] Thomas H. Lee Handout #2: EE214 Fall 2002A, *Review of MOS Device Physics*, https://web.stanford.edu/class/archive/ee/ee214/ee214.1032/Handouts/HO2.pdf

[7] Spice Model Parameters of MOSFETS, University of Pennsylvania Department of Electric Engineering https://www.seas.upenn.edu/~jan/spice/spice.MOSparamlist.html

Outage Probability of FSO System Using Optimal System Parameters in the Presence of Atmospheric Turbulence

Yordan Kovachev[1], Nikolay Kolev[2] and Tsvetan Mitsev[3]

Abstract – **In this paper the outage probability of FSO system using optimal system parameters in the presence of atmospheric turbulence is studied. Various statistical models for turbulent channels are explored and closed form expressions for calculating the outage probability in the presence of atmospheric turbulence are derived. Using these expressions numerical simulations of the outage probability of FSO system using optimal system parameters are performed.**

Keywords – **FSO, Gamma-gamma, Exponentiated Weibull, Outage probability, Log-normal distribution**.

I. INTRODUCTION

FSO (Free Space Optics) systems are high speed wireless optical communication systems that can be used as a backup of the traditional wireless technologies or as a last-mile solution in the high-speed local area networks (LAN). Their main disadvantage is the unpredictable reliability, as laser beam propagation through the atmospheric channel is highly influenced by various random atmospheric factors [1, 2].

One of the main factors that causes degradation in FSO availability is atmospheric turbulence [4 – 12]. Large turbulent eddies cause fluctuations in the initial beam direction, and small turbulent eddies cause scintillation. In this paper we evaluate the FSO channels by investigating their outage probability caused by atmospheric turbulence. We will study mainly the influence of turbulence induced fading (caused by scintillation) on the FSO system performance.

Over the years different statistical models have been proposed to describe the optical channel characteristics with respect to the atmospheric turbulence. The most commonly used are the log-normal and gamma-gamma distributions [13]. They showed to be suitable for modelling optical channels for weak-to-moderate and moderate-to-strong atmospheric turbulence. The gamma-gamma model shows better fit when used for modelling moderate-to-strong atmospheric turbulence channels. In the last few years the exponentiated Weibull is preferred for modelling moderate-to-strong turbulence channels, when aperture averaging is used [4, 6].

In this work we derive closed form expressions for calculating the outage probability of FSO links over atmospheric turbulence-induced fading channels modelled by the aforementioned distributions.

The rest of the paper is organized as follows: in Section 2 we present a short theoretical explanation of the atmospheric turbulence, introduce the channel models and the derivation of some system parameters needed for calculating the outage probability. In Section 3 the simulation results and some analysis of the results are presented.

[1]Yordan Kovachev is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: dakatapz@gmail.com.

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

[3]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

II. THEORY

A. Atmospheric Turbulence

In general atmospheric turbulence is caused by inhomogeneties in both temperature and pressure in the atmosphere, causing air cells or air pockets that are differently heated. This results in changes in the index of refraction, which in turn changes the path that the optical beam takes while it propagates through the atmospheric channel. Because the air pockets are not stable in time or space, the change of the index of refraction is random. A good measure of atmospheric turbulence is the refractive index structure coefficient C_n^2.

There are three main effects that optical beams experience when propagating through turbulent atmosphere. First, the laser beam direction can deviate randomly through the changing refractive index (beam wander). Second, the phase front of the optical beam can vary, producing intensity fluctuations or in other words scintillation. Third, the optical beam can spread more than diffraction theory can predict [13 - 14].

Beam Wander

This occurs in the presence of large turbulent eddies, or cells of turbulence that can be equal to or larger than the beam diameter. The radial variance σ_r can be presented as a function of wavelength λ and distance, z, as follows:

$$\sigma_r = 1.83 C_n^2 \lambda^{-1/6} z^{17/6} \quad (1)$$

This relationship implies that longer wavelengths are less affected by large turbulence eddies (they will have less beam wander).

Scintillation

FSO system performance is most affected by scintillation. The random interference with the wave front can cause peaks and dips, resulting in receiver saturation or signal loss. This is presented on Fig. 1:

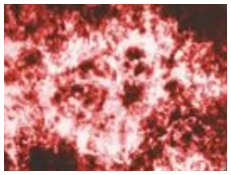


Fig. 1. Scintillation effect on the optical intensity

The scintillation causes redistribution of the optical intensity in the receiver's plane making it from normally distributed to log-normal or gamma-gamma distributed. Scintillation effects are characterized by the variance σ_I as shown in Eq. 2:

$$\sigma_I^2 = 1.23 C_n^2 k^{7/6} z^{11/6}, \quad (2)$$

where $k = 2\pi/\lambda$. Eq. 2 shows that larger wavelengths would experience smaller variance, all other factor being equal.

Beam Spreading

This is characterized by the effective radius $\rho_{z,eff}$, the distance from the center of the beam (z = 0) to where the relative mean intensity has decreased by 1/e. The effective radius is given by the expression:

$$\rho_{z,eff} = 2.01\left(\lambda^{-1/5} C_n^{6/5} z^{8/5}\right) \quad (3)$$

B. System Parameters Calculation

The needed system parameters can be calculated as follows. First we need to calculate the desired bit error rate (*BER*), so that the FSO system can function properly [11]:

$$BER = \frac{1}{2} erfc\left(\frac{SNR}{2\sqrt{2}}\right) \quad (4)$$

From Eq. (4) we can easily derive the value of the needed signal to noise ratio (*SNR*) required to maintain *BER*. Having the *SNR* defined, the minimal intensity needed at the plane of the receiving aperture can be calculated using the equation:

$$I_{min} = \frac{\Phi_{PD|SNR=const}}{\pi \tau_r R_r^2}, \quad (5)$$

where $\Phi_{PD}$is the optical power at the photo detector's aperture, needed to keep the *BER* calculated with Eq. (4). This optical power can be derived by:

$$\Phi_{pd} = \frac{1}{2}\left[\frac{SNR^2 . C_I . e^-}{R_I} + \left(\left(-\frac{SNR^2 . C_I . e^-}{R_I}\right)^2 + \frac{4 SNR^2 . C_I}{R_I}\left(\frac{2 k_B . T . A}{R_I . R_{Fb}} + e^- . \Phi_B\right)\right)^{\frac{1}{2}}\right] \quad (6)$$

where $R_I = 8.06.10^5 \eta(\lambda_0)\lambda_0$ and $\eta(\lambda_0)$ are the integral sensitivity and the quantum efficiency of the photodetector, C_I is the channel capacity. A is a constant of the receiver; R_{FB} is the value of the resistor in the feedback of the preamplifier and $e^- = 1{,}602.10^{-19} C$ is the electric charge of the electron. The background optical flux Φ_B is defined by the brightness of the background radiation $L_{\lambda B}$, the transmission wavelength of the interference filter before the photodetector $\Delta_{\lambda F}$, the losses in the transmitting antenna τ_t, and the parameters of the receiver: radius of the receiver's aperture R_r and its angular width θ_r:

$$\Phi_B = \pi^2 \tau_r L_{\lambda,B} R_r^2 \theta_r^2 \Delta\lambda_F \quad (7)$$

Having Eqs. (5) and (6), the optimal system parameters for maintaining I_{min} and respectively *BER* can be calculated. The optimal beam radius $\rho_{z,opt}$ is:

$$\rho_{z,opt} = \sqrt{\frac{2\tau_a \tau_t \Phi_L}{\pi e I_{min}}}, e = 2.7182 \quad (8)$$

and the optimal beam divergence angle $\theta_{t,opt}$ is:

$$\theta_{t,opt} = \frac{1}{z}\sqrt{\frac{2\tau_a \tau_t \tau_r R_r^2 \Phi_L}{e \Phi_{PD}|_{SNR=const}}} \quad (9)$$

where τ_a is the atmospheric transparency and τ_r denotes the losses in the receiver's antenna. Φ_L is the transmitted optical power.

These optimal system parameters (Eqs. (8) and (9)) allow for the compensation of the errors in the FSO system caused by random angular vibrations in the transmitting antenna. These angular vibrations correspond to linear misalignments between the optical beam axis and the center of the receiving antenna. The maximal linear misalignments that can be compensated using optimal beam radius and divergence angle are:

$$\rho_{max} = \frac{1}{\sqrt{2}}\rho_{z,opt}\sqrt{\frac{2\tau_a \tau_t \Phi_L}{\pi \rho_{z,opt}^2 (1 - e^{-2}) I_{min}}} \quad (10)$$

C. Channel Models

The FSO channel with atmospheric turbulence is well described by log-normal and gamma-gamma distributions. The optical intensity distribution at the plane of the receiver is presented with the following expression in the case of log-normal distribution [3, 10, 13]:

$$f(I) = \frac{1}{\sqrt{2\pi} I \sigma_I} exp\left[-\frac{(\ln(I) - \sigma_I^2/2)^2}{2\sigma_I^2}\right], \quad (11)$$

where σ_I is the standard deviation of the log-normal distribution and depends on the channel characteristics.

$$\sigma_I^2 = exp\left[\frac{0.49\delta^2}{(1 + 0.18d^2 + 0.56\delta^{12/5})^{7/6}} + \frac{0.51\delta^2}{(1 + 0.9d^2 + 0.62d^2\delta^{12/5})^{5/6}}\right] - 1 \quad (12)$$

where

$$d = \sqrt{\frac{kR_r^2}{4z}}, k = \frac{2\pi}{\lambda}, \delta^2 = 1.23 C_n^2 k^{7/6} z^{11/6} \quad (13)$$

In the case of gamma-gamma distribution:

$$f(I) = \frac{2(\alpha\beta)^{\alpha+\beta/2}}{\Gamma(\alpha)\Gamma(\beta)\langle I\rangle}\left(\frac{I}{\langle I\rangle}\right)^{\alpha+\beta/2-1} K_{\alpha-\beta}\left(2\sqrt{\frac{\alpha\beta I}{\langle I\rangle}}\right), \quad (14)$$

where

$$\alpha = \left[exp\left(\frac{0.49\delta^2}{(1 + 0.18d^2 + 0.56\delta^{12/5})^{7/6}}\right) - 1\right]^{-1}$$
$$\beta = \left[exp\left(\frac{0.51\delta^2}{(1 + 0.9d^2 + 0.62d^2\delta^{12/5})^{5/6}}\right) - 1\right]^{-1} \quad (15)$$

In the presence of moderate-to-strong atmospheric turbulence and when aperture averaging is used exponentiated Weibull distribution gives the best fit for the optical intensity distribution [4-6]:

$$f(I,\alpha,\beta,\eta) = \frac{\alpha\beta}{\eta}\left(\frac{I}{\eta}\right)^{\beta-1} exp\left[-\left(\frac{I}{\eta}\right)^{\beta}\right]\left\{1 - exp\left[-\left(\frac{I}{\eta}\right)^{\beta}\right]\right\}^{\alpha-1} \quad (16)$$

where α and β are the second scale parameter and the scale parameter respectively and η is the shape parameter of the Weibull distribution:

$$\begin{aligned} \beta &= 1.012(\alpha\sigma_I^2)^{-13/25} + 0.142 \\ \alpha &= 3.93\left(\frac{R_r}{\delta}\right)^{-0.519} \\ \delta &= (1.46C_n^2k^2z)^{-3/5} \\ \eta &= \frac{1}{\alpha\Gamma(1+1/\beta)g_1(\alpha,\beta)} \end{aligned} \quad (17)$$

D. Outage Probability Calculation

In this work, by outage probability we mean that the system will work with bit error rate greater than the one calculated with Eq. (4). In other words, if we calculate *SNR* and I_{min} to maintain *BER* = 10^{-8}, then by system outage we will consider any time the bit error rate gets greater than 10^{-8} due to atmospheric turbulence.

Having the optical intensity distribution defined in Eq. (11) through Eq. (17), the outage probability P_{out} will be equal to $P(I < I_{min})$ or, in other words, the cumulative density function CDF or $F(I)$ of the corresponding pdf. This corresponds to:

$$P_{out} = P(I < I_{min}) = \int_0^{I_{min}} f(I)dI \quad (18)$$

In the case of log-normal distribution, the outage probability is derived by replacing Eq. (11) in Eq. (18) [13, 10]:

$$P_{out} = F(I_{min}) = \frac{1}{2}erfc\left(-\frac{\ln(I) + 0.5\sigma_I^2}{\sqrt{2}\sigma_I}\right) \quad (19)$$

When gamma-gamma distribution is used, the outage probability is [15, 16]:

$$P_{out} = \frac{(\alpha\beta)^{\alpha+\beta/2}}{\Gamma(\alpha)\Gamma(\beta)}(I)^{\alpha+\beta/2} \times G_{1,3}^{2,1}\left[\alpha\beta(I)\Big|_{\frac{\alpha-\beta}{2},\frac{\beta-\alpha}{2},-\frac{\alpha+\beta}{2}}^{1-\frac{\alpha+\beta}{2}}\right] \quad (20)$$

And for the case, when exponentiated Weibull distribution is used [4 - 6], the outage probability is:

$$P_{out} = F(I_{min}) = \left\{1 - exp\left[-\left(\frac{I}{\eta}\right)^{\beta}\right]\right\}^{\alpha} \quad (19)$$

III. Simulation Results and Discussion

The numerical simulations are performed using the following FSO system parameters: quantum efficiency of the photodetector material $\eta(\lambda_0)$ = 0,7; central wavelength λ_0 = 1,55 μm; T = 300 K; aperture coefficient A = 5; value of the resistor in the feedback of the preamplifier, R_{Fb} = 1 kΩ; $\tau_r = \tau_t$ = 0,85; R_r = 5,5 cm; transmission wavelength of the interference filter before the photodetector $\Delta_{\lambda F}$ = 10 nm; background radiation, $L_{\lambda,B} = 10^{-2}$ W/m2.sr.Ång (corresponds to bright day); angular width of the receiving antenna θ_r = 5 mrad and z = 2km.

From here-on after as outage probability we will consider the probability for $I < I_{min}$. This doesn't mean that the system will stop working it just indicates that the FSO will work with bit error rate larger than the one calculated with Eq. (4). The outage probability is calculated using the log-normal distribution model for the atmospheric turbulence channel (Eq. (11)). It is chosen because of its mathematical simplicity; $C_n^2 = 2.3.10^{-13}$.

Fig. 2 shows the outage probability of FSO system using optimal system parameters depending on different values of *SNR*. Fig. 3 represents the dependence of the outage probability of FSO system using optimal system parameters on the channel capacity (C_I).

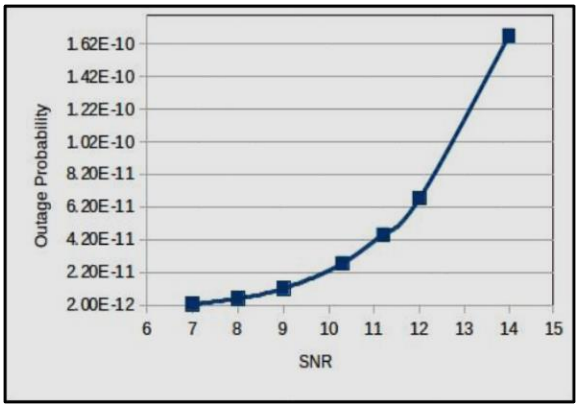


Fig. 2. Outage probability of FSO system using optimal system parameters depending on SNR; R_r = 5.5 cm = const; C_I = 1.25 Gbps = const

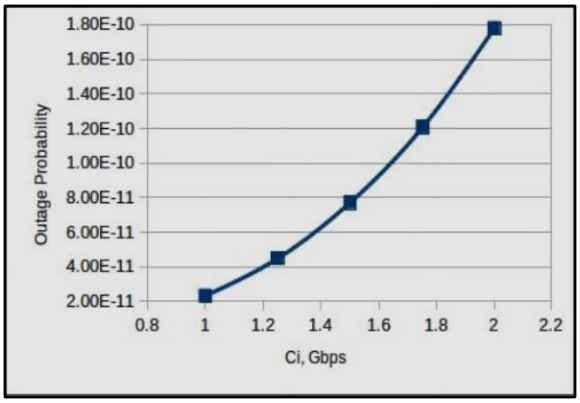


Fig. 3. Outage probability of FSO system using optimal system parameters depending on the channel capacity C_I; *SNR* = 11.2 = const R_r = 5.5 cm = const.

It is observed in Fig. 2 that the higher the *SNR* the higher the outage probability. This is because as seen in Eq. (5) greater *SNR* values (respectively lower *BER*) require greater value of I_{min}, which means, that because of the turbulent

channel (and the scintillation), there is a greater chance the intensity at the plane of the receiver drops bellow the required I_{min}, needed to keep constant *BER* calculated with Eq. (4). In Fig. 3 higher channel capacity requires higher values of I_{min}, which again means greater probability that the optical intensity at the plane of the receiver could be lower than the required I_{min}, needed to keep *BER* = const as calculated by Eq. (4).

Fig. 4 depicts the dependence of the outage probability of FSO system using optimal system parameters on the receiver's aperture R_r:

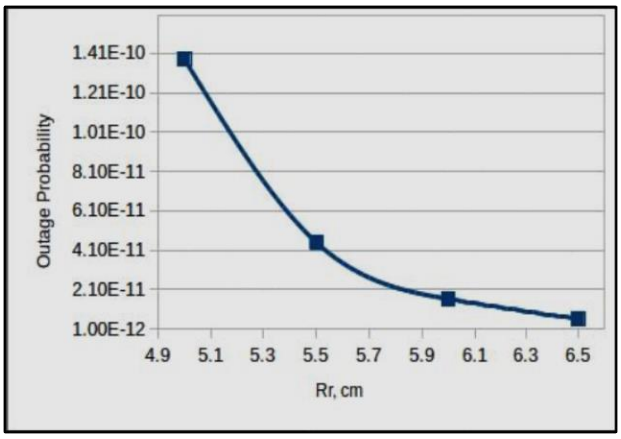


Fig. 4. Outage probability of FSO system using optimal system parameters depending on receiver's aperture R_r; *SNR* = 11.2 = const; C_I = 1.25 Gbps = const

In Fig. 4 the bigger the aperture radius (R_r) the more optical power gets in the receiver, so lower values of I_{min} are required to keep *BER* = const, which means lower outage probability.

It is observed (in Fig. 3 and Fig. 4), that because the FSO system uses optimal values of the beam radius and the beam divergence angle respectively, there are less fluctuations in the outage probability. It varies in the interval [10^{-11}, 10^{-10}]. This means that using optimal system parameters can guarantee some level of predictability of the FOS system's availability in the presence of atmospheric turbulence over the optical channel.

IV. Conclusion

In this paper the outage probability of FSO system using optimal system parameters in the presence of atmospheric turbulence is studied. For the purpose various statistical models for turbulent channels were explored. Closed form expressions for the outage probability in the presence of atmospheric turbulence were derived.

The outage probability of the FSO system depending on various parameters was simulated and the results were graphically represented in Fig. 2 through Fig. 4. The results can be used when designing FSO systems using optimal system parameters.

References

[1] Kedar D., Arnon S.: "Urban optical wireless communication network: The main challenges and possible solutions", IEEE Optical Communications, 2004, pp. S1–S7.

[2] Y. Kovachev, Ts. Mitsev, A Brief Survey of the Methods for Increasing FSO Availability, In-Sillico Intellect, Vol. 1 , No. 1, pp 41-44

[3] H.E. Nistazakis1 T.A. Tsiftsis G.S. Tombras, "Performance analysis of free-space optical communication systems over atmospheric turbulence channels", IET Commun., 2009, Vol. 3, Iss. 8, pp. 1402–1409

[4] Ping Wang, Jiao Qin, Lixin Guo, and Yintang Yang, "BER Performance of FSO Limited by Shot and Thermal Noise Over Exponentiated Weibull Fading Channels", IEEE Photonics Technology Letters, Vol. 28, NO. 3, February 1, 2016, pp. 252-255

[5] Md. Mobasher Ahmed, Kazi Tanvir Ahmmed, "An Expression for Average Bit Error Rate of Free Space Optical Links Over Exponentiated Weibull Atmospheric Turbulence Channel", 3rd International Conference on Informatics, Electronics & Vision 2014

[6] Haitham S. Khallaf, Abdulaziz E. El-Fiqi, Hossam M. H. Shalaby, and Salah S., A. Obayya, "Average SER of MPPM Technique over Exponentiated Weibull Fading FSO Channels Considering Fog and Beam Divergence", ICTON, 2017

[7] Neha Mishra, D Sriram Kumar, "Outage Analysis of Relay Assisted FSO Systems over K-Distribution Turbulence Channel", International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), 2016

[8] Norhanis Aida M. Nor, Zabih Ghassemlooy, Stanislav Zvanovec, Manav R. Bhatnagar, "Comparison of Optical and Electrical based Amplifying-Forward Relay-Assisted FSO Links over Gamma-Gamma Channels", 10th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP), 2016.

[9] J.-Y. Wang, J.-B. Wang, M. Chen, Y. Tang, and Y. Zhang, "Outage analysis for relay-aided free-space optical communications over turbulence channels with nonzero boresight pointing errors," IEEE Photon. J., vol. 6, no. 4, Aug. 2014, Art. ID 7901815.

[10] Ales Prokes, "Modelling of Atmospheric turbulence Effect on Terrestrial FSO Link", Radioengineering, Vol. 18, No. 1, April 2009, pp. 42-47.

[11] Ts. Mitsev, Y. Kovachev, "Availability of Free-Space Optical Systems Depending on Atmospheric Conditions and System Parameters", International Journal of Wireless Communications and Networking Technologies, Volume 5, No. 3, April – May 2016

[12] T.A. Tsiftsis, "Performance of heterodyne wireless optical communication systems over gamma-gamma atmospheric turbulence channels", Electronics Letters 28th February 2008 Vol. 44 Issue 5, pp. 372-373

[13] Larry C. Andrews, Ronald L. Phillips, "Laser Beam Propagation through Random Media", SPIE PRESS, 2005.

[14] Heinz Willebrand, Baksheesh S. Ghuman, "Free-Space Optics: Enabling Optical Connectivity in Today's Networks" , Sams Publishing, 2002.

[15] Adamchik, V.S., and Marichev, O.I.: "The algorithm for calculating integrals of hypergeometric type functions and its realization in REDUCE system" Proc. Int. Conf. on Symbolic and Algebraic Computation, Tokyo, Japan, 1990, pp. 212–224 S.

[16] Gradshteyn, I.S., and Ryzhik, I.M.: 'Table of integrals, series, and products' (Academic Press, New York, 2000, 6th edn.)

Performance Assessment of IEEE 802.11a 54 Mbps WEP Laboratory Links

José A. R. Pacheco de Carvalho[1],

Cláudia F. F. P. Ribeiro Pacheco[2], Hugo Veiga[3], António D. Reis[4]

Abstract – **The importance of wireless communications, involving electronic devices, has been widely recognized. Performance is a fundamental issue, resulting in more reliable and efficient communications. Security is also critically important. Laboratory measurements were performed about several performance aspects of Wi-Fi IEEE 802.11a 54 Mbps WEP links. Our study contributes to performance evaluation of this technology, using available equipments (HP V-M200 access points and Linksys WPC600N adapters). New results are presented and discussed, namely at OSI level 4, from TCP and UDP experiments. TCP throughput is measured versus TCP packet length. Jitter and percentage datagram loss are measured versus UDP datagram size. Results are compared for both point-to-point and four-node point-to-multipoint links. Conclusions are drawn about performance of the links.**

Keywords – **IEEE 802.11a, Multi-Node WEP Links, TCP packet size, UDP datagram size, Wi-Fi, WLAN Laboratory Performance.**

I. INTRODUCTION

Electromagnetic waves in several frequency ranges, propagating in the air, have made possible the development of contactless communication technologies. Wireless fidelity (Wi-Fi) and free space optics (FSO) are examples of such technologies. Microwaves and laser light are used, respectively. Their importance and utilization have been growing worldwide.

Wi-Fi completes traditional wired networks. The main configuration is infrastructure mode. Here, a WLAN (wireless local area network) is formed where an access point, AP, enables communications of Wi-Fi electronic devices with a wired based LAN, through a switch/router. At the private home level a WPAN (wireless personal area network) permits personal devices to communicate. Frequency bands of 2.4 and 5 GHz are usable, with IEEE 802.11a, b, g, n standards [1]. Nominal transfer rates up to 11 (802.11b), 54 Mbps (802.11 a, g) and 600 Mbps (802.11n) are stated. Carrier sense multiple access with collision avoidance (CSMA/CA) is the medium access control. Point-to-point (PTP) and point-to-multipoint (PTMP) microwave links are applied. The intensive use of the 2.4 GHz band has caused growing interference. The 5 GHz band solves this problem, at the expense of higher absorption and shorter range.

802.11a,g use a multi-carrier modulation scheme called orthogonal frequency division multiplexing (OFDM) that allows for binary phase-shift keying (BPSK), quadrature phase-shift keying (QPSK) and quadrature amplitude modulation (QAM) of the 16-QAM and 64-QAM density types. One spatial stream (one antenna) and coding rates up to 3/4 are possible and a 20 MHz channel.

There are studies on wireless communications, wave propagation [2],[3], practical setups of WLANs [4], performance analysis of the effective transfer rate [5], performance in crowded indoor environments [6].

Communication performance is a crucial issue, giving higher reliability and efficiency. Requirements are given for new and traditional telematic applications [7].

Wi-Fi security is critically important for confidentiality. Several encryption methods have been developed to provide authentication such as, by increasing order of security, wired equivalent privacy (WEP), Wi-Fi protected access (WPA) and Wi-Fi protected access II (WPA2).

Several performance measurements have been published for 2.4 and 5 GHz Wi-Fi Open [8], WEP [9], WPA[10] and WPA2 [11] links, as well as very high speed FSO [12]. Studies are published on modelling TCP throughput [13]. A formula that bounds average TCP throughput is available [14]. Studies have been given for 5 GHz 802.11n and 54 Mbps 802.11a Open links [15],[16].

The motivation of this work is to evaluate and compare performance in laboratory measurements of WEP PTP and four-node point-to-multipoint (4N-PTMP) 802.11a links at 54 Mbps using new available equipments. This new contribution permits to increase the knowledge about performance of Wi-Fi (IEEE 802.11 a) links [16]. The problem statement is that performance needs to be evaluated under several TCP and UDP parameterizations and link topologies under WEP encryption. The solution proposed uses an experimental setup and method, permitting to monitor signal to noise ratios (SNR) and noise levels (N), measure TCP throughput (from TCP connections) versus TCP packet size, and UDP jitter and

[1]José Pacheco de Carvalho is with the APTEL Research Group and the Physics Department at the University of Beira Interior, R. Marquês d'Ávila e Bolama, 6201-001 Covilhã, Portugal, E-mail: pacheco@ubi.pt.

[2]Cláudia Pacheco is with the APTEL Research Group at the University of Beira Interior, R. Marquês d'Ávila e Bolama, 6201-001 Covilhã, Portugal, E-mail: a17597@ubi.pt.

[3]Hugo Veiga is with the APTEL Research Group 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.

[4]António Reis is with the APTEL Research Group 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.

percentage datagram loss (from UDP communications) versus UDP datagram size.

The remaining of the paper is organized as follows: Section II is about the experimental conditions i.e. the measurement setup and procedure. Results and discussion are given in Section III. Conclusions are drawn in Section IV.

II. Experimental details

In here we have used a HP V-M200 access point [17], with three external dual-band 3x3 MIMO antennas, IEEE 802.11 a/b/g/n, software version 5.4.1.0-01-16481, a 1000-Base-T/100-Base-TX/10-Base-T layer 2 3Com Gigabit switch 16 and a 100-Base-TX/10-Base-T layer 2 Allied Telesis AT-8000S/16 switch [18]. Three PCs were adopted having a PCMCIA IEEE.802.11 a/b/g/n Linksys WPC600N wireless adapter with three internal antennas [19], to enable 4N-PTMP links to the access point. In every type of experiment, an interference free communication channel was used (ch. 36). This was centrally checked through a portable computer, equipped with a Wi-Fi 802.11 a/b/g/n adapter, running Acrylic WiFi software [20]. WEP encryption was turned on in the AP and the wireless adapters of the PCs, with 128 bits and a key composed of twenty six hexadecimal characters. The experiments were made under far-field conditions. Power levels of 30 mW (15 dBm) were not exceeded, as the wireless equipments were proximate.

A functional laboratory arrangement has been planned and set up for the measurements, as shown in Fig. 1. Up to three wireless links to the AP are viable. At OSI level 4, measurements were made for TCP connections and UDP communications using Iperf software [21]. For a TCP client/server connection (TCP New Reno, RFC 6582, was used), TCP throughput was obtained for a given TCP packet size, varying from 0.25k to 64k bytes. For a UDP client/server communication with a given bandwidth specification, UDP jitter and percentage loss of datagrams were determined for a given UDP datagram size, varying from 0.25k to 64k bytes.

One PC, with IP 192.168.0.2 was the Iperf server and the others, with IPs 192.168.0.6 and 192.168.0.50, were the Iperf clients (client1 and client2, respectively). Jitter, which is the root mean square of differences between consecutive transit times, was constantly computed by the server, as specified by the real time protocol RTP, in RFC 1889 [22]. A control PC, with IP 192.168.0.20, was mainly intended to control the AP configuration. The net mask of the wireless network was 255.255.255.0. Three types of measures are feasible: PTP, using the client1 and the control PC as server; PTMP, using the client1 and the 192.168.0.2 server PC; 4N-PTMP, using simultaneous connections/communications between the two clients and the 192.168.0.2 server PC.

The server and client PCs were HP nx9030 and nx9010 portable computers, respectively. The control PC was an HP nx6110 portable computer. Windows XP Professional SP3 was the operating system. The PCs were arranged to provide maximum resources to the present work. Batch command files have been re-written for the new TCP and UDP tests.

The results were obtained in batch mode and recorded as data files to the client PCs disks. Every PC had a second Ethernet adapter, to permit remote control from the official IP APTEL (Applied Physics and Telecommunications) Research Network, via switch.

III. Results and discussion

The wireless network adapters of the PCs were manually configured for a nominal rate of 54 Mbps. WEP encryption was activated in the AP and the wireless network adapters of the PCs. Transmit and receive rates were typically 54 Mbps, as monitored in the AP. For every TCP packet size in the range 0.25k-64k bytes, and for every corresponding UDP datagram size in the same range, data were acquired for WEP 4N-PTMP and PTP links at OSI levels 1 (physical layer) and 4 (transport layer) using the setup of Fig. 1. For every TCP packet size an average TCP throughput was calculated from a series of experiments. This value was taken 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 obtained in the AP. Typical values are shown in Fig. 2. The links had good, high, SNR values. The main average TCP and UDP results are summarized in Table I, for WEP 4N-PTMP and PTP links. The statistical analysis, including calculations of confidence intervals, was made as in [23]. In Fig. 3 polynomial fits were made (shown as y versus x), using the Excel worksheet, to the TCP throughput data both for both links, where R^2 is the coefficient of determination. It gives the goodness of fit. It is 1.0 for a perfect fit to data. It was found that, on average, the best TCP throughputs are for PTP links (Table I). In passing from PTP to 4N-PTMP throughput reduces to 23%. This is due to increase of processing requirements for the AP, so as to maintain links between the PCs. Fig. 3 shows a fair increase in TCP throughput with packet size. Small packets give a large overhead, due to small amounts of data that are sent in comparison to the protocol components. Frame has a very heavy role in Wi-Fi. For larger packets, overhead decreases; the amount of sent data overcomes the protocol components.

In Figs. 4 and 5, the data points of jitter and percentage datagram loss were joined by smoothed lines. The vertical axes have log 10 scales. It was found that, on average, the best jitter performance is for PTP links. There are large error bars mainly in the 8k, 16k, 32k data points of the 4N-PTMP curve (Fig. 4), needing further investigations. For PTP it can be seen that, for small sized datagrams, jitter is small. There are small delays in sending datagrams. Latency is also small. For larger datagram sizes jitter increases.

Concerning average percentage datagram loss, performances were found on average significantly better for PTP than for 4N-PTMP links (Table I). There are large error bars mainly in the 16k, 32k data points of the 4N-PTMP curve, needing further investigations. Fig. 5, mainly for 4N-PTMP, shows larger percentage datagram losses for small sized datagrams, when the amounts of data to send are small in comparison to the protocol components. There is considerable processing of frame headers and buffer management. For larger datagrams, percentage datagram loss

is lower. However, large UDP segments originate fragmentation at the IP datagram level, leading to higher losses.

TCP throughput, jitter and percentage datagram loss were generally found to show performance degradations due to link topology, in passing from PTP to 4N-PTMP, where processing requirements for the AP are higher so as to maintain links between PCs. As CSMA/CA is the medium access control, the available bandwidth and the air time are divided by the nodes using the medium. In comparison to Open links [16], TCP throughput did not show significantly sensitive to WEP within the experimental error. In passing from Open to WEP links, where data length increases due to encryption, jitter and percentage datagram loss have shown considerable performance degradations.

In comparison to previous results for 5 GHz 802.11n Open links [15] the present results show that 5 GHz 802.11n gives better TCP, jitter and datagram loss performances than 802.11a.

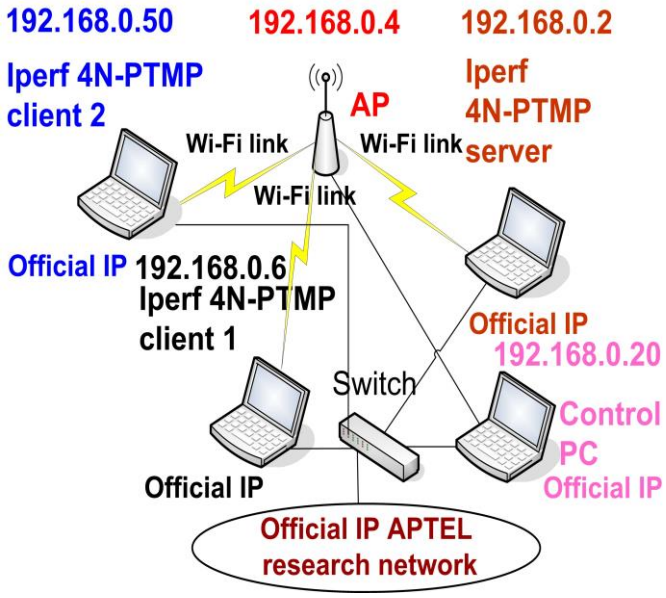


Fig. 1. Wi-Fi laboratory arrangement.

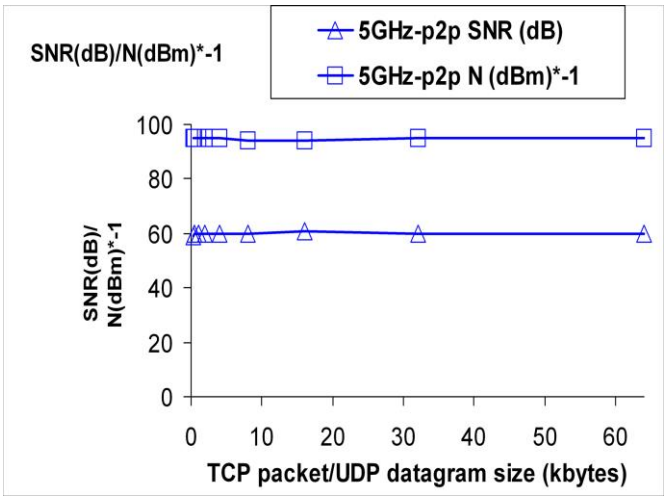


Fig. 2. Typical SNR (dB) and N (dBm).

TABLE I
AVERAGE WI-FI (IEEE 802.11A) RESULTS
WEP PTP AND 4N-PTMP LINKS

Parameter/Link type	PTP	4N-PTMP
TCP throughput (Mbps)	23.2 +- 0.7	5.4 +- 0.2
UDP-jitter (ms)	3.7 +- 0.3	28.6 +- 17.6
UDP-% datagram loss	1.3 +- 0.1	30.4 +- 9.5

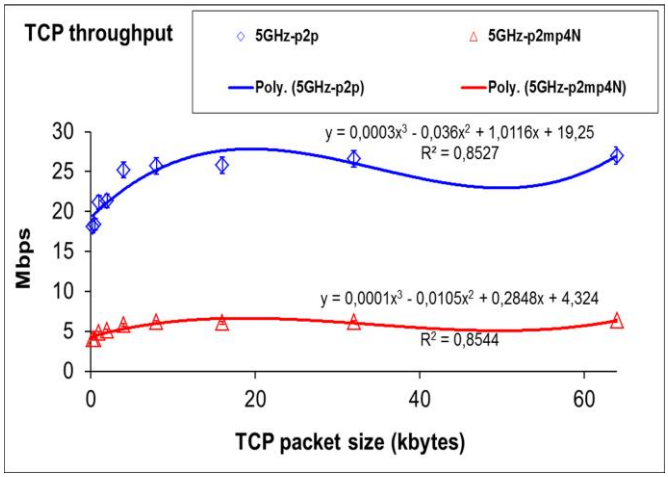


Fig. 3. TCP throughput (y) versus TCP packet size (x). WEP links.

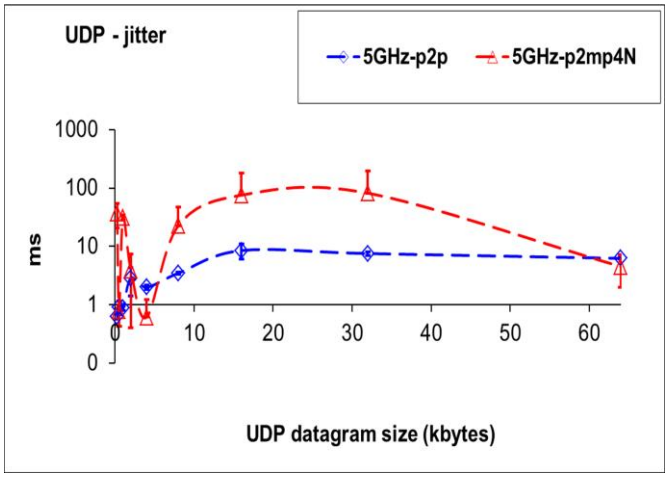


Fig. 4. UDP - jitter versus UDP datagram size. WEP links.

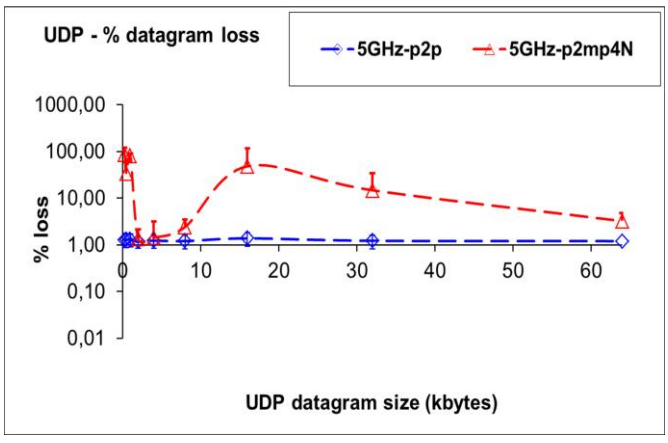


Fig. 5. UDP – percentage datagram loss versus UDP datagram size. WEP links.

IV. CONCLUSION

In the present work a versatile laboratory setup arrangement was devised and implemented, that permitted systematic performance measurements using new available wireless equipments (V-M200 access points from HP and WPC600N adapters from Linksys) for Wi-Fi (IEEE 802.11 a) in 54 Mbps WEP PTP and 4N-PTMP links. Through OSI layer 4, TCP and UDP performances were measured versus TCP packet size and UDP datagram size, respectively. TCP throughput, jitter and percentage datagram loss were measured and compared for WEP PTP and 4N-PTMP links. TCP throughput was found to increase with packet size. For PTP jitter, for small sized datagrams, is found small. It increases for larger datagrams. Concerning percentage datagram loss, it was found high for small sized datagrams. For larger datagrams it diminishes. However, large UDP segments originate fragmentation at the IP datagram level, leading to higher losses. In comparison to PTP links, TCP throughput, jitter and percentage datagram loss were found to show significant performance degradations for 4N-PTMP links, where the AP experiments higher processing requirements for maintaining links between PCs. Unlike jitter and percentage datagram loss, TCP throughput has not shown significant sensitivity to WEP. The present results show that 5 GHz 802.11n gives better TCP, jitter and datagram loss performances than 802.11a.

Further performance studies are planned using several standards, 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 2017.

[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] F. Boavida, E. Monteiro, Engenharia de Redes Informáticas, 10th ed., FCA-Editora de Informática Lda, Lisbon, 2011.

[8] J. A. R. Pacheco de Carvalho, H. Veiga, C. F. Ribeiro Pacheco, A. D. Reis, "Extended Performance Research on Wi-Fi IEEE 802.11 a, b, g Laboratory Open Point-to-Multipoint and Point-to-Point Links", in Transactions on Engineering Technologies, Sio-Iong Ao, Gi-Chul Yang, Len Gelman, Eds. Singapore: Springer, 2016, pp. 475-484.

[9] J. A. R. Pacheco de Carvalho, H. Veiga, N. Marques, C. F. Ribeiro Pacheco, A. D. Reis, "Wi-Fi WEP Point-to-Point Links-Performance Studies of IEEE 802.11 a,b,g Laboratory Links", in Electronic Engineering and Computing Technology, Series: Lecture Notes in Electrical Engineering, Sio-Iong Ao, Len Gelman, Eds. Netherlands: Springer, 2011, Vol. 90, pp. 105-114.

[10] J. A. R. Pacheco de Carvalho, H. Veiga, C. F. Ribeiro Pacheco, A. D. Reis, "Extended Performance Studies of Wi-Fi IEEE 802.11a, b, g Laboratory WPA Point-to-Multipoint and Point-to-Point Links", in Transactions on Engineering Technologies: Special Volume of the World Congress on Engineering 2013, Gi-Chul Yang, Sio-Iong Ao, Len Gelman, Eds. Gordrecht: Springer, 2014, pp. 455-465.

[11] J. A. R. Pacheco de Carvalho, H. Veiga, C. F. Ribeiro Pacheco, A. D. Reis, "Performance Evaluation of IEEE 802.11 a, g Laboratory WPA2 Point-to-Multipoint Links", Lecture Notes in Engineering and Computer Science: Proceedings of the World Congress of Engineering 2014, WCE 2014, 2-4 July 2014, London, U. K., pp. 699-704.

[12] J. A. R. Pacheco de Carvalho, N. Marques, H. Veiga, C. F. F. Ribeiro Pacheco, A. D. Reis, "Performance Measurements of a 1550 nm Gbps FSO Link at Covilhã City, Portugal", Proc. Applied Electronics 2010 - 15th International Conference, 8-9 September 2010, University of West Bohemia, Pilsen, Czech Republic, pp. 235-239.

[13] J. Padhye, V. Firoiu, D. Towsley, J. Kurose, "Modeling TCP Throughput: A Simple Model and its Empirical Validation", Proc. SIGCOMM Symposium Communications, Architecture and Protocols, August 1998, pp. 304-314.

[14] M. Mathis, J. Semke, J. Mahdavi, "The Macroscopic Behavior of the TCP Congestion Avoidance Algorithm", ACM SIGCOMM Computer Communication Review, vol. 27, Issue 3, July 1997, pp. 67-82.

[15] J. A. R. Pacheco de Carvalho, H. Veiga, C. F. Ribeiro Pacheco, A. D. Reis, " Performance Measurements of Open 5 GHz IEEE 802.11n Laboratory Links", Lecture Notes in Engineering and Computer Science: Proceedings of the World Congress on Engineering 2016, WCE 2016, 29 June-1 July, 2016, London, U.K., pp. 607-611.

[16] José A. R. Pacheco de Carvalho, Cláudia F. F. P. Ribeiro Pacheco, Hugo Veiga, António D. Reis, "A New Performance Determination of IEEE 802.11a 54 Mbps Open Laboratory Links", *Proc. ICEST 2017 – LII International Scientific Conference on Information, Communication and Energy Systems and Technologies*, P28.6, pp. 257-260, Nis, Serbia, 28-30 June, 2017.

[17] Web site http://www.hp.com; HP V-M200 802.11n access point management and configuration guide; 2010; accessed 3 Jan 2018.

[18] Web site http://www.alliedtelesis.com; AT-8000S/16 level 2 switch technical data; 2009; accessed 10 Dec 2016.

[19] Web site http://www.linksys.com; WPC600N notebook adapter user guide; 2008; accessed 10 Jan 2012.

[20] Acrylic WiFi software; 2016; http://www.acrylicwifi.com; accessed 8 Jan 2018.

[21] Iperf software; 2016; http://iperf.fr; accessed 16 Feb 2018.

[22] Network Working Group. "RFC 1889-RTP: A Transport Protocol for Real Time Applications", http://www.rfc-archive.org; 1996; accessed 13 Apr 2018.

[23] P. R. Bevington, Data Reduction and Error Analysis for the Physical Sciences, Mc Graw-Hill Book Company, 1969.

Design of microstrip phased array for autonomous flying vehicles

Peter Z. Petkov[1], Rosen Vitanov[2] and Ivaylo Nachev[3]

Abstract – **In this paper considers patches designed for a doppler radar allowing him enchanced coverage and beam-steering capabilities is presented. Design of antennas is based the idea of integrating them in HB100 module, provides small physical package, low price and high reliability, makes it possible for a autonomous flying vehicle to autodetect obstacles in flight without need of changing position or orientation.**

Keywords – **radar, antenna, patch antenna, patch array, phased array, transmission line phase shifter, quadcopter, autonomous flying vehicle**

I. INTRODUCTION

We are witnessing a technological boom, where more and more automatic devices do complex operations and services in areas, recently possible only for humans. Autonomous flights by unmanned air vehicles (UAV) is one of those. The great challenge is to detect and overcome and obstacle in midflight, without being interrupted. For this purpose – a microwave radar is used with enhanced range and integrated beam-steering functionality is implemented. In this way UAV may test the upcoming trace for any obstacles and avoid collisions during a flight.

Most antennas have a fixed radiation pattern, which is a inconvenience, when used in drones, because it should stop in one place and test the area by turning around using its propellers. A practical solution to this problem is usage of array antenna with scan.

In this case most appropriate is to use a series patch array. It may sweep the direction of the beam by varying electronically the phase of the radiating element, thereby producing a moving pattern with no moving parts.

II. CORPORATE FEED NETWORK PATCH ARRAY DESIGN

An important step when designing patch antenna is to take into account dielectric substance on which patches are made. The employed substrate for the proposed design is Rogers RO 4003 with thickness h=0.508mm and Dissipation factor tanδ = 0.0027**.** In order to improve the bandwidth and efficiency of antennas, a design method called "suspended substrate" is chosen, where the thickness of the gap between ground plane and antenna is Δ=1mm. In this method we use the equivalent Dielectric constant (7). Distribution of the elements of the simulated model, as with radars frequency f_c = 10.525GHz we have wavelength λ = 28.50mm, and λ_g = 22.6150mm. For calculation of antenna dimensions a straightforward basic algorithm is proposed:

Step 1: Calculation of width (W): Width of the Micro strip antenna is given by the equation:

$$W = \frac{c}{2fc}\left(\frac{\varepsilon_r + 1}{2}\right)^{-\frac{1}{2}} = 12.2208mm, \frac{W}{h} > 1 \quad (1)$$

Where c is the speed of the light c=3*10^8, ε_r is a dielectric constant of substrate, f_c is radar frequency and h is substrate thickness.

Step 2: Calculation of Effective Length (L): Length of the microstrip radiator is given by:

$$L = \frac{c}{2fc\sqrt{\varepsilon_r}} = 11.3mm \quad (2)$$

Where f_c is radar frequency and ε_{eff} is effectively dielectric constant of substrate

Step 3. Calculation of the Length extension (L_{ext}): The actual length is obtained by the equation (accounting for the fringing fields):

$$L_{ext} = 0.412h\frac{(\varepsilon_{eff} + 0.3)(\frac{W}{h} + 0.264)}{(\varepsilon_{eff} - 0.258)(\frac{W}{h} + 8)} = 0.8825mm \quad (3)$$

Step 4: Calculation of the actual length of the Patch (L_{eff}): The actual length of the patch is given by:

$$L_{eff} = L + 2L_{\exp} = 12.9650mm \quad (4)$$

Determination of effective dielectric permittivity ε_{eff}:

$$\varepsilon_{eff} = \frac{\varepsilon_r + 1}{2} + \frac{\varepsilon_r - 1}{2}\left(1 + 10\frac{h}{W}\right)^{0.5} = 1.5886 \quad (5)$$

The length of the guided wave:

[1] Peter Z. Petkov is Associate Professor with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: pjpetkov@tu-sofia.bg.

[2] Rosen Vitanov is PhD student with the Faculty oft Telecommunications at Technical University of Sofia, 8 Kl. Ohridski, Blvd,Sofia 1000, Bulgaria, E-mail: rosenvitanov@gmail.com

[3] Ivaylo Nachev is a student at Faculty of Telecommunications att Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000,, Bulgaria, E-mail: ivaylonachev@yahoo.com

$$\lambda_g = \frac{\lambda}{\sqrt{\varepsilon_{eff}}} = \frac{c}{fc\sqrt{\varepsilon_{eff}}} = 22.6150mm \quad (6)$$

Influence of the gap is taken into account:

$$\varepsilon_{r\,eq} = \frac{\varepsilon_r * (h + \Delta)}{\varepsilon_r + h * \Delta} = 1.72 \quad (7)$$

Where ε_r is a dielectric constant of substrate and Δ is thickness of the gap. The new dielectric constant in technology with suspended substrate is marked with $\varepsilon_r eq$, where the characteristics of the air are also taken into account.

Connection between radiating antenna and phase delay line control system is achieved using SMA connectors with imedance Z_0=50Ω which is low enough to not interfere with antenna`s parameters.

In table 1 are presented the values obtained for dimensions of the antenna components. Fig.1 shows simulation model of corporate feed network patch array. [1] [2][3]

TABLE 1
VALUES OBTAINED FOR DIMENSIONS OF THE ANTENNA COMPONENTS

	W [mm]	L_{eff} [mm]
Calculated	12.2	12.96
After simulation	13.8	11

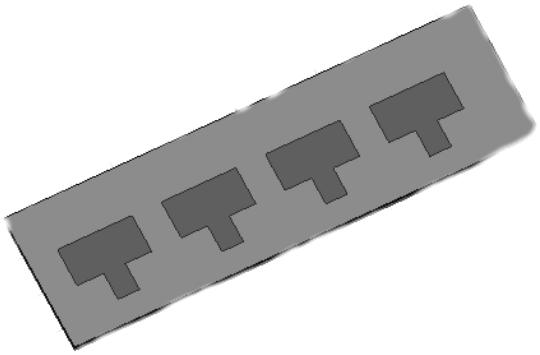


Fig. 1 corporate feed patch array – simulation model

III. TRANSMISSION LINE PHASE SHIFTER

Described phased antenna array uses a hybrid T junction system that divides the input power in 4 equal parts. In this paper is described how - by changing electric length of microstrip lines (ML), changes the phases of fed signals, thus steering the beam (fig.2). It is required to change the position of the beam in 3 different degrees: -15; 0; +15 deg. page.

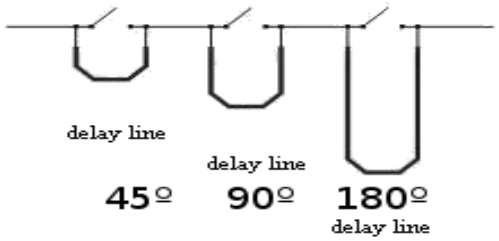


Fig. 2. Switched line phase shifter

In the switched-line phase shifter, phase shifting is obtained by varying the length of the antenna feed line. The phase delay in the feed line is calculated from:

$$L_{effective\ length} = \frac{\beta c}{2\pi f\sqrt{\varepsilon_{eff}}} \approx 2.877mm \quad (8)$$

$L_{effective\ length}$ - Change in length;
β - Propagation constant;
c - Speed of light;
ε_{eff} - Effective dielectric constant;
f - Operating frequency;

Formula (8), leads to conclusion that phase differences between each patch should be 2.877mm, and using electromagnetic simulator equals 45 deg. That's confirm from simulations in the design process Simulation showed best results with each phase line doubled the length on the previous. This is shown in table 2. [4][5]

TABLE 2
RELETION BETWEEN BRAM STEERING AND PATCH PHASES

Bean steering	Patch1	Patch2	Patch3	Patch4
0^O	0^O	0^O	0^O	0^O
$+15^O$	0^O	45^O	90^O	180^O
-15^O	180^O	90^O	45^O	0^O

IV. PHASE DELAY LINE CONTROL SYSTEM

On figure 3 a graphical relation between radiating antenna, feed, pahse delay lines and hybrid T junctions is shown. Building such a system requires input power to be divided in 4 eaqual parts, described below. Initially, the signal „P" enters at the input point (1). Then, using a 50 Ohm microwave line (2) is passed to the first T junction (3), designed to divide the energy into two equal parts „P / 2". This is accomplished by two 100 Ohm MLs(4). Then the impedance is again converted to 50 Ohm (6) by step – work like impedance transformer (5). The energy separation is repeated again by a second group of f T junctions (7) to 100 Ohm microwave liness (8), the final result is „P / 4". Using ML (9), the impedance is again lowered to 50 Ohms and „P / 4" passes through delay line (10) to the radiating element. Beam steering is accomplished by switching

betwen microstrip lines with different lengths , using PIN diodes. This leads to required beam scanning (tabl. 2).

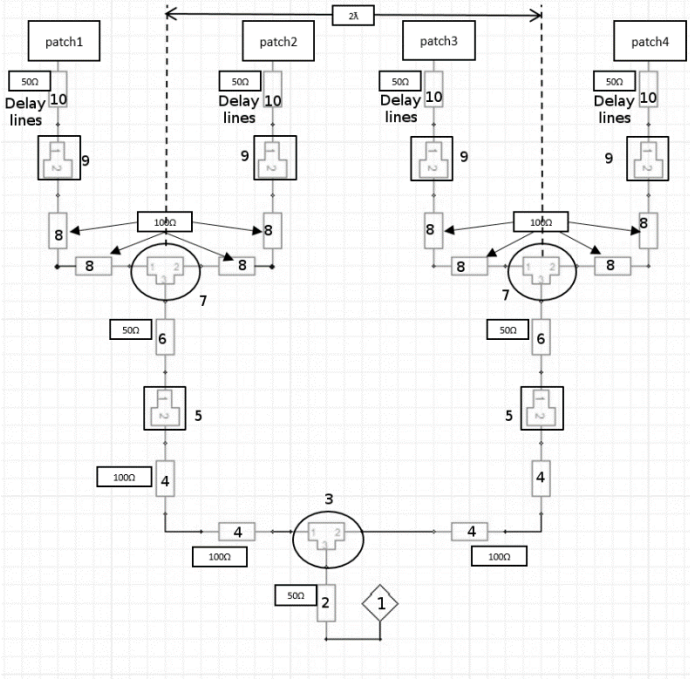


Fig. 3. Block diagram of radiating antenna

V. SIMULATIONS RESULT

Shown are results from simulations. In fig. 4 is displayed the return loss of the antenna. Figure 5, shows elevation radiation pattern. Figs. 6a, 6b,6c show represent the beam in azimuth, steered a the three available angles – 0 deg, -15 degs and +15 degs. In table 3, a comparison of gains in beam`s different state.

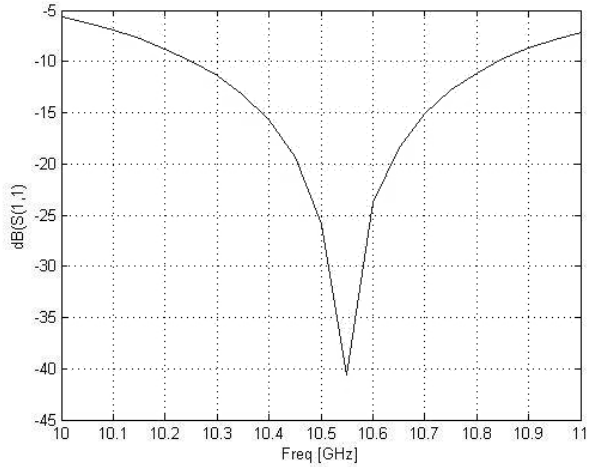


Fig. 4. Antenna bandwidth (S11)

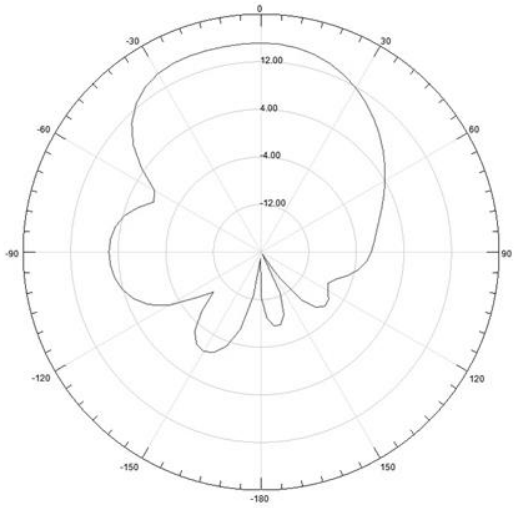


Fig. 5. Radiation Pattern elevation

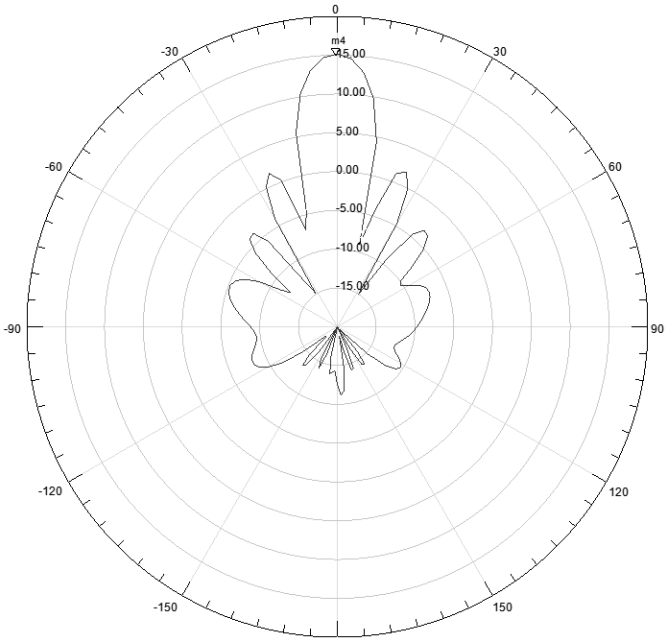


Fig. 6a. Radiation Pattern azimuth 0deg.

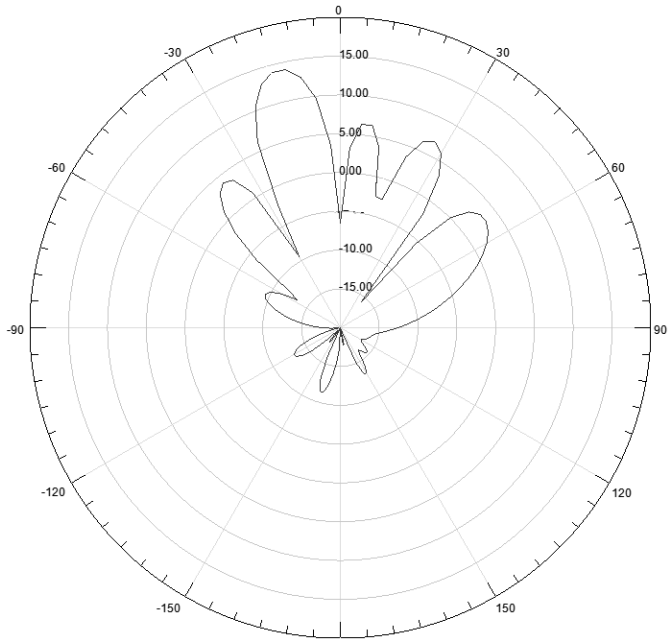


Fig. 6b. Radiation Pattern azimuth -15deg.

.

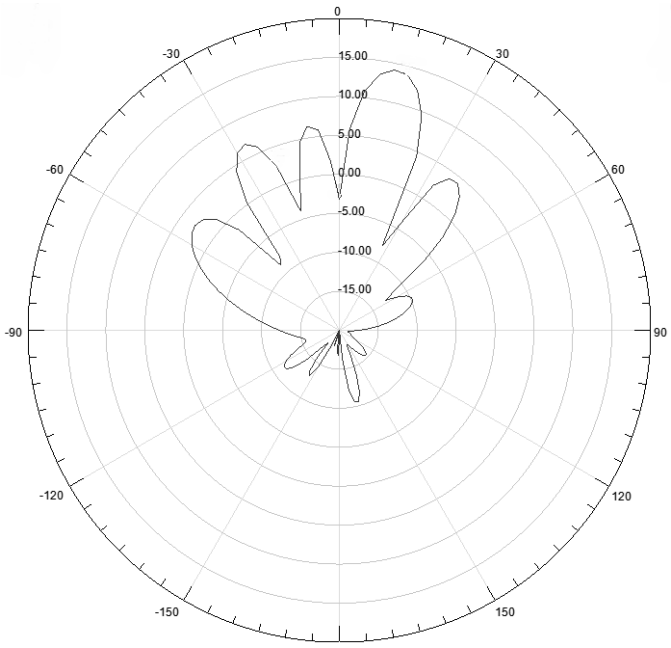


Fig. 6c. Radiation Pattern azimuth +15deg.

TABLE 3
COMPARISION OF GAINS IN BEAMS DIFFERENT STATE

	Gain[dB]
Patch array 0deg.	15.1
Patch array -15deg	14.7
Patch array +15deg	14.67

VI. Conclusion

Using microwave delay line phase shift in systems such as considered allows an electronic scan of the antenna beam. Integrated antennas in the HB100 module can be used for automatic and non-intrusive obstacle detection system for quadcopters and other unmanned vehicles.

References

[1] HB100 datasheet. https://www.limpkin.fr/public/HB100/HB100_Microwave_Sensor_Module_Datasheet.pdf.

[2] Chang, Kai; "RF and Microwave Wireless Systems", 2000, ISBN: 0-471-35199-7.

[3] Rogers RO4003 specification: https://www.rogerscorp.com/documents/726/acm/RO4000-Laminates---Data-sheet.pdf.

[4] Akhtar A., Mateen Alahi H., Sehnan M., ""Simulation of Phased A rrays with Rectangular Microstrip Patches on Photonic Crystal Substrates", 2012.10.12

[5] J. Ehmouda, Z. Briqech, and A. Amer, "Steered Microstrip Phased Array Antennas", World Academy of Science, Engineering and Technology, issue: 49, 2009, p. 321 - p. 323

Mobile system for monitoring of air quality and gas pollution

Rosen Miletiev[1],Emil Iontchev[2], and Rumen Yordanov[3]

Abstract - **The current paper represents the mobile system for monitoring of air quality by detecting the concentrations of most of the industrial pollutants such as ammonia, carbon oxide, carbon dioxide, volatile organic compound and nitrogen oxides. The presence of these gases in the outdoor or indoor air is very important for the human health and safety because some of them may be dangerous in high concentrations. This portable system is designed to meet the requirements of the health protection in urban environments.**

Keywords – **air quality, gas pollution, wireless communications**

I. INTRODUCTION

Continuous air pollution component measurements for precursor gases, ions, PM and carbon dioxide have been available since the late 1990s. Pollutants for which air quality standards have been established in many countries (called "criteria pollutants" under the statute defined by the US EPA, i.e. CO, NO_2, SO_2, ozone, $PM_{2.5}$ mass, PM_{10} mass, TSP, and lead) are monitored in populated areas to determine compliance with the ambient air quality standards (AAQS) [1]. Most of the measurements systems detect only one of the possible air pollutants while it is very important to scan all very poisonous gases such as NO_x, SO_2, CO and NH_3.

Reliable measurements of atmospheric trace gases are necessary for safety reasons because some of them are very toxic (NO_x) [2], indetectable to humans, being both tasteless and odorless (CO) [3] and irritating in small concentration or in high concentrations is an immediate hazard to life (NH_3) [4].

The mobile measurement of the air quality and gas concentrations is a very actual problem due to the great human mobility, pollution motion and the absence of measurement stations in most places especially for the closed spaces. The requirements to the proposed systems are based on the device functionality such as small size, battery powered with recharge capabilities, data filtering and analysis. Also the system have to combine different types of communication devices such as short-range and long-range ones to ensure the continuous data transmission and long-term operation time during the sensor motion and data acquisition. These requirements are fulfilled by using the latest communication chips which combine GSM and Bluetooth connectivity, integrated GPS receiver, dual SIM card integration and chip antennas.

[1] 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

[2] Emil Iontchev is with the Higher School of Transport "T. Kableshkov" 158 Geo Milev Street, Sofia 1574, Bulgaria, E-mail: e_iontchev@yahoo.com

[3] 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

II. SYSTEM DESCRIPTION

The proposed measurement and communication system has to fulfill the following requirement:

1/ measurement of the poising gases such as carbon oxide (CO), ammonia (NH_3), nitrogen oxides (NO_x)

2/ measurement of air quality respect to the concentration of carbon dioxide (CO_2) and organic compounds (tVOC)

3/ battery powered device with low current consumption and recharge possibilities via common used mini USB interface

4/ measurement of the device position via GPS service

5/ send the air quality and navigation data to the remote server via Bluetooth or GPRS link

The mobile device consists from two main blocks (Figure 1) – navigation and communication board, which includes GPS receiver, GSM/GPRS modem and Bluetooth transceiver, and measurement board, which integrates the sensors - Gas sensors, air quality (AQ) sensor or particle measurement (PM) sensor and user indication.

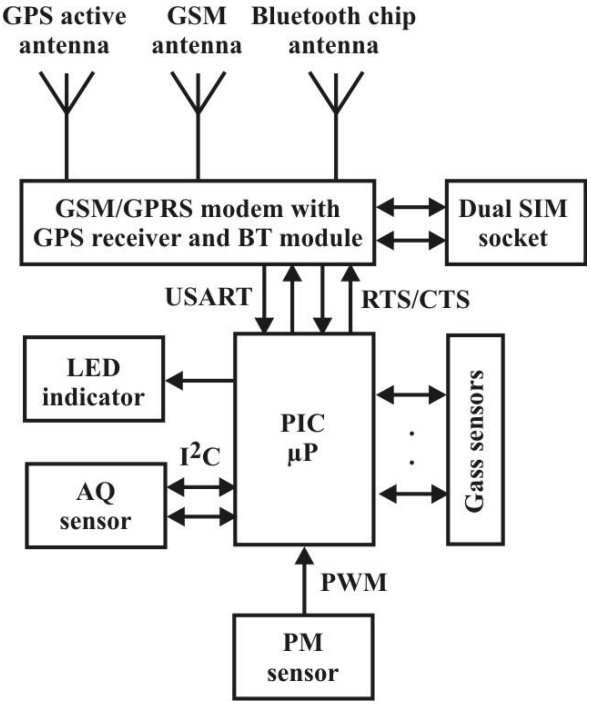


Fig.1. Block diagram of the system

The system connectivity is ensured by MC60 as GSM/GPRS modem which supports Bluetooth interface, it is fully compliant with Bluetooth specification 3.0 and supports SPP

(Serial Port Profile). It also features GPRS class 12 and is distinguished with an integrated TCP/UDP, FTP and PPP protocols. The current consumption is as low as 1.3 mA in SLEEP mode. The portable size of the system and the best network coverage are provided by the integrated chip-antenna WE-MCA 7488910245 for the frequency range 2400-2500MHz (Bluetooth module) and external GSM antenna via female SMA connector. The chip antenna peak gain and average gain are equal to 3dBi and 1.0dBi respectively. The system microcontroller is based on PIC16F family (PIC16F1825) and has built-in I^2C interface as a part of Master synchronous serial port (MSSP) module and EUSART with an autobaud interface, which is connected to the GSM/GPRS modem with RTS/CTS hardware flow control. The microcontroller oscillator scheme uses SMD (QFN) crystal at 16.000MHz as a low-profile resonator. The frequency tolerance is equal to approximately ±15ppm@25ºC.

The power supply (Figure 2) is based on LDO (low-dropout) regulator BA00DD0WHFP with adjusted output voltage which may set in the range from 3.8V to 4.2V to power the GSM/GPRS modem. This chip has a high-precision output voltage of ±1% and may deliver up to 2A with maximum dropout voltage of 0.5V. It also is distinguished with a built-in over-current protection, over-voltage protection and thermal shutdown circuit. The LDO regulator is chosen to allow an USB charging.

The output voltage of 4.2V is additionally used to power LDO chips for the microcontroller, sensor board and GPS receiver. These LDO are recognized as MCP1700T-3.0, MCP1700T-3.3 and TC1224 respectively. The LDO chips are high accuracy (typically ±0.5%) CMOS low dropout regulators and are designed specifically for battery-operated systems, while the CMOS construction eliminates wasted ground current, significantly extending battery life. The TC1224 regulator is distinguishes with an enable input which is controlled via AT command by GSM/GPRS modem to on/off GPS receiver to save power if GPS module is not needed or the device is stationary.

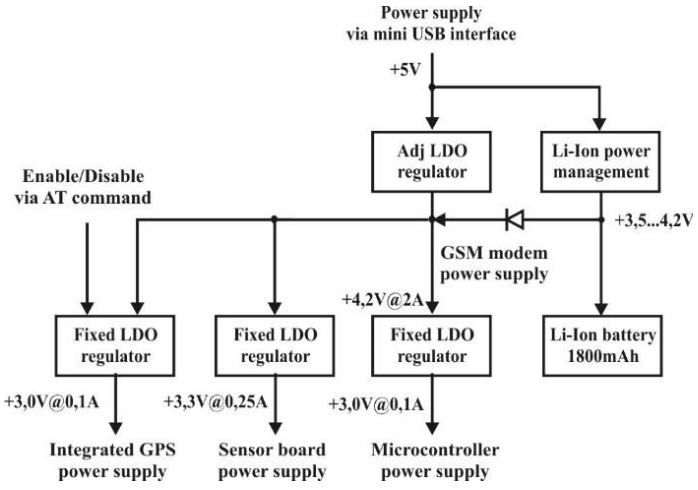


Fig.2. Power supply schematic

There chargeable Li-Ion battery is used as a backup power supply and is supported by MCP73831T-2ACI/OT ICs as a Li-Ion charge management controller. The charging current is set via an external resistor and its maximum value is equal to 500mA. The power management chip employ a constant-current/constant-voltage charge algorithm and charge termination.

The sensor board contains the gas sensors such as MiCS-6814, which is the compact MOS sensor with three independent sensing elements for ammonia, carbon oxide and nitrogen oxides on one package. Each sensor includes a heating circuit and sensing element (Figure 3). The detecting layer changes its resistance according to the pollutant concentration. The sensing resistance in air varies from 10 to 1500kΩ for the NH_3 sensor, therefore the sensor requires a calibration to measure the real concentrations.

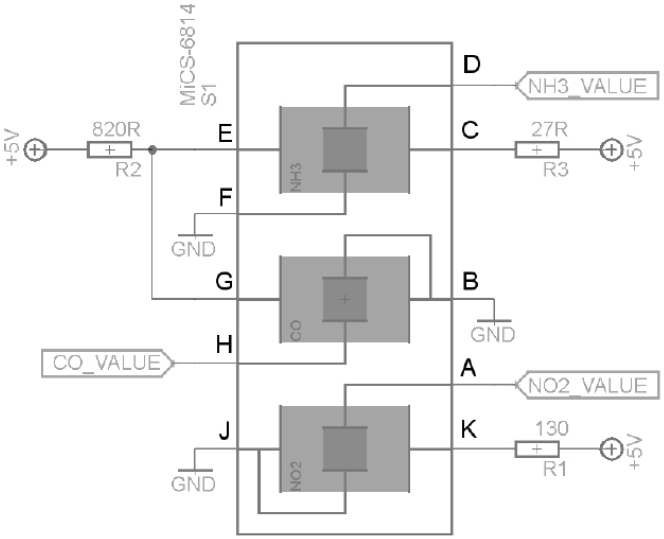


Fig.3. Gas sensor [5]

The sensor board also contains the air quality (AQ) sensor which is recognized as a MiCS-VZ-89TE of the same company (Figure 4). It is a digital output sensor for carbon dioxide (CO_2) and tVOC concentrations. It has two independent digital outputs – I^2C and PWM.

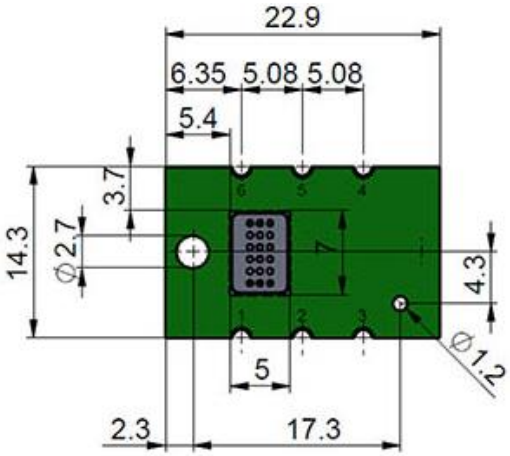
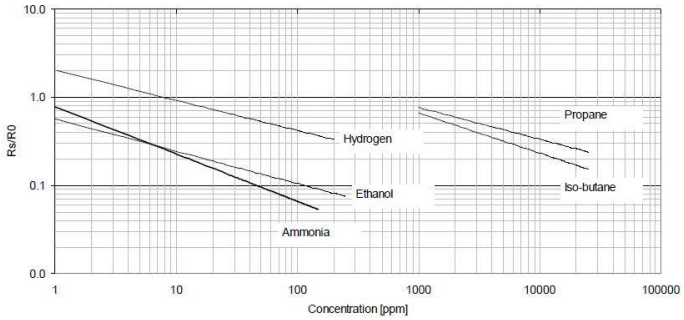


Fig.4. Air quality sensor.

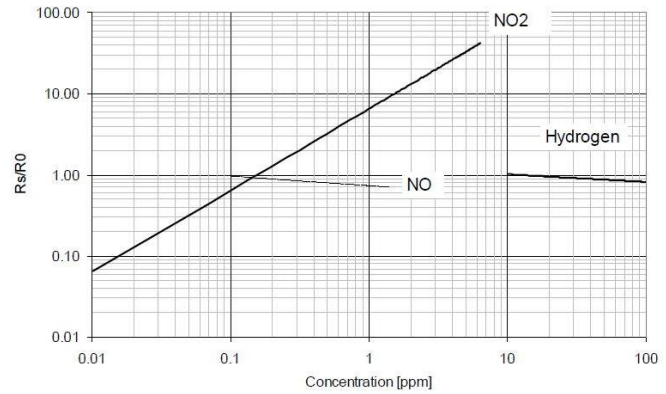
III. EXPERIMENTS

The first step of the experiment is the sensor calibration due to the high range of the sensor resistance in air. The calibration is made in the environment of a pure nitrogen at normal temperature by measuring the sensing resistance R_0. When the sensor is exposed to the air the resistance is change to Rs value which corresponded to the gas concentration according to the Rs/R_0 to Concentration graphics (Figure 5) [5]. The sensing resistance Rs is converted to voltage using resistor divider and then the integrated 10-bit ADC converts this voltage to a digital word. The second resistor in the divider is chosen to be equal to R_0, measured in the calibration step, to guarantee the maximum amplitude response to the gases.

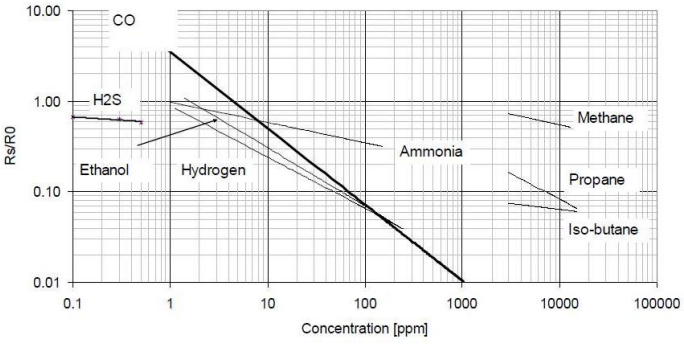
The sensing resistance may be affected by the temperature, humidity and time [6], so the calibration procedure have to be accomplished if an optimum accuracy is required.



NH3 sensor, continuous power ON, 25°C, 50% RH



OX sensor, continuous power ON, 25°C, 50% RH



RED sensor, continuous power ON, 25°C, 50% RH

Fig.5. Ratio Rs/R_0 as a function of concentration [5]

According to the manufacturer curves [5], we calculate the gas concentration by the equation lgY=a*lgX +b, where X and Y denote the gas concentration in ppm and Rs/R_0 value respectively and a and b are constants. We establish the following constants for the dependence of the Rs/R_0 value towards the concentration:

1/ CO sensor: a = -0,82709; b = 0,526057

2/ NO_2 sensor: a = 1.032031; b = 0,831015

3/ NH_3 sensor: a = -0,50084; b = -0,1728

The CO_2 and tVOC concentrations are established via I^2C interface according to the equations [7]:

tVOC[ppb]= (tVOC- 13) * (1000.0 / 229)

CO_2[ppm] = (CO2 - 13) * (1600.0 / 229) + 400

The experiments are made in the indoor situations and the concentration of the carbon dioxide (CO_2) and volatile organic compound (tVOC) are shown at Figure 6 while the gas sensor data for NH_3, CO and NO_x as a ADC word and concentration in ppm are given at Figure 7..

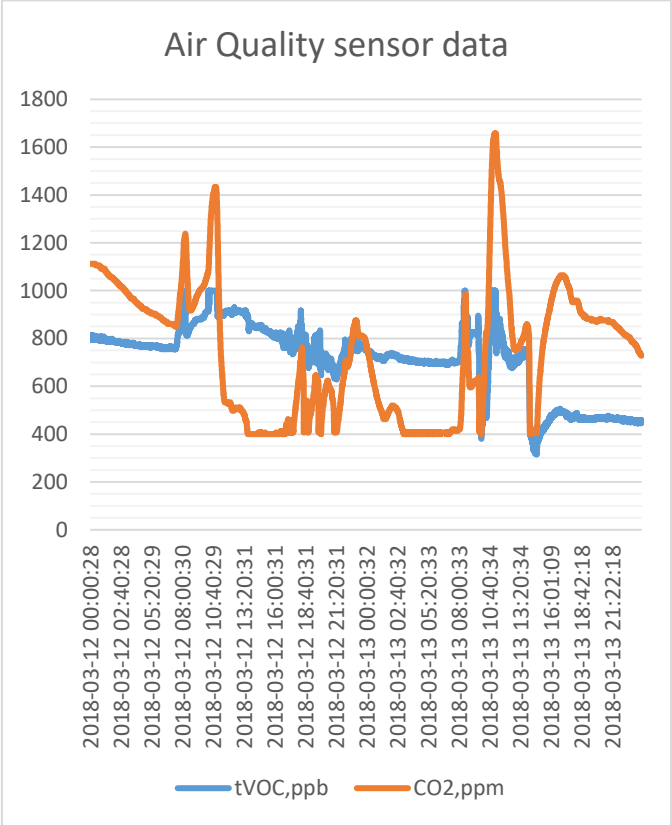
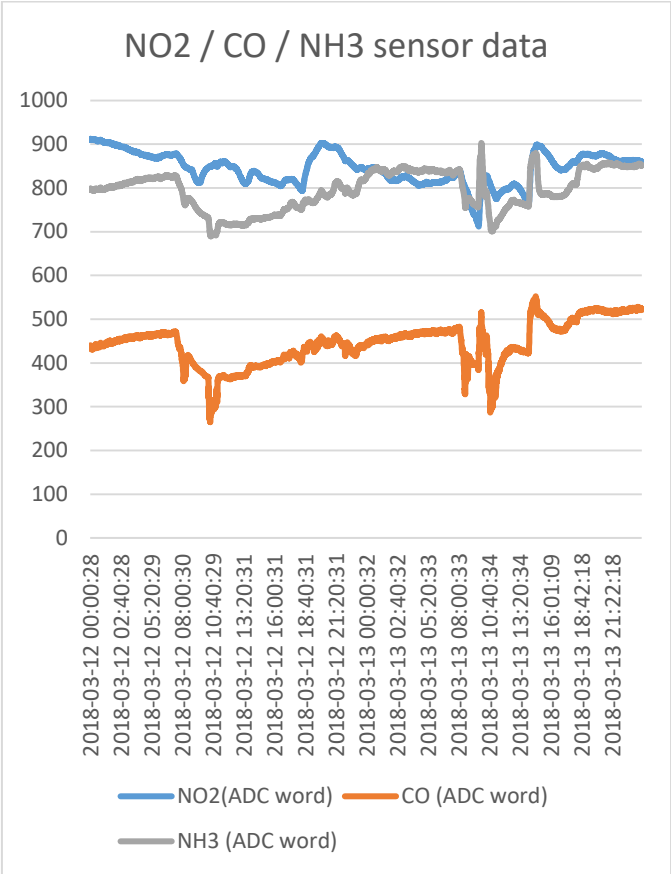


Fig.6. Air quality sensor data

The data represent a continuous 48 hour measurement of the air quality and gas concentrations of all gas sensors. The represented data shows that NH_3 concentration is under 25ppm limit (Detectable by smell. Maximum Permissible Exposure Limit (PEL)), CO concentration also is under the maximum recommended indoor CO level of 9ppm (ASHRAE) [8] and the NO_2 value is approximately equal to 0 ppm.



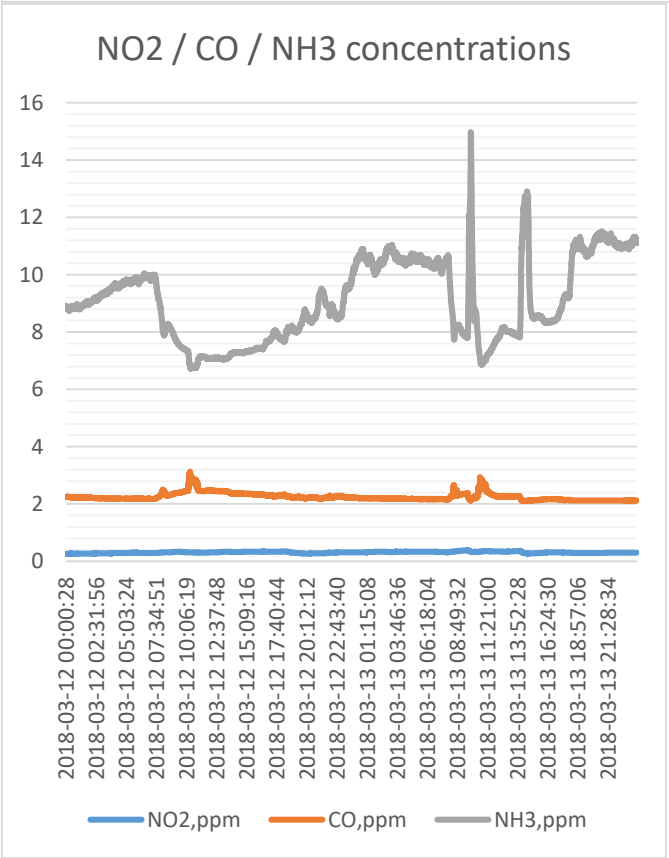


Fig.7. Gas sensor data for NH_3, NO_x and CO as ADC word and concentration ppm

IV. Conclusion

The proposed mobile system is capable to measure most of the poisonous gases such as carbon oxide, ammonia and nitrogen oxides by three independent sensors in one package and the air quality as a function of the carbon dioxide and volatile organic compound concentration. The air quality sensor may be replaced or combined with PM2.5 and PM10 sensor, which is the goal of the future system development.

References

[1]. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, No. 109 - Outdoor air pollution, 2016
[2]. Jorg Kleffmann, Guillermo Villena Tapia, Iustinian Bejan, Ralf Kurtenbach, and Peter Wiesen - NO2 Measurement Techniques: Pitfalls and New Developments, Chapter 2, Springer, 2016
[3]. Mehdi Fazlzadeh, Roohollah, Rostami Sadegh, Hazrati AliRastgu - Concentrations of carbon monoxide in indoor and outdoor air of Ghalyun cafes, Atmospheric Pollution Research, Volume 6, Issue 4, July 2015, Pages 550-55
[4]. Al Heber, D. Jones and A. Sutton. 2002. Controlling Ammonia Gas In Swine Buildings, February 2004
[5]. SGX Sensortech – MiCS-6814 datasheet 14143 rev 8
[6]. SGX Sensortech - SGX Metal Oxide Gas Sensors, AN-0172, Issue 1, 14-Jul-2014
[7]. SGX Sensortech - MiCS-VZ-89TE datasheet
[8]. Thomas H. Greiner, Extension Agricultural Engineer Department of Agricultural and Biosystems Engineering, Iowa State University, August, 1997

CONTROL SYSTEMS AND REMOTE ECOLOGICAL MONITORING – POSTER SESSION

Fire safety and fire dynamics of a school room simulated in FDS computer program

Svetlin Antonov[1] and Borislav Borisov[2]

Abstract – **Fire safety and fire dynamics are very important in these days. Building constructions and planning must be up to the requirements of different laws in force and instructions. They are in result of the scientific exploitation and human experience for ages. This paper present such treatment of a real building case in a Bulgarian high school located near to the Technical university in Sofia. A fire arise out of the chemical laboratory which is located in the underground floor of the school. A PYROSIM (FDS) program is used to simulate the fire dynamics- fire arise, flame and smoke movement and the even self-extinguish. The reason for this result is the usage of a new woodwork which helps the fire to be extinguished because the room is run out of oxygen if the door and windows are closed. The temperature on the premises is under the level of breaking the glass, which would lead to oxygen income. The result helps to predict the situation of fire calamity and the time for evacuation. This method may be used in different buildings with different sizes and different building plans..**

Keywords – **Thermal impact, Fire dynamics, Fire safety, PYROSIM, School building.**

I. DANGEROUS FIRE FACTORS

In case of fire, there are enough hazards that force people to leave the room, often taking the decision independently and spontaneously.

It is known that in most cases the time for which dangerous fire factors acquire critical meanings (so-called critical evacuation time) is less than the time required for the arrival of fire departments.

Smoke is a mixture of combustion products and particulate matter with the air. The composition of the smoke depends on the composition of the burning substances and the combustion conditions. When burning organic materials (wood, straw, paper, gas, etc.) in the composition of the smoke may be carbon monoxide, water vapors, carbon dioxide and soot. In the thermal decomposition and burning of plastics, rubber, synthetic fibers and others in the composition of the smoke is found nitric oxide, hydrogen chloride, hydrogen sulphide, sulfur dioxide, phosgene, silicic acid and other toxic substances. The concentration of these products in smoke depends on the intensity of the gas exchange and the quantity of these products that are separated from the 1 m2 burning area. In the case of internal fires due to incomplete combustion, the amount of these products is greater than in the case of fires in the open, where full combustion occurs. The smoke density is determined by the amount of solid particles in unit volume and is measured in gr / m3. Density (smoke rate) can be regulated by gas exchange. The high density complicates the orientation and actions of the fire departments in the extinguishing of the fires. Products of incomplete combustion (carbon, carbon monoxide, etc.), which are in the composition of smoke mixed with air, can create not only combustible but also explosive mixtures. The high concentration of combustion products in the smoke composition reduces the percentage of oxygen. At 14-16% of oxygen in the air there is oxygen starvation, and at 9% there is a danger to life. (Under normal conditions, the amount of oxygen contained in the air is 21% by volume.), The density and the toxicity of the smoke depend on the properties of the burning material and on the conditions in which the combustion process takes place. The volume of products that are released during combustion of the burner is called smoke formation.

Combustion of organic substances and materials (wood, paper, gasoline, etc.) in the composition of the smoke contains carbon monoxide, carbon dioxide, water vapor and soot. In thermal decomposition and combustion of plastics, synthetic rubber, synthetic fibers, etc. in the composition of the smoke is found: nitrogen oxide, hydrogen chloride, hydrogen sulphide, sulfur oxides, sialic acid and other toxic substances.

The concentration of these products in smoke depends on the smoke intensity and the quantity of these products, which is separated from one m2 of combustion area. In internal fires, due to incomplete combustion, the amount of these products is greater than in the case of open fires where the combustion of the substances is complete.

Smoke concentration is the amount of combustion products contained in a unit of the room volume (gr/m^3, gr/l or in volume percentages).

The high concentration of internal combustion products and the combustion process are the reason for reducing the percentage of oxygen.

Heat load is the thermal energy expressed in SI units, which is released by the burning of all materials in a given space, including the lining of the walls, partitions, floors and ceilings.

The self-evacuating process of heat transfer in a space with a non-uniform temperature field is called heat exchange.

Many heat transfer processes are accompanied by bulk transfer. The cooperative process of heat and mass exchange is called heat exchange. The sum of the instantaneous temperature values at all points of the considered space / body (coordinates x, y, z) at time t is called the temperature field.

Each temperature can be expressed either in degrees Celsius [°C] or in Kelvin [K], since in the Celsius and Kelvin

[1] Svetlin Antonov is with the Faculty of Telecommunications at the Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria. e-mail: svantonov@yahoo.com.

[2] Borislav Borisov is with the Directorate Fire Safety and Civil Protection at the Ministry of Interior, Sofia 1000, Bulgaria. e-mail: b.borisov8788@gmail.com.

rocks the linear dimensions corresponding to one degree are the same.

The temperature field can be characterized by isothermal surfaces - a geometric location of points with the same temperature at a given time. Isothermal surfaces corresponding to different temperatures cannot cross. They are either closed or ending on the surface of the body. When crossing a plane (for example, in a drawing), isothermal surfaces leave traces in the form of families of curves - isotherms.

The fire is a collection of time-consuming processes. The dynamics of the phenomenon and the impacts on structures in case of fire are described by the non-stationary heat exchange equations.

The density of the heat flow is proportional to the temperature gradient. The energy transfer is always from a higher location to a place with a lower temperature and continues until equilibrium is reached [1].

In the case of a fire, a process of non-stationary heat transfer from a moving environment (gas) to a solid medium (building structure) through a separating surface is realized. Temperature influence on a structure for each point of fire development in a building is quantified by the density of the heat flow to the surface of the elements (the amount of heat passing per unit of time over a unit of area of an isothermal surface.

Temperature mode is the change in the average volume temperature over time [1]. Normalized temperature modes represent either a steadily increasing time function (continuous fuel delivery, eg standard fire) or an increase followed by a constant stretch (hydrocarbon and external fire curve). There is still no international technical consensus on the issue of harmonizing the conditions for thermal radiation in the implementation of standardized temperature regimes. In the conditions of a free (non-extinguishing) fire in buildings, the temperature rises, after reaching the maximum for a short time, it holds approximately a constant value, followed by a drop due to the burning of the fire load. The standard temperature curve adopted for international scale (ISO 834) (for Bulgaria BDS 6316-81) simulates the period of active combustion of fire in a room containing solid combustible materials with a calorific value Hu = 20,9 [MJ / kg] and a mass burning rate equal to $13.9.10^{-3}$ [kg / (m^2.s)]. The idea of a standard fire curve arises in 1903, the computation formula was proposed by Briton S.N. Ingberg in 1928, in 1961 was adopted by ISO. Mathematical models for the determination of temperature regime of fire in buildings provide the possibility to predict the change of the dangerous factors of the real object phenomenon and the transition from a comparative assessment of the fire hazard of the building materials / structures (based on the standard fire) to its forecasting of the actual operating conditions . Attempts for a mathematical description of fire behavior date back to 1737 and continue to this day due to the complexity of the physico-chemical processes accompanying the phenomenon.

Knowledge of fire regulations in premises is the basis for solving issues related to the fire resistance of building structures, providing evacuation, expert assessment of buildings, etc. Integral analysis of the fire by experimentally determined aggregate characteristics provides sufficient accuracy for building practice. Differential analysis of the phenomenon is more detailed but requires the application of equations of the solid environment mechanics (including the rheological law of Stokes), the Fourier heat law, the laws of diffusion and radiation transmission in the gaseous environment, etc. The system of equations describing the change in density, pressure, temperature and composition of the gaseous medium in time / space is complex / difficult to solve [2].

The normative fire load is the sum of the temporary (burning content of the temporary load in buildings) and constant (the combustible content of the building elements, including linings / coatings) [2].

In the norms of Switzerland, Great Britain, Sweden, France, etc., as well as in the European Union Construction Design Guide, the classification of the thermal potential [MJ/m^2] of the normative temporary fire load, determined according to the functional purpose of premises.

II. DYNAMICS OF THE INTERNAL FIRE PHENOMENON IN NON-HERMETIC BUILDINGS

The development of the fire involves spreading the burning process on the fire load and changing the parameters of the phenomenon in space and time. A good picture of the free development of a fire in a room is as follows. After combustion occurs at a point in the volume of the room, flame propagation on the surface of a burning material causes a sequence of continuous flames of the gases / vapors released from the material. In the initial period the dimensions of the area of the fire are small and the combustion develops at the expense of the air in the local volume.

After heating the surface of the enclosing structures begins to radiate heat. The floor temperature increases from the thermal conductivity. The ceiling and the walls are heated by the convective gas streams, a convective gas jet is formed above the hearth. Gas exchange between the combustion zone and the air from the room volume is established. The area and the thermal impact of the fire are increasing.

On combustion, the substances emit a volume of gas larger than the volume of air required to burn per unit mass. The pressure in the room increases, the glazing is heated and destroyed (at an average temperature of 480 - 580 ° C). The airflow increases sharply, air exchange is established with the outside environment or adjacent premises, the fire parameters are intensified.

The development of the fire accelerates to a certain limit, then ceases due either to the exhaustion of the fuel or to the insufficient speed of the exhaust air.

In the period of active combustion, 70 - 80% of the fire load is absorbed and the maximum gas temperature is reached.

An important role in studying gas exchange processes is the knowledge of the fluid dynamics.

In case of a fire, it is necessary to consider a number of additional factors that affect the gas exchange: the weight of

the burned material, the rate of change of the room temperature, etc. It is also known that in small sections of the inlet and inlet openings, and at a high rate of increase in room temperature, overpressure may occur in relation to the outside air. These features must be known because of their great practical significance.

III. COMPUTER MODELLING OF FIRE

PyroSim - is the user interface of Fire Dynamics Simulator (FDS) software [3]. FDS model can predict the spread of smoke, temperature, carbon monoxide and other hazards during a fire.
The results of the simulation are used to ensure the safety of the buildings in the design, to determine the safety of existing buildings, to reconstruct fires in the investigations, and to assist in the training of firefighters.

Fire processes are very complex and complex phenomena consisting of combustion, radiation, strong turbulence, processes, and is related to physical and chemical processes as well as phase transformations.

Contemporary requirements for the occurrence and development of fires are based on numerical simulations. They allow, using minimal means, to anticipate the course of the fire, resp. Smoke formation on the building volume, the increase of the combustion during the fire, including eventually the development.

The use of computer modeling of fires and precise fire protection analyzes has become increasingly applicable in recent years, with prognostic capabilities covering a wide range of applications. In general, basic algebraic ratios can be included in modeling and simulation of fire to determine model zone parameters and computational fluid dynamics (field models) that are used to predict or report different fire-fighting phenomena within established set of boundary conditions.

In recent years, computer fire simulation has been used as the economically cheapest method to study the processes of fire development and their visualization. This approach is particularly applicable in the case of fires in tunnels, car parks and buildings, because full-scale ground fire experiments in these structures can cause serious damage to the material and technical base. Nowadays, the development of the fire environment allows for a relatively good knowledge of fluid and gas dynamics.

Existing software tools provided by CFD simulators also allow them to visualize their development. The use of (fluid mechanics) of theoretical CFD and its practical knowledge has a wide application in aerodynamics, fluid dynamics, combustion engineering and other areas. CFDs have been applied over the last decades. In these models, the field of fire is divided into separate fire zones, so each of these current processes is self-contained. The theoretical basis of these methods consists of the laws of mass and energy conservation. The whole space is divided into two spatially homogeneous zones: the warm upper volume containing heat and smoke, and the lower part significantly less affected by heat and smoke. For each zone, mass and energy balances are implemented with additional models describing other physical processes such as the development of fire through doors, windows and other technological openings, radiation and convective heat transfer and solid fuel pyrolysis.

Computational Hydrodynamic (CFD) models were introduced in the 1990s and have reached significant development and relatively widespread use in various areas of human activity. The rapid growth of computational calculation resources has led to the development of CFD based field models by solving the Reynolds equations, an average format of the Novae-Stokes equations. The use of CFD models is allowed to describe fires in complex geometries involving a wide variety of physical phenomena related to the development of the fire.

In the FDS model, low-speed, thermal-driven equations are solved. FDS solves the numerical form of the Nova-Stockx equations, suitable smoke and heat spreaders in a fire. The equations in the FDS are presented as a set of so-called private differential equations with appropriate simplifications and approximations.

Basic equations of mass, time and energy conservation of Newtonian gas are equations relating to the dynamics of the gas mixture.

This class of equations and their variations are Anderson's, which represent a set of private differential equations consisting of six equations with six unknowns, all of which are functions of three spatial dimensions and time: the density, the three components of velocity, temperature and pressure. Full calculation can be simultaneously considered as a Direct Numerical Simulation in which sputtering members are calculated directly or with the help of Large Eddy Simulation (LES) in which large-scale swirling movements are computed directly and at the same time modeling network scattering processes. The numeric algorithm is shaped so that LES becomes DNS when the network is cleared. Most FDS applications are LES.

The FDS version accepts that the pressure may vary from volume to volume. In the case of a given volume inside the computing sphere, it is isolated from the other volumes, except for a fluid leak, for example from ventilation ducts, it is referred to as "zone pressure" and assigned a given origin pressure.

IV. DESCRIPTION OF THE PROGRAM PYROSIM

PyroSim - is the user interface of Fire Dynamics Simulator (FDS) software [3]. The FDS model can predict the spread of smoke, temperature, and other hazards during a fire. The results of the simulation are used to ensure the safety of the buildings in the design, to determine the safety of existing buildings, to reconstruct fires in the investigations, and to assist in the training of firefighters.

FDS - a powerful fire simulation tool developed by the National Institute of Standards and Technology (NIST). The program allows you to enter interactive input data and validate the source file format for FDS.

PyroSim provides convenient tools for creating geometry in 2D and 3D mode, such as diagonal walls, to facilitate drawing, object grouping, flexible display setup, and copy and

change objects. You can import DXF files with 3D faces or 2D lines, from which you can then create 3D objects in PyroSim.

PyroSim includes tools for creating and validating multiple networks. Networks allow parallel processing of information to speed up the solution.

V. OBJECT AND CONDITIONS OF RESEARCH

The numerical simulation was made in a training corpus in the building of the Professional high school of Telecommunications (Fig.1). The building consists of a basement and four floors. At both ends of the building there are internal staircase cells from the basement to the 4th floor.

Fig. 1. Photo picture of the building

The purpose of the present work is to attempt numerical modeling and to show the spread of smoke and indoor pollution by displaying areas with critical emission, smoke and fire parameters. The source of smoke and debris generation is burning objects in learning rooms. The first step towards the solution of the task is to construct a geometric model of the room (Fig.2).

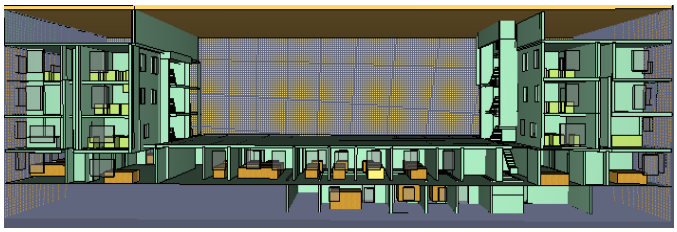


Fig. 2. Geometric model of the building

Room (school room) is selected in the basement of the building with dimensions: 9m long and 7m wide or total area 63m^2. According to Regulation 1971 STPNBP, the heat load for study rooms is 72 kW.h/m^2. We take heat load of 80 kWh/m^2, being increased by 10% of the normative.

VI. RESULTS AND ANALYSIS

According to the heat load given above, the fire has a power of 12600 kW. The fire occurs quickly and in the first minutes it covers the entire room. Subsequently, it reduces its intensity, which can be explained by the rapid combustion of combustible materials and the reduced amount of oxidant.

The smoke in the first minutes through the corresponding slits between the room in which the fire and the surrounding area arose, emerges out of the fireplace. It almost fills the basement floor, covers a part of the first floor and reaches the staircase, leaves the first floor through the open outside door. This indicates that a possible evacuation of the residents on the ground floor should take place in the first 5-10 minutes. When the outer door is closed, the smoke will fill the entire ground floor and will "rise" on the staircase, which will complicate the situation during evacuation and the actions of the fire service.

The fire rate in the room (kW), depending on the time in (min), is given in Fig. 3. In the first five minutes (3 minutes), the fire reaches its maximum power, corresponding to the full range of combustion materials in the room. Here is a section with characteristic turbulent combustion and average power about half the maximum. After 19-20 min. there is a gradual attenuation. This is due to a lack of combustible material and an oxidant (O_2) in the room.

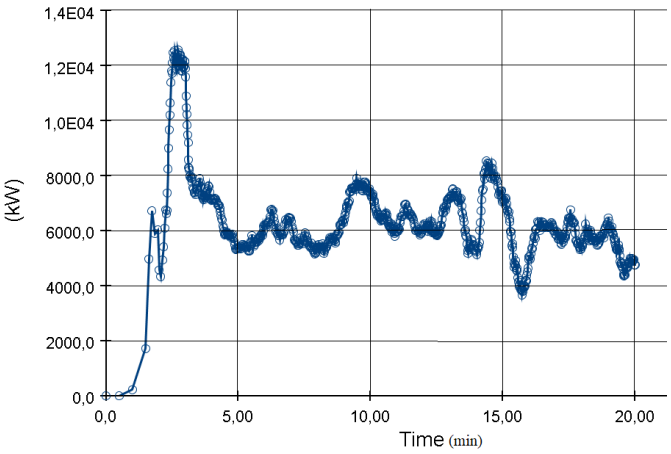


Fig. 3. The fire rate in the room in time domain

VII. CONCLUSION

This article shows an opportunity with the methods of computational fluid mechanics to investigate the development of a fire occurring and developing in a limited space- in this case a study corpus of the Professional high school of Telecommunications.

The FDS program allows for fairly accurate results on the development of fire, smoke, heat exchange with the environment, and provision for measures to limit the occurrence and spread of fire and smoke. This also allows the evacuation routes to be identified with great precision. This is possible if the smoke movement is shown.

The simulation done refers to a specific actual site that was provided with information about the building's layout.

REFERENCES

[1] Rыzhov A.M., I.R. Hasanov, A V. Karpov i dr. ,Primenenie polevogo metoda matematicheskogo modelirovaniia pozharov v pomeshteniiah. Metodicheskie rekomendacii. M. ,VNIIPO, 2003.

[2] Delev.K, Ogneustojchivost i ognezashtita na stroitelni konstrukcii, AVS-Tehnika, 2000.

[3] https://www.thunderheadeng.com/pyrosim/

Flight Safety Sensor System of Unmanned Aerial System

Krume Andreev[1] and Rumen Arnaudov[2]

Abstract – **Over the last decade has seen a rapid development of unmanned aerial systems (UAS). Different solutions have been sought to collect information from difficult access points or in cases where there is a risk to operators and technicians. This report discusses an option for an in-vehicle implementation of the flight safety sensor system for an unmanned aeral system to ensure effectively complete of the mission and respond to the problems and uncertain environmental conditions. The features that these systems must meet are explored. Proposed architectural realization of a flight control platform for unmanned aerial vehicle systems. This system can detect and diagnose failures and violations of safety or performance rules during the flight of a UAS.**

Keywords – **UAS, Unmanned aerial systems, Flight safety, Sensor system.**

I. INTRODUCTION

The collection of information from difficult access points or in cases where there is a risk for operators and techniques within the boundaries of the ground level between 200 and 500 feet requires the use of special technical devices (Figure 1). One option to solve this problem is the use of unmanned flight systems. They can be classified as "fixed wing" UAS, "rotation wing" UAS and lighter-than-air UAS (airships). They are characterizes by high mobility, unlimited time and space access to information and monitoring of their condition [1]. All of this stimulates the development of the ideas for civil use of UAS, making them one of the fastest growing branch of remote sensing technology. Currently, they can list a number of tasks, the research and solution of which presupposes the use of UAS, each having an operational and / or permanent nature.

Unmanned Aerial Systems (UAS) can find the following applications [2]:

• In the event of accidents and disasters, the detection of injured persons, the monitoring of the situation in a distressed region, etc.;

• In agriculture for crop inspection, search for lost animals, view of the farm, damage caused by hail, spills, droughts, diseases, pests, etc.;

• In the ecology for monitoring the status of flora and fauna, monitoring the status of certain areas, etc.;

• In the surveillance systems for ensure large territories;

[1]Krume Andreev Ph.D. student is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: andreev.k@abv.bg.

[2]Ph.D. Rumen Arnaudov Professor is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: ra@tu-sofia.bg.

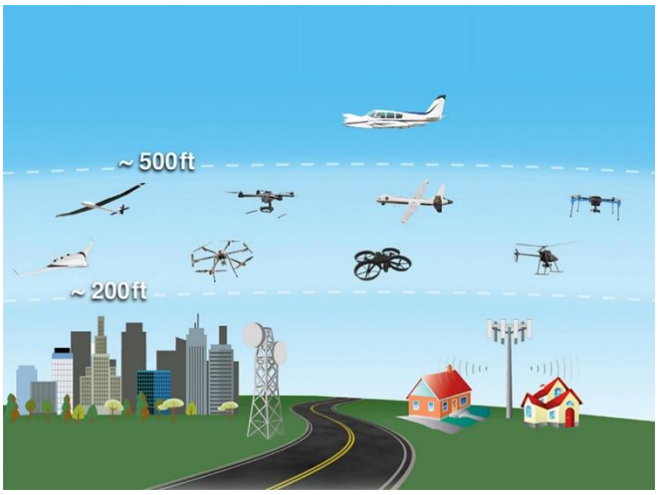


Fig. 1. Use of unmanned aerial vehicles for gathering information from difficult access points or where there is a risk to operators and technology

• To assess damage from areas affected by natural disasters: spills and floods, overflow of dams, forest fires, earthquakes, industrial accidents, destroyed bridges / roads, gasses in certain areas, etc.;

• When capturing newly built infrastructure sites and integrating them into a database, monitoring the state of bridges / viaducts, etc.;

• In the energy sector: monitoring and recording of damages / icing on the power grid, recording of damages along the routes of the oil and gas transmission network, etc.

II. BUILDING A PLATFORM FOR UNMANNED SYSTEM

A. Multi-Purpose Development Platform for Collecting and Processing Data

Unmanned flight systems can only be used if they can effectively complete their mission and adapt to problematic and precarious environmental conditions [3]. Safety must also be maintained with respect to other aircraft as well as people and land. Currently, there is a regulatory law in the US[4] that distinguishes between the use of UAS for specific missions. In the European Union, including Bulgaria, there are baselines that distinguish and define the use of Unmanned Aerial Systems. They are expected to be completed and published soon.

That is why it is necessary to discuss and develop reliable systems and algorithms for flight safety and monitoring of the technical condition of UAS.

Such a system should be able to perform the following tasks:

• To measure and control multiple sensor and software data in dynamic mode;

• On the basis of the collected measurement information from on-board systems, to be possible of perform and report on the control and diagnosis of defects;

• When UAS left the range coverage of the operator range to be possible automatically switch to autopilot or semi-autopilot mode;

• Electronic systems used in unmanned aeral system must meet the essential electromagnetic compatibility requirements.

B. Subsystems for building of the Flight Sensor Safety System of the Unmanned Aeral System

The first subsystem for the construction of UAS is the choice of a construction solution for the system. As a construction solution may be used "fixed-wing" UAS, "rotation wing" UAS and lighter-than-air UAS (airships). "Fixed-wing" UAS (airplanes, sailplanes ...) have relatively high load-bearing capacity and flight duration, but do not have much maneuverability. "rotation wing" (copters) devices are characterized by great maneuverability but do not have high payload and flight duration. Devices lighter than air (airships) are slow and difficult to maneuverability [5].

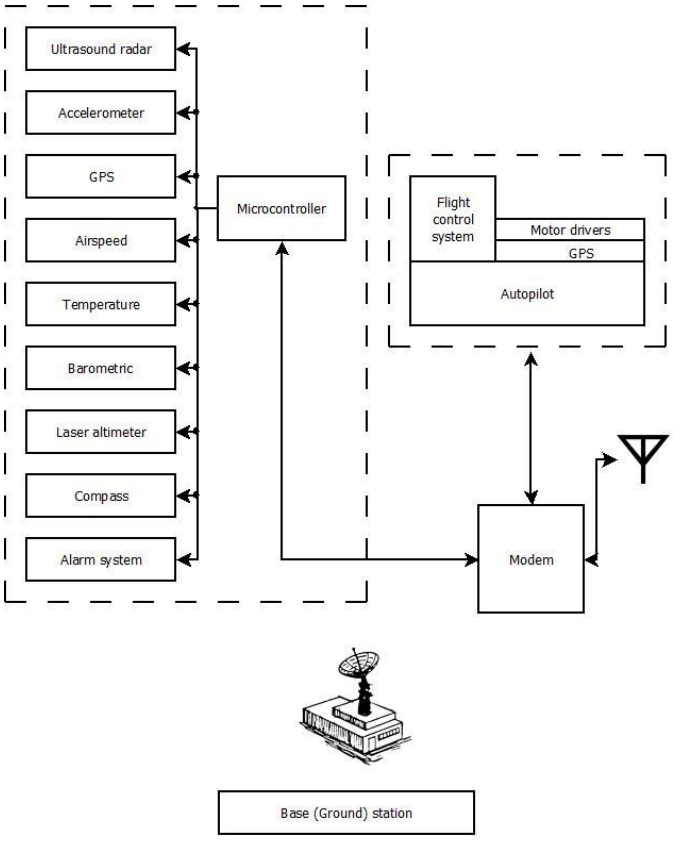


Fig. 2. Architecture Platform for Flight Safety Sensor System

The second subsystem of the UAS is the navigation system that allows a single (or assisted) controlled flight to be performed on a predetermined route or route that is constructed at the time of flight. The presence of this subsystem is the basic subsystem that distinguishes BMA from radio-controlled aircraft models used in hobby practice. The navigation subsystem allows a controlled flight to be monitored, captured, measured, enabling the retrieval of scientific information about the presence, condition or condition of the subject or phenomenon being investigated. The navigation subsystem includes: an autopilot controlling the aircraft's controllable planes to conduct a flight along a route consisting of a number of positions in the space, making a closed flight around a point and at a certain height, returning the aircraft to the point from which has started the flight, etc. ; a positioning device (GPS) reporting the reached position along the route as well as the position of the aircraft (height and direction of flight) at the time of capture or measurement; device (s) for assessing the position of the aircraft relative to the horizon, etc [6].

The third subsystem of UAS is made up of different instruments and tools for capturing, recording and measuring, parameters for assessing spatial dimensions, position or state of various objects and phenomena. Depending on the objectives, this subsystem may be deployed only on-board the aircraft or be composed of two segments: an on-board segment and a ground segment. In the latter case, the on-board segment is complemented by a transmitting element, which allows the real-time telemetry information to be transmitted to the receiving segment from the ground segment. In the presence of a transmitting element, the function of the navigation system can also be enriched by providing a feedback allowing visualization of the flight parameters and the operation of the aircraft [7].

It can be seen that the task is very complex in itself, and in addition it becomes even more complicated when considering the tasks that UAS should perform. This is also the reason for the complexity of compiling standards and norms in the field of trembling.

C. Architecture of the Flight Sensor Safety System of the Unmanned Aeral System

Fig. 2 shows the architecture of a platform for flight safety sensor system of a UAS using a fixed-wing. It contains: measuring sensors, microcontroller for data processing and measurement system management, flight control system, modem, autopilot, manual control channel for UAS, telemetry for data transmission and ground station. With the help of the measuring sensors, the condition of the aircraft, its coordinates and the environmental parameters are monitored. For this purpose the following sensors are used:

• GPS receiver. It helps determine the location of the aircraft.

• Temperature sensor. In this case, the presence of hot air currents is monitored.

• Air velocity sensor. This parameter is important because it can also control the speed of UAS.

• Barometric pressure sensor. This height and speed can be monitored via the Z axis.

• Laser altimeter for height measurement.

• Ultrasonic radar to prevent the unmanned airplane from colliding with other or other flying objects.

• Accelerator accelerometer for acceleration of unmanned aircraft in case of a sudden loss of lifting force or impact of a UAS on another object.

• A magnetic field sensor that measures the Earth's magnetic field as an inertial navigation signal;

• Alarm system. With this system, in the event of a fall of UAS, it signals the uncontrolled fall of the unmanned flight system.

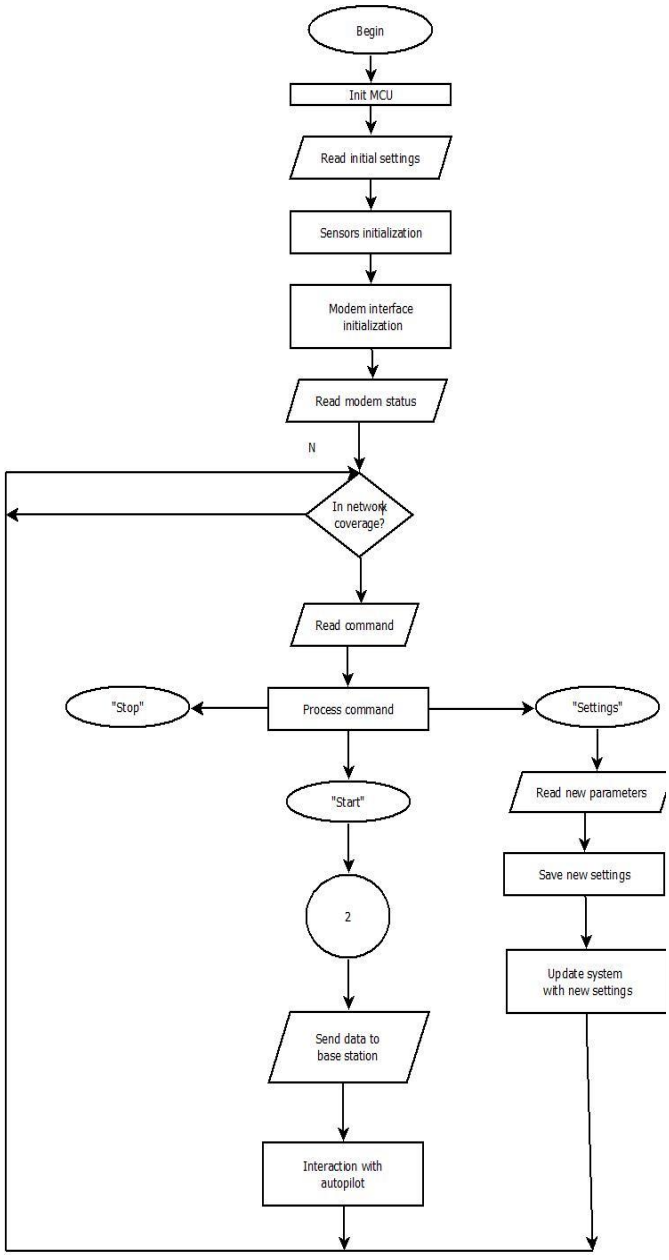


Fig. 3. Algorithm for operation of the system for flight control and technical condition for collecting and processing data from sensors onboard the UAS

Knowing the capabilities of different systems allows for proper selection in preparation for a particular task, and the same goes for each task requiring a proper selection of the UAS as a flight system, scientific and navigation equipment.

The basic principles for building UAS are [8]:

• Creating a modular, light platform with low power consumption;

• Real-time monitoring of sensors and software signals;

• Analysis of received information and in-flight preprocessing.

Another important feature in the choice of UAS and the flight safety sensor system is the ability of the established platform to comply with the flight rules in the respective country. At this stage in Bulgaria, these rules are being developed. Flight rules are defined by national and international institutions, for example, the US Federal Aviation Administration [9]. Some of the parameters in this policy are the minimum height that aircraft after take-off, maximum flight speed, landing time, etc. have to reach;

D. Algorithm of the Flight Safety System for Collecting and Processing Data from Sensors on Board of the Unmanned Aeral System

After starting the device follows an initialization of the MCU [10]. Initialization is performed on the controller, i.e. self-initialization. After initialization follows read of the initial settings and initialization on all sensors on board that directly depend on the system and model interface initialization. Then system proceeds with the modem read status. If the system is in the network coverage the next step is expected acceptance of command. If it's not the system is waiting and checking for the connection with the ground(base) station. Commands are three, as follows, for start, stop and system settings. For a given command start the system begins to read the data sent by sensors on board of the UAS and began to perform the functions of the system (the system functions described above in the text-point section Introduction). After receiving the data from sensors follows the process on the received data and the system makes the right decision. This will be showed on Fig. 4 and described in the description for the algorithm. The steps are as follows send data to the base (ground) station and interaction with the other systems on the UAS. In case of need for a interaction of the system for flight control and technical condition of UAS with the other sub-system on board of the UAS i.e. in this case the system for flight control or autopilot. In this case the decision taken in the form of data is sent to the flight control system i.e. to the UAV's autopilot. [11] For a given command stop the system will stop working until the operation will turn it again on. For a given command settings the system will read the new parameters set up by the operator. Save the new set up parameters or save the new settings and will automatically update the system with the new set up parameters.

Fig. 3 shows the algorithm of the flight safety sensor system of an unmanned aeral system for collecting and processing data from sensors on board.

For a given command begin system starts working. The first step is reading of the input data. After reading of the data follows the data processing or filtration of the data. Next step is to check if the filtered data is in statistical range, depends on the stored system parameters data. If the data is in the range the system continue working. If the data is not in the range of the stored system parameters data the entire test is repeat until the system reach limit. If the limit is reached the system is sending warning signal data to the base (ground) station for non-fuctional sensor.

Fig. 4 shows the algorithm for data processing and decision making algorithm of the system for flight control and technical condition for collecting and processing data from sensors on board.

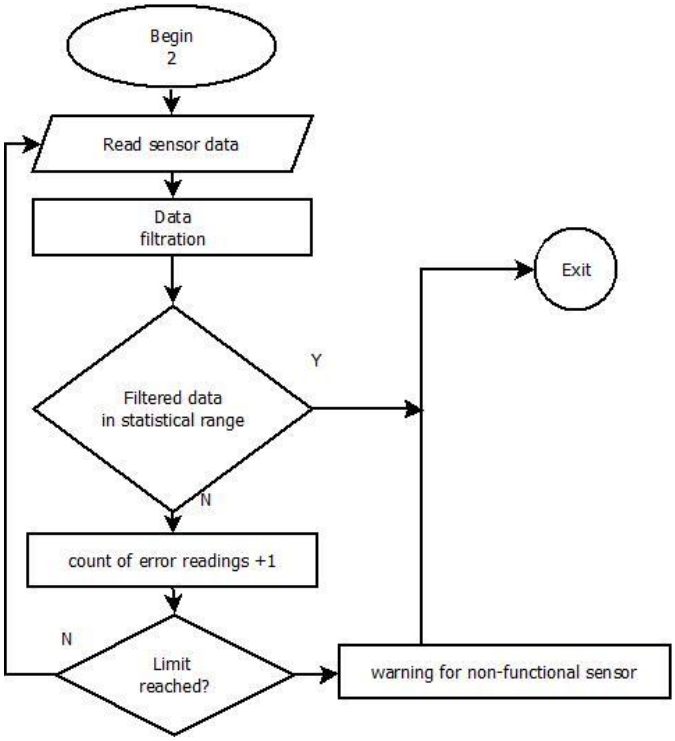


Fig. 4. Algorithm for data processing and decision making algorithm of the system for flight control and technical condition for collecting and processing data from sensors on board the UAS

III. Conclusion

This report proposes an architecture and algorithm for flight safety sensor system of an unmanned aeral system for collecting and processing data from sensors on board. They are the basis for the design and development of intelligent UAS and to develop a regulatory framework for standardization and rules for their flight. Such systems may find wide application in other areas, especially where the combine requirements for high operational reliability and service while making quick decisions.

This report introduce a system that meets this requirement with the design of a real-time onboard system for flight safety capability to continuously monitor sensors and hardware components. This system can detect and diagnose failures and violations of safety or performance rules during the flight of a UAS.

Acknowledgement

The proposed architecture and algorithm have been developed with the support of NIS at TU-Sofia, through participation in research projects to help graduate "System for flight control and technical condition of unmanned aerial systems" (session 2018) № 172ПД0005-07.

References

[1] D. Basin, F. Klaedtke, & E. Z˘alinescu, “Algorithms for monitoring real-time properties.” pp. 220–235, 2017.
[2] H. Latchman, T. Wong "Software Documentation for Airborne Traffic Surveillance Systems", pp. 10-17, 2007.
[3] J. Schumann, O.J. Mengshoel, and T. Mbaya, "Integrated Software and Sensor Health Management for Small Spacecraft", 2011 IEEE Fourth International Conference on Space Mission Challenges for Information Technology, pp. 77-84, 2011.
[4] Federal Aviation Administration. (2013). Federal Aviation. Regulation §91.
[5] D. Basin, F. Klaedtke, and E. Zalinescu, "Algorithms for Monitoring Real-time Properties", pp. 260-275, 2011.
[6] D. W. Murphy "The Air Mobile Ground Security and Surveillance System", pp. 20-23, 2004.
[7] Richard A. Di Lorenzo, Michael A. Bayer, “A Prognostics and Health Management System for an Unmanned Combat Aircraft System – A Defense Acquisition University Case Study”, pp.1-12, 2009.
[8] Johann Schumann, Kristin Y. Rozier, Thomas Reinbacher, Ole J. Mengshoel, Timmy Mbaya, and Corey Ippolito, “Towards Real-time, On-board, Hardware-supported Sensor and Software Health Management for Unmanned Aerial Systems”, pp.1-27, 2014.
[9] Federal Aviation Administration. (2013). Federal Aviation. Regulation §91.
[10] Johann Schumann, Kristin Y. Rozier, Thomas Reinbacher, Ole J. Mengshoel, Timmy Mbaya, and Corey Ippolito “Towards Real-time, On-board, Hardware-supported Sensor and Software Health Management for Unmanned Aerial Systems”, pp.1-21, 2013.
[11] Reg Austin, “Unmanned Aircraft Systems –UAVS Design, Development and Deployment”, pp. 16-56, 2010

System Information Analysis of the Human Factor Within Control Systems

Zoya Hubenova [1], Peter Getsov[1], Georgi Sotirov[1]

***Abstract* –In the paper hereby, topics related to contemporary information control systems analysis are discussed taking into account main technical aspects such as economical, technical and psychological. In addition, the current study topicality is demonstrated relevant to human factor within such a system. The operator reliability indicators are also estimated such as error free, precision, and timeliness.**

Keywords – **information and control system, human factor, reliability, energy systems**

I. INTRODUCTION

At the present development stage of the scientific – technological progress, the task of research and development of a theory of human factor (HF) reliability and efficiency within the control systems and systems "man – machine" is put forward. Working out a solution of such a problem is possible provided that a sophisticated system – informational analysis is carried out both in technical point of view and physiological, psychological, and engineering – genetic problems of the human factor.

The formalization, modelling and analysis of the man in the sophisticated information and control systems in that regard turns out to be *important interdisciplinary scientific problem*, requiring the use of theory and methods in many branches of science: the Cybernetics, Physiology, Ergonomics, Mathematics (fuzzy sets, mathematical linguistics, semi-markov processes, etc.), system analysis, biomechanics, computer sciences etc. When analysing the function of such sophisticated systems, it becomes increasingly evident, that reliable results cannot be obtained without taking into account the human factor, because a person is an active part in them, who defines to a large extent the achievement of its objectives in its operation and development. The man, as an element of the control, participates in every stage of its formation – perception, recognition, prediction, adoption of a decision and implementation [1].

The present work focuses on the existing possibilities for transferring of theory reliability of technical objects in the development of technologies for reporting the reliability of operators in complex hierarchical systems.

The purpose from position of the system approach is to analyze the quality of modern energy systems functioning as well as the requirement for high reliability of the human being as a unit in them. The factors influencing the reliability of the human operator (HO) and their quantification are analyzed. The problem is particularly relevant in connection with the fact that a number of operations are characterized by extreme conditions that can not only reduce reliability but also cause harmful and dangerous human impacts.

[1]Space Research and Technology Institute–BAS, Acad. G. Bonchev Str., bl. 1, 1113 Sofia, Bulgaria, E-mail: zhubenova@space.bas.bg director@space.bas.bg; gsotirov@space.bas.bg

II. SYSTEMATIC ASPECTS OF INFORMATION SYSTEMS FUNCTIONING

In the technical cybernetics the "big artificial system" is called the aggregate of a big number of hierarchically dependent complex sub-systems, composed of collectives of people and machines with a certain level of organization and independency, united on the base of an active hierarchy of goals and means of the organization, commonly with energetic, substantial and informational connections for the insurance of the purposeful functioning of the entire system as "a whole".

The means of organization include knots for management, where the process of decision making takes place, and executive organs, realizing the information of the made decision in action, directed to achieving the set goals from the management. Commonly, in the big system, the knots of management and the executive organs represent complex human-mechanic complexes. For the purposes of quantity analysis of the big system, it is necessary for all its defined components to be formalized. Presently, the least studied of them, particularly quantity-wise, are the organization, the hierarchy of goals, the information and the independency, as well as the relations between man and machine, composing the knot of management. In relation to the goals of functionality, the big artificial system must ensure the completion of assigned substantial-energetic processes. But its functionality, development and existence as a whole are determined by the processes of transfer, refinement and transformation of information. Therefore, in the big artificial systems it is productive to separate three aspects: substantial-energetic, economical and informational. With the informational approach and the observation of informational-management systems, each studied system is represented by three levels: field, management, informational – meaning, in an abstract type, it is represented through a hierarchical structure on the lower level, where the technological process sections are located, and on the higher levels, the management knots, connected with the objects of managements and with each other, through channels of the network communication, shift places.

The information, circulating in the system, may manifest itself in three forms: informative – directed with priority from the objects of management to the relevant management knots; managerial – in reverse direction; transformational – defining

the regularities of behavior of the management knot and the algorithm of functioning of its individual elements [2].

The management knots transform the informative data into managerial with the help of the transformational information, contained in the algorithms and the structure of the management knot. The different sections of the technological process are considered as generators of primal informative information. As it moves up the hierarchy, the information gradually gets synthesized, transformed in the different management knots and arrives in the main management knot on the top of the hierarchy. Using the acquired informative information, this knot generates managerial information, which gets more detailed as it moves down to the lower knots. The smaller the volume of the necessary higher knot information to form managerial information in the "i" knot, the more autonomous this knot is.

To achieve the goals of management, it is of great importance that only the necessary (valuable) information reaches the respective knots. That is why in the informational-managerial systems, the semantic and value statistics of the information are displayed with priority. The management of complex technological processes (such as energy) always occurs in an environment of information deficiency. This is mainly caused by the fact that in the management knots in these systems, representing a unification of people and machines, these systems are only partially observable, partially manageable and partially knowable.

The energy system is an example of such large artificial systems. They, with their variety in equipment, huge amount of sub-systems and complex connections between them, are a natural subject for the application of a multileveled hierarchic approach in the building of their organization and management, including a wide spectrum of organizational activities, planning, management, building, infiltration, exploitation, support and development of information systems and technologies. With the contemporary organization of the production of electricity and its distribution and allocation through systems for access and transportation, organized by various wholesale companies, some exceptionally elaborate energetic complexes emerge, uniting systems in narrow correlation. Due to the high requirements for reliability and safety of functioning, their management is unthinkable without quick and timely decisions, since possible failures in the energetic system may have serious economical and social ramifications. This requires development and upgrade of new approaches towards organization and control of such huge energy systems.

Hierarchy of management in the contemporary electro-energetic complex can be represented with a series of prominent levels: European, regional, national, local; through a wholesale object – electrical station, sub-station, network region. Each of these levels is composed of its own internal hierarchy [3].

The low level of management, particularly if an electric station is being considered as a subject of management, has the most strongly emphasized internal hierarchy: dispatcher management (commands for commutation switching, correction and control of the work of group and autonomous management systems); common (secondary) automatic management of stations in normal and alert modes (distribution of active and reactive power, regulation of frequency, management of commutation operations, function of anti-alert automatics, adaptation of settings of regulators); group complex regulation, "zero" autonomic regulation, ensuring the assigned carrying out of the technological process of preparation of the energetic system and etc. [4].

Typical for the energetic system is the existence of a complex information structure; hierarchic dependency between the knots of management and in them; colossal in volume streams of information over the communication channels, between the knots of management; correlation between information and processes of its transformation relative to the goals, (which defines the semantic and pragmatic properties of the information); big variety of different information properties, transformers of information (different in their purpose regulators, controllers and etc.; people from different fields of expertise in the knots of management).

A classical example of a multileveled hierarchic system is the problem with dispatching of active power in the energetic system (Fig. 1)

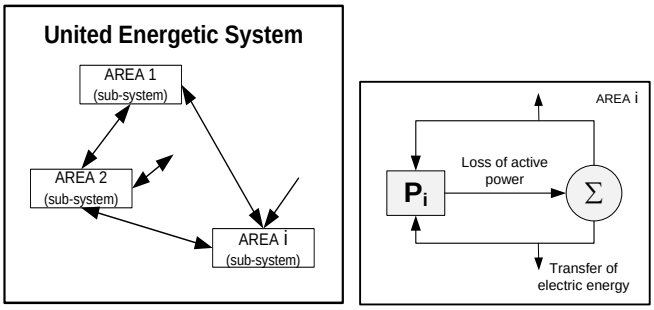


Fig. 1. Hierarchical power system

The system is divided in *n* correlated sub-systems. Usually, the boundaries of the respective areas are chosen in a way that each one of them represents a separate association or company [5].

Each area has a series of generator stations and a big variety of consumers, but as long as the transfer of electrical energy between the areas is a major interest, it can be assumed, that each area is characterized only by the following parameters: C_i – load up (full) in area *i*; X_i – active power, generated by the elements in area *i*; Y_i – loss of active power in area *i*; U_i – transfer of powers through the energetic lines, connecting area i with the other areas.

With a set load, the loss in the *i* area depend on the power, generated in the area itself, as well as on the transfer of powers, meaning Y_i is a function of X_i and U_i.

$$Y_i = P_i(X_i, U_i) \tag{1}$$

The equation, defining the balance of powers can be written in the following way:

$$P_i(X_i, U_i) + C_i - U_i + X_i = 0. \tag{2}$$

There can be an *n* amount of equations of this kind. Also, the transfer of powers between the individual areas must be balanced:

$$U_1 + U_2 + ... + U_n = 0 \qquad (3)$$

Equation (2) describes the sub-processes, and equation (3) – their relation.

Based on this, it can be assumed that the problem of dispatching of active power in a unified energetic system lies in the need to be able to define, on all lines, the power X_i and the volume of transfers U_i where the value of the produced electric energy would be minimal. For that purpose, the function $F_i(X_i)$ is introduced, called function of value of generation in each of the areas with power X_i. Then the overall value of the generated power will be:

$$F(X_1,.....,X_n) = F_1(X_1) + ... + F_n(X_n) \qquad (4)$$

The problem with optimal dispatching is reduced to a minimization of example (4), provided that the variables $X = (X_1,.....X_n)$ and $U = (U_1,....U_n)$ satisfy the equations of balance (2) and (3). It can be noted that after equation (3) the number of the independent values U_i is $n-1$. The task and goal of each hierarchic organization is connected with the solving of the problem with minimization. In relation to this raises the question how is this task supposed to be distributed between the computers of the different areas and the central computer. The multileveled management is preferred in comparison to the centralized approach, due to a number of reasons of technical, economical and exploitational nature.

III. USE OF SYSTEM-INFORMATION METHODS FOR HUMAN FACTOR EVALUATION

It is known that tasks definition of the operator in such a case of automated control systems is reduced practically to the following: timely detection the automated system inability in order to deal with the violations occurring during the process, determination of cause of the fault and making up for the consequences [6].

Human activity is structured and organized as a set of unit operations and is based on concepts: valuables, sense, objectives, actions, and operations. Regulatory activities are associated with the rules, standards and regulations. Elementary activities presented as elemental activity unit have spatial – temporal characteristics, [6]. The information activity model is described through a *Standard Operating Procedure* (*SOP*). It is structured on the basis of (in relation to the information search time) decision – making time, time for action (inaction), and time to obtain this result in accordance with the plan of action.

Operator's reliability can be defined as the probability of accurately and adequately performing the tasks in accordance with the instructions. Accuracy means implementation in accordance with the standard limitations in time and space. Proper operation a1 is performed within a standard operating procedure in moments of time t_1, t_2, …, t_i with reserve excess ±Δt. Proper or normal operation is performed in case of timely detection either a signal or a set of signals by the operator which is/are necessary for action. Next, what follows is correspondence identification of signals according to standard configurations, making a decision, the action itself, obtaining and evaluation of the outcome. The concept of time reserve excess is associated with relevance or importance of the procedure, [7].

$$C(\bar{t}_r) = e^{-\lambda \bar{t}_r} \ , \qquad (5)$$

where $C(\bar{t}_r)$ means importance in terms of content and time constraints whereas $e^{-\lambda \bar{t}_r}$ denotes a continuous stream of events.

Proper operation a1 within the standard procedure ps is performed at a time t1 with reserve excess ±Δt. Proper performance of actions is possible in case of early detection of a signal or a set of signals by the operator need for action. What follows is signal identification for conformity to the Standard Operating Procedures, decision-making, action itself and outcome identification, which is determined in the same order. The procedure time Tp = Ta + Tr increases with the reaction time Tr (Ta – the time of action).

Accuracy could be determined as a degree of approximation of the actual process parameter value to its nominal value, i.e. as result compliance. The accuracy of operator actions depends on the systematic and random causes. They are determined with confidence $\beta = 0{,}95$.

$$\delta_s = M + 2\sigma_s \, , \qquad (6)$$

where $M = m_1 + m_2 + ... + m_k$ is the systematic error of the system whereas

$$\delta_s = \sqrt{\sigma_1^2 + \sigma_2^2 + ... + \sigma_k^2} \ , \qquad (7)$$

is the random error standard deviation of the system elements, [8].

As a category, contrary to the accuracy the term imprecision is adopted. Absolute accuracy of the actions is practically unattainable, therefore a total acceptable imprecision and accuracy reserve are determined:

$$\delta_G = D_j - \delta_{\min} \, , \qquad (8)$$

where D_j is tolerance limits of the parameter j through D_j whereas δ_{min} is minimum total error by parameter j consisting of imprecision of the device during parameter measurement and parameter estimation imprecision made by the operator and his actions. Accuracy reserve excess is determined by the largest admissible imprecision.

A good example of this is the unmanned systems where safety is of paramount importance and is directly connected to selection, assessment and reliability of operators [9].

Due to the widespread implementation of computerized control systems, the operator's field perception narrows to the monitor screen(s) (change numeric values, colors and shapes ripple image) in practice as well as audio messages and

violations. Operator's control forces are carried out through a choice of set elements displayed on the screen or keyboard keys and functional devices buttons (joystick, trackball, tablets, etc.), thus depriving the operator's motorial action from a significant motive component. Taking into account the man-operator's specific activity, related to a complex dynamics of observing monotonous parameters and logic elements, it could be said that intellectual tasks predominate before perceptual and motor ones.

Researches on the human brain activity are focusing on the fact that a reasonable person copes with unfamiliar situations and makes rational decisions because he can extract new knowledge from existing experience and can consider the consequences of those decisions. People analyze not only accurate, predefined data, but also incomplete information which often has not a numeric expression. This means that a person meets challenges of unstructured type with non-algorithmic solution on daily basis; the quantities he operates with cannot be set in numerical form; their solution requires processing of information which is ambiguous and changes dynamically; purposes of the mathematical problem cannot be expressed by exact objective function. The human intellect is complex biological phenomenon and it is not limited only to problem solving, structured or unstructured. But in the present level of knowledge and instruments not all types of human reasoning are well studied and therefore it is impossible to be modelled (for instance like creativity, intuition, imagination).

IV. CONCLUSION

As we noted so far, the operator's activity differs from the other type of activity with this that he resolves issues on control, management, transmission or transformation of information, interacting with the external environment or technical devises not directly, but by the assistance of various means to display the information and by the relevant control authorities. General characteristics of the activity of all operators is the collection, evaluation and processing of information for technical equipment, technological and other processes, dynamic objects; taking the relevant operator's decisions based on the evaluation of information; actions on their implementation; monitoring of the effectiveness.

In this paper is made an analysis in the system-information plan of the HF as an element of complex information-control systems. The proposed structure of the hierarchical system is the basis for further fundamental research and development of the activity of dispatchers and operators with intensive professions. The expected results are related to the creation of activity models to allow monitoring of the operator's work.

Reliability and efficiency of the man – machine systems have to do with the particular important features. Not only is the Man – Operator reliability within the human-machine systems determined by psycho – physiological endurance regarding harsh loads but the human intellectual properties to make up for a variety of distorted informational processes during interaction with technology is also important. The development of these systems require more research to be carried out on the processes of information perception, processing, and storage by human as well as decision making mechanisms in different cases, psycho – physiological factor impact on reliability and effectiveness of such systems.

Acknowledgements: *This paper is financed by Operaional Programme "Development of competitiveness of Bulgarian economy" 2007-2013. Cofinaced by the European fund for Regional Developmen - Project BG161PO003–1.2.04–0053 "Inforrmation complex for aerospace monitoring of environment".*

REFERENCES

[1] A. Andonov, Z. Hubenova, Functional Stability of Information Control Complexes in Case of Their Critical Applications, University of Transport, Sofia, 2011.

[2] Z. Hubenova, "Method of information assurance operators' activities during control processes within complex ergatic systems", 20th International Scientific Conference, Academic journal "Mechanics, Transport, Communications", Issue 3/2011, p.VIII-76-83.

[3] Green Paper: Towards a European strategy for the security of energy supply, http://www.seea.government.bg/documents/Green-Book.pdf

[4] Energy Strategy of the Republic of Bulgaria til 2020, june 2011, http://www.strategy.bg//Images/logo_ok.jpg

[5] Bradford Travis, The Energy System, Technology, Economics, Markets, and Policy. Hardcover, 2017.

[6] Handbook of Human Factors, edited by G. Salvendy, in 6 volumes, Purdue University. 1998.

[7] Hollnagel. Human reliability analysis: context and control / E. Hollnagel // London; San Diego, CA: Academic Press, 1993

[8] V. Bodrov, V. Orlov , Psychology and dependability: Human in control systems of technical equipment. – M.: Academy of Science", 1998.

[9] P. Gezov, Z. Hubenova , "Training and Evaluation of UAV Operators by Means of C-Star simulator", Journal of the Bulgarian Academy of Sciences , M .Drinov, Vol. 5, 2017, p.27-33.

Objects Identification in a Loop with PID controller

Jordan Badev[1], Angel Lichev[2] and Veselin Stanchev[3]

Abstract – **This paper presents an experimental method for collecting data on processes identification. The data is acquired by an experiment, conducted in a closed loop with a PID controller. For the purposes of the experiment, a critical operating mode is achieved. The analytical expressions of the amplitude frequency characteristic and the phase-frequency characteristic are used for the definition of two parameters.**

Keywords – **Object control by a PID controller, Critical operation mode in a closed loop, High order models**

I. INTRODUCTION

During the identification of the controllable objects (CO), there is a search for an analytical model, based on experimental readings. The analytical models are suitable for repeatedly reproduction of experiments with the real physical objects with reasonable accuracy - independently of or jointly with measuring and/or controllable devices (CD). Analytical models are most frequently used for laboratory setting of the elements of the controllable device, by the CD system simulation with CO or for producing scaled (reduced or enlarged) physical models [1].

Collecting informative data is a very important stage of the identification, by which an adequate analytical model of the physical processes can be evaluated [2]. The proposed identification method has the following characteristics (advantages):

Informative identification data is collect by experiment, for which a conventional PID controller [3, 4] is sufficient in a circuit for constant operating with CO

The method is applicable to high order models [5] with transportation lag (Straits models).

II. THEORETICAL SETUP OF THE IDENTIFICATION METHOD

The proposed identification method calculates the parameters in the model of CO, using the experimental results from the critical process in the loop and the analytical expressions for: amplitude frequency characteristics (AFC) and phase-frequency characteristics (PFC).

The analitical expressions of the frequency characteristics (FC) for the proposed model are found according to the rules of the control theory (CT) and they contain manifestly and most frequently non-linear the unknown parameters.

The experimental results for the amplitude and the phase are measured from a critical operation mode of the object in the closed loop. Eventually, the task becomes solving a system of non-linear equations, according to the unknown parameters. Fig. 1 shows the indications of the signals in the loop CD and CO. These indications are used in the paper.

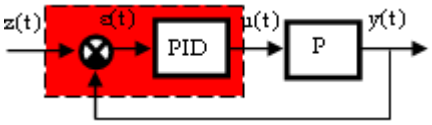


Fig. 1. Indications of the signals in a loop with a PID controller

It is indicated in the figure:
P – Controllable object;
PID – Control device (PID controller);
z(t).- Reference value (set point);
u(t).- Control signal sent to the system(control signal) ;
y(t) – The measured output of the system(process variable);
y(t)-z(t)= ε(t) – error value (system error).

The identification method of CO uses experimental data collected from the critical operating mode of the loop. The critical operating mode is well presented in CT but the well-known Nyquist stability criterion for stability of linear systems presents best the critical mode. Fig. 2 shows the amplitude-phase-frequency characteristics of the open structure:

$W_{OC}(j\omega) = W_{PID}(s)*W_P(s)$,

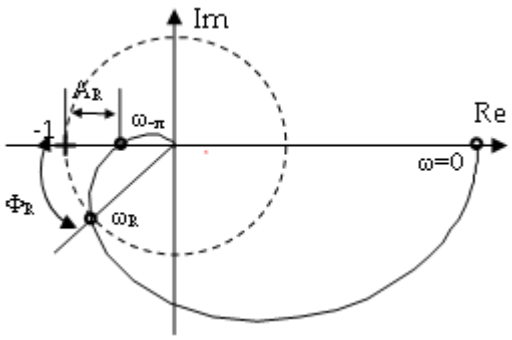


Fig. 2. APFC characteristics of the open structure

In Fig. 2 by A_R and Φ_R are marked the corresponding stability margins.

According to the Fig. 2.the critical operating mode represents the case when amplitude-phase-frequency

[1] Jordan Badev is with the Department of Automation, information and control systems, Technical Faculty, University of Food Technologies, 26 Maritza Blvd., 4002 Plovdiv, Bulgaria, E-mail: jobadev@abv.bg.

[2] 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

[3] Veselin Stanchev is with the Department of Automation, information and control systems, Technical Faculty, University of Food Technologies, 26 Maritza Blvd., 4002 Plovdiv, Bulgaria, E-mail: vsstanchev@abv.bg

characteristics (APFC) crosses the real axes, in the point*(-1,j0)* (theNyquist point). It is known from the CT, that the crossing is impossible in case of first and second order transfer functions (TF) $W_{OC}(j\omega)$ without transportation lag. Therefore it is considered the case of third and higher order TF.

On the other hand, it could be written for the common TF (controller + object) P mode,

$$W_{OC}(j\omega) = W_{PID}(s)*W_P(s)=K_{PID}*K_P*\frac{B(s)}{A(s)} \quad (1)$$

where: K_{PID}– static coefficient of the controller in P mode.
K_P– static coefficient of the controllable object.
$\frac{B(s)}{A(s)}$ – fractional-rational part of the common TF.
It follows from the last expression:

1. We could change the APFC by increasing (decreasing) K_{PID}, so it could cross the real axes in the Nyquist point, or we could set the loop by the contoller to operate in the critical mode.
2. If the controller is only proportional (only *P*), the rest of the TF is exactly the TF of the object, or $K_P*\frac{B(s)}{A(s)}=W_P(s)$, which gives us a reason to think that the parameters of the critical operating mode are only parameters of the object, on which the proposed method is based.

The basic parameter which characterizes the critical operating mode is the critical frequency - $\omega_{кр}$, that is the frequency at which:
The module$|\ W_{oc}(j\ \omega_{кр})| = A(\omega_{кр}) = 1$
and the phase $arg(\ W_{oc}(j\ \omega_{кр})) = \Phi(\omega_{кр}) = -\pi$.
The last considerations lead to the conclusion that:
The loop could be set to a critical operating mode by increasing K_{PID}, so APFC could cross the real axes in the point *(-1, j0)*. In this operating mode, the output process *y(t)* oscillates with a constant amplitude and is in antiphase with the input control signal *u(t)*.

The last fact is used for simple experimental realisation of the critical operating mode and experimental data collection for identification.

In case of operation with classical PID controller, the critical operating mode shall be set up as follows:

1. The PID controller is set for operation in P mode, and the integral and the derivative terms are excluded, by setting the appropriate values of the parameters for setting ($T_и\to\infty$, $T_д\to 0$).
2. An average value is set with regard to the maximum control range.
3. A small value is defined of the coefficient of proportionality K_{PID}, for which the process of the output y(t) decreases. The coefficient K_{PID} is increased smoothly, and y(t) is monitored (the characteristic is built). The process is supposed to change to a slower attenuation and to an increase of the oscillation. The increasing continues until fluctuations with constant amplitude Δy_{kp}, constant cycle, $T_{кр}$ and in antiphase with the input *u(t)* are established.
4. The oscillations of the *y(t)*are being written over several cycles. The amplitude Δy_{kp} and the established oscillations $T_{кр}$. are defined from the diagram.

The analytical expressions of TF, AFC and PFC of the model by which the controllable object of the proposed method could be defined are shown in Table. 1.

TABLE I
MODELS OF APFC AND PFC

№	TF of the model	System of PFC and AFC
1.	$\frac{K_p}{Ts+1}*e^{-\tau s}$	$\pi=\omega_{кр}\tau+arctg(\omega_{кр}T)$ $A_{cr}=\frac{K_p}{\sqrt{1+(\omega_{кр}T)^2}}=\frac{\Delta y_{кр}}{\Delta u}$
2.	$\frac{K_p}{(Ts+1)^2}*e^{-\tau s}$	$\pi=\omega_{кр}\tau+2*arctg(\omega_{кр}T)$ $A_{cr}=\frac{K_p}{1+(\omega_{кр}T)^2}=\frac{\Delta y_{кр}}{\Delta u}$
3.	$\frac{K_p}{(Ts+1)^n}$	$\pi=n*arctg(\omega_{кр}T)$ $A_{cr}=\frac{Kp}{(\sqrt{1+(\omega_{кр}T)^2})^n}=\frac{\Delta y_{cr}}{\Delta u}$

It is typical for all the presented formulas, that they contain three unknown parameters: (K_P, T, τ и K_P, T, n).

The following parameters are needed for the identification, which are defined by the experimental diagram of the critical process, shown in the Fig. 3:

- Critical frequency

$$\omega_{кр}=\frac{2\pi}{T_{кр}} \quad (2)$$

- Fluctuation range of the controllable (output) variable

$$\Delta y_{кр}=y_{\max}-y_{\min} \quad (3)$$

- Average value of the controllable variable

$$y_{ср}=\frac{y_{\max}+y_{\min}}{2} \quad (4)$$

- Average value of the control impact

$$u_{ср}=\frac{\Delta u}{2} \quad (5)$$

- Variation range of the control impact

$$\Delta u=u_{\max}-u_{\min} \quad (6)$$

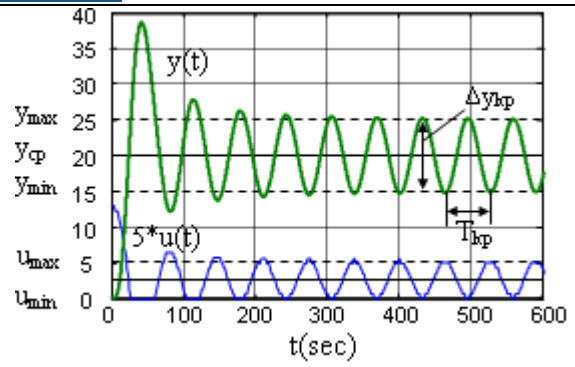


Fig. 3. PID controller operation with object with TF $\frac{50}{(10s+1)^4}$

III. AN ALGORITHM FOR THE CALCULATION METHOD OF THE PARAMETERS OF MODELS

1. Preparation and realization of the experiment and construction of the process in Fig.3
2. Calculation of the static coefficient of the object by the formula:

$$K_p = \frac{y_{cp}}{u_{cp}} \tag{7}$$

3. Calculation of the critical frequency by the formula:

$$\omega_{кр} = \frac{2\pi}{T_{кр}} \tag{8}$$

4. Calculation of the critical module by the formula

$$A_{кр} = \frac{y_{кр}}{\Delta u} \tag{9}$$

5. For the models 1 and 2 in Table 1., it is easy T and τ to be defined – first T from AFC and after that τ from PFC.

6. For the models 3, the system according to T and n is:

$$n = \frac{\pi}{arctg(\omega_{кр}T)} = f_1(T) \tag{10}$$

$$n = \frac{\lg(\frac{K_{об}}{A_{кр}})}{\lg(\sqrt{1+(\omega_{кр}T)^2})} = f_2(T) \tag{11}$$

It is obvious, that the system is non-linear. Methods for solving different non-linear equations systems are known from mathematics.

We propose a decision, which can be easily achieved in computing environment of MATLAB, but it must be taken into account that:

- the unknown n is number of first order lags in the model and can accept equivalent and positive units only, which are greater than or equal to 3, i. e. ($n \geq 3$)
- the unknown T is meant as a time constant and thus it can accept positive values only, i. e. $T \geq 0$.

IV. VALIDATION OF THE METHOD AND RESULTS

The presented method is validated in the computing environment of MATLAB. Numbers of loops with objects are simulated. In all cases, the aforementioned algorithm is used. For high order models and PID controller a non-linear equations system (9) and (10) is solved. The non-linear equations system is solved building the graphics of the two equations *(T, n)*, and the coordinates of the intersection of the two graphics are the system solution. So, the definitive value of the parameter *n* is supposed to be the nearest equivalent unit.

To illustrate the method, two exampels are shown – with simulated loops with PID controller and fourth and sixth order objects (by the circuit in Fig. 1.), respectively by transfer functions of CO:

$$W_{P1(s)} = \frac{50}{(10s+1)^4} \tag{12}$$

the desired set point is:

$$z_1(t) = 25 * 1(t) \tag{13}$$

and

$$W_{P2(s)} = \frac{5}{(s+1)^6} \tag{14}$$

the desired set point is:

$$z_2(t) = 0.5 * 1(t) \tag{15}$$

The critical operating modes are achieved, respectively when $K_{PID1} = 0.08$ *and* $K_{PID2} = 0.474$.

The following quantities are reported from the last graphics of the Fig. 3 and Fig.4: Δy_{kp}, T_{kp}, y_{cp}, Δu, u_{cp}.

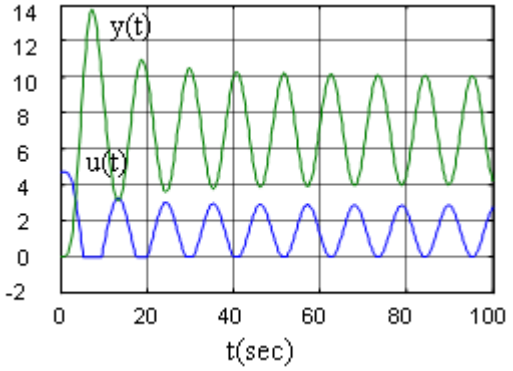


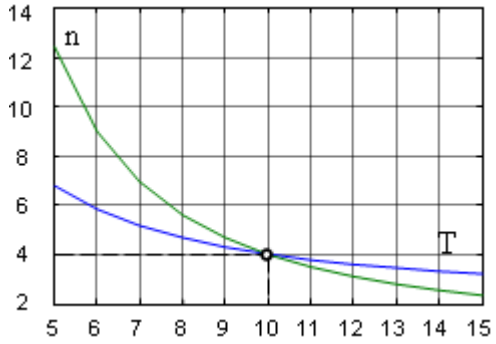
Fig. 4. PID controller operation with object with TF $\frac{5}{(s+1)^6}$

It is calculated from the reported values in Table 2: $K_P = \frac{y_{cp}}{u_{cp}}$; $\omega_{kp} = \frac{2\pi}{T_{kp}}$; $A_{kp} = \frac{\Delta y_{kp}}{\Delta u}$.

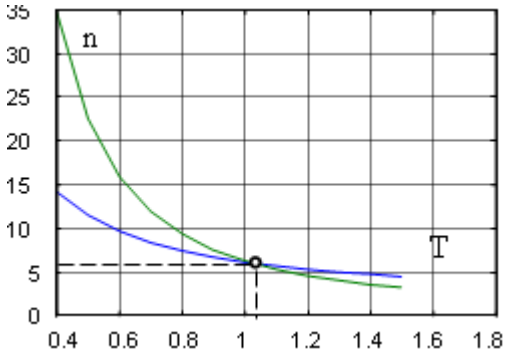
TABLE II

	K_P	$\Omega_{кр\ (r/sec)}$	$A_{кр}$
a)	49.38	0.0977	12.469
b)	4.9754	0.562	2.109

The system of equations (9) and (10) is solved graphically (for the cases (a) and (b)) by the obtained values in Table. 2.



a)



b)

V. CONCLUSION

On the basis of the results which are obtained (the coordinates of the intersections in a) and b)), the following conclusions can be drawn:

1. The case a) – it is obvious from the graphic that: T ≈ 10s and n ≈ 4, which is an accurate result for the fourth order model
2. The case b) – it is obvious from the graphic that: $T = 1.02\ (\approx 1)s$ and $n = 5.91\ (\approx 6)$, which is an acceptable accuracy for the sixth order model.
3. In its examination of the identification method, visual reportings and calculations with reasonable accuracy are made for all the models in Tabl. 1 and with different values of the parameters.When MATLAB possibilities for precisely reporting and more accurate calculations are used, the more accurate result would be achieved. It is not a problem for the reader to check that.
4. An advantage of the identification method is that, there is no need of additional operational units for its realisation, except these, used in the loop for practical realisation. It is obvious that the presence of a systemic error in established mode does not influence the results (K_P).

REFERENCES

[1] I. Dragotinov, "Technological Processes Automation", UFT, 2003.

[2] J. Badev, A. Lichev, G. Terziyski, "Identification Method for Objects by High Order Models in Closed Loop", LII International Scientific Conference on Information, Communication and Energy Systems and Technologies ICEST 2017, pp. 401 – 404, Nis, Serbia, 2017

[3] P. Cominos, N. Munro, "PID controllers: recent tuning methods and design to specification", IEE Proceedings – Control Theory and Applications, vol. 149, no. 1, pp. 46-53, 2002

[4] G. J. Silva, A. Datta, S.p. Bhattacharyya, "New Results on the Synthesis of PID Controllers", IEEE Transactions on automatic Control, vol. 47, no. 2, pp. 241 – 252, 2002

[5] F. Golnaraghi, B. C. Kuo, "Automatic Control Systems", Wiley Publishing, 2009

Platform for exploring the possibilities of creating an intelligent elevator system

Rosen Vitanov[1] and Krume Andreev[2]

Abstract – **In recent years heavily were discussed problems related to elevator equipment – their reliability, modernization and accident-prevention techniques. In this paper an autonomous data collecting platform for lifts movement parameters analysis is presented. As it consists of multiple sensors – acceleration, gyroscope, magnetometer, temperature, doppler and distance – a variety of parameters can be collected thus providing the grounds for a vast research in prevention of accident. Included are wireless communication modules, power independence and non-volatile memory - further eases research efforts by remote data collection, local-based storage for off-the network locations, script-based analysis algorithms and power-loss prone system.**

Keywords – **Lift, accident, prevention, sensor platform, lift equipment, data collection, data analysis.**

I. INTRODUCTION

In recent two years problems related to elevators emerged – as several different generations of lifts are in operation - relay, electronics and microprocessors based. Apart from some being obsolete, depreciated, started to break down due to poor, maintenance and deliberately inflicted damage. Most human involved elevator accidents leaded to injuries and some of which with lethal end - are mainly caused due to technical deficiencies in construction, human errors , improper operation, etc [1].

Due to SAMTS, 98000 are the known and registered elevators in Bulgaria. Up to December, 2016, 7063 lifts were technically reviewed, 299 discontinued work, and more than 4000 had technical issues. In 2013 there were 20 official signals for accidents – 20 with material damaged, 20 people were injured and 3 leaded to deaths. In 2014 – 25 accidents, 5 with material damages, 13 injured people and 3 deaths. In 2015 – 18 accidents, 5 with material damages, 10 injured and 2 deaths [2].

Most of those were due to missing or non-functional load control devices, misaligned elevator cabin to floor, worn friction discs and sleight, broken operational buttons, oil leakage [1]. Most of which can detected by monitoring cabin movement. In order to limit injuries and human panic in case of an accident, all elevators were equipped with phone based devices for direct conversation with support team [3]. Nowadays all elevators have some kind of intercom system usable only after an accident occures. Still, in case of major accident a direct connection to 112 is not available. The eCall standard, related to traffic and vehicle crashes, implements such functionality, and thus is implementable in lifts as well. To minimize the risk of an accident, a system that predicts and warns for such an event is needed. The first step is to build a data collecting system that will enable us to "see" what is going on in life of an elevator by develop lift movement analysis algorithm.

Purpose of this paper is to describe a microcontroller based platform for algorithm synthesis and development related to lift movement analysis (using multiple sensors - doppler radar, accelerometers, magnetometer, temperature-humidity sensors, gyroscopes, distance sensors, etc.) to predict and prevent accident in the near future. If conditions for such and event are in place – a warning shall be send to support team to take appropriate precaution measures.

II. OVERVIEW OF THE SYSTEM

On fig. 1 a block-diagram of elevator-monitoring platform is presented.

Hardware may be divided in three main subsystems. The first one is power supply unit. It provides the required voltages for the system. Power may come from the grid or from an 12V battery. In case of an power outage, the system shall continue its work.

The microcontroller is the brain of the system. It analyses data from the sensors, constantly monitors the health status of elevator and in case of an emergency, a message is sent to support team using either Wi-Fi, GSM call or 3G network. During its run, data is also collected and stored in non-volatile memory, as to be accessed at anytime by developers and engineers. In this case Atmel SAM3X8E Arm Cortex M3 processor is used with micro SD card support, a Wi-Fi, LoRA and GSM/3G modules.

To collect data multiple sensors are available to the system:

- Compass sensor – compass sensor is used in calibration process to determine the initial orientation of the device – which way is "UP", where is "NORTH" etc. HMC5883 is used – a 3 axial magnitoresistive sensor with 12 ADC and I2C bus.
- A DHT22 sensor is used to collect temperature and humidity data alongside with MLX9061. It is used may be used for compensation, error correction and fire recognition.
- MPU6050 is accelerometer with build-in gyroscope – both 3-axial. Its main purpose is to monitor for vibrations while the lift travels up and down, constantly monitoring for any

[1]Rosen Vitanov is PhD student with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: rosenvitanov@mail.bg.

[2]Krume Andreev Ph.D. student is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: andreev.k@abv.bg.

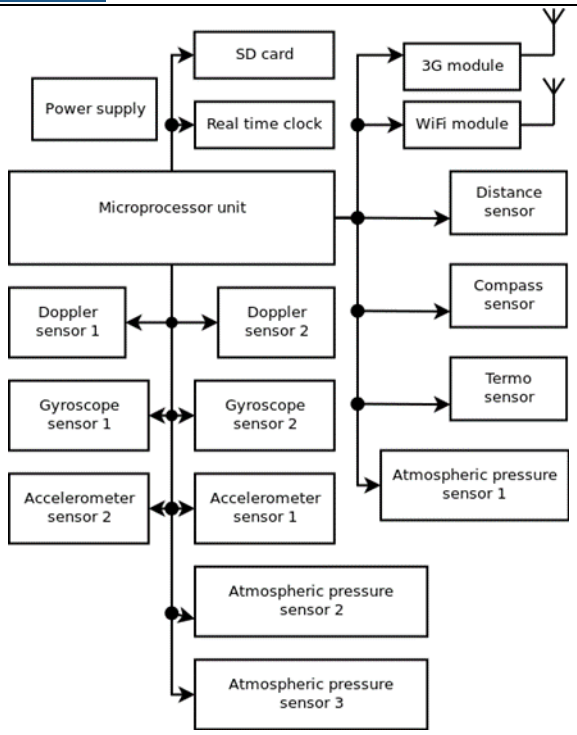


Fig. 1. Hardware block diagram

upcoming failures, thus preventing any accidents. Two of it are used – for error correction algorithms.

- Doppler sensor – a microwave radar HB100 alongside with laser distance meter and atmospheric pressure are used to detect speed, movement, direction and elevation of the lift carriage. Again data is collected from multiple sensors to allow development of failure-detection and prevention techniques.

III. Software algorithms

Following are examples of algorithms, implemented using presented hardware features. The first and most important step is to gather initial statistical data.

A. Calibrating initial data

Elevator engineer should check for any problems. If is satisfied with lift's condition the system is initiated and enters in Initial data calibration mode. This is done for the algorithm to "exam" and "learn" what is a "normal condition" for the current elevator.

The first step is to determine the orientation of the platform itself. This is done while the cabin is stationary and there are no people inside. First "down" direction is determined by detecting Earth's acceleration. Used sensor MPU6050 is 3 axial linear accelerometer which detects 9.8 m/s2 on corresponding axis. X and Y axis, used for interpreting gyroscopic data, are determined by compass sensor. The procedure is shown on figure 2.

After phase one is completed, the system should learn the "normal" way of travel. This is done by collecting statistical data while the cabin travels from the lowest to the highest and in reverse. This procedure is repeated twice more – with stops at each floor . In this way, system also "learns" how the elevator stops – accelerations, vibrations etc.

After all initialization steps completed, system is ready to collect data. It is stored on an external micro SD card as well to be used for development of new algorithms, emergency rules, warning conditions, etc. Using Internet access it is accessible any time by support team. LoRa connectivity allows new remote sensors to be attached to the system and collect data from them too.

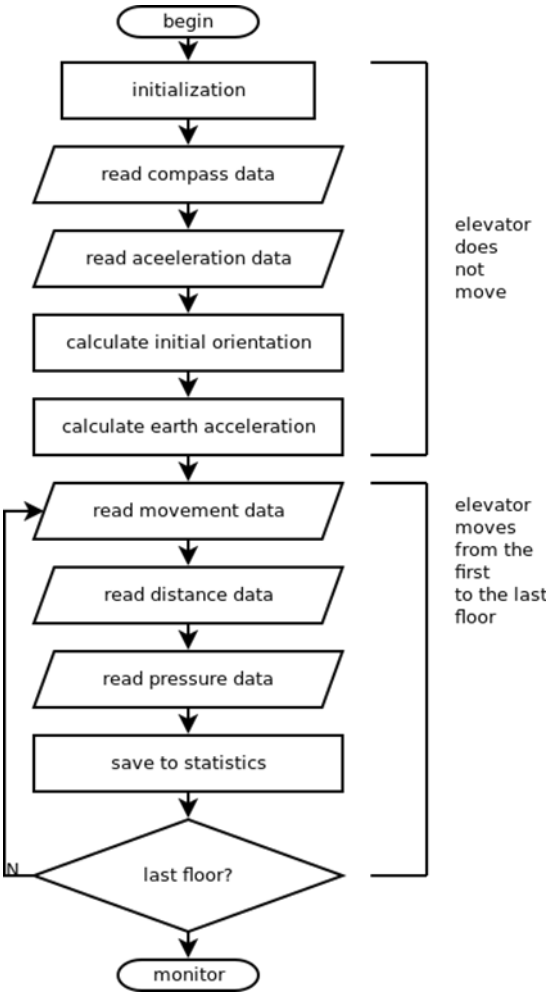


Fig. 2. Software diagram for initial data calibration

B. First floor recognition algorithm

One of the algorithms that may be accomplished with such system, that is vital for the proper functioning" is recognition of first floor. This is shown on figure 3. Using data from pressure sensors it is quite easy to determine where the cabin is located. An algorithm is shown on figure 3. To accurately

detect the first floor, no movement should be registered. This is

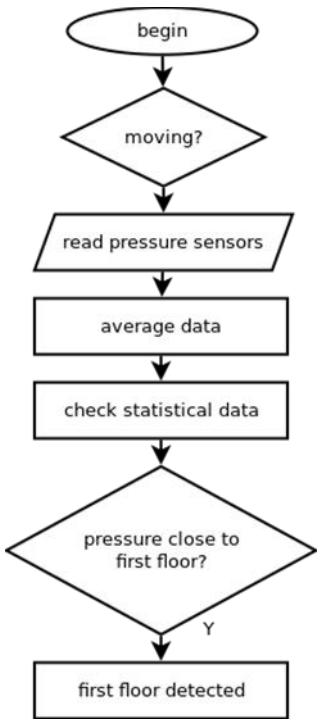


Fig. 3. Software diagram for detecting first floor

done by using doppler sensor.

If no movement is detected all pressure sensors are sampled. To reduce sensor errors a moving average algorithm is used, where windows seize equals previously determined by the operator number. This smoothen value is compared to any values stored in controller's memory (previously learned in calibration mode). Closest one is selected and the floor is recognized.

This would be an ideal situation if the weather and pressure were constant. As the pressure changes over time difference between current one and that stored in statistics for the corresponding floor varies as well. If the margin reaches predetermined value (for example 20 per cent) value in statistics is corrected. All recorded pressure values for the rest of the floors are recalculated and corrected if needed.

C. Free fall recognition algorithm

A free fall recognition algorithm may be realized, but first a closer look into acceleration data should be taken. On figure 4 is shown data collected from Z axis of an accelerometer.
In first couple of milliseconds lift cabin is stationary. Sensor is linear thus reads 9.8 m/s2 – Earth's gravity acceleration is not compensated. When free fall starts, measured acceleration reaches 0 m/s2. On figure 5 an algorithm for recognizing free fall is shown. First calibration data is read. Microcontroller constantly reads data from both acceleration sensors and calculates average value. Finally a Moving average is applied for filtering purposes. Free fall condition is detected when 0 m/s2 is read for more than 250ms. According to relation

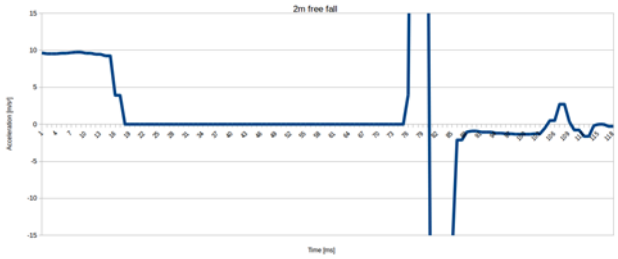


Fig. 4. Free fall data accelerometer MPU6050. X axis – time; Y axis – acceleration m/s^2

between acceleration and distance (formula 1) the cabin would have fallen with approximately 30 cm. This is necessary do filter out vibrations, jumps, etc. In such case an warning shall be sent to support and rescue team.

In addition if multiple vibrations are detected – this would be sign for an abuse and property damage. Again support team shall be called.

$$S=0.5at^2=0.5*9.8(0.25)^2=0.30625m \quad (1)$$

ADDITIONAL REMARKS

Described algorithms are an example of the capabilities of the proposed platform and cover a parameters, collected during general usage of elevators. They may be modified.

This prototype is still in development phase. As a platform, it shall provide ability for researchers to develop many new algorithms that will prevent disasters and save people's lives.

IV. CONCLUSION

Proposed platform is a great tool to help scientists to investigate, research and develop new life saving and disaster prevention techniques. A working version of the prototype shall be available in the following months. More information and measured data will be provided to all institutions, research centers, developers and students interested in this project.

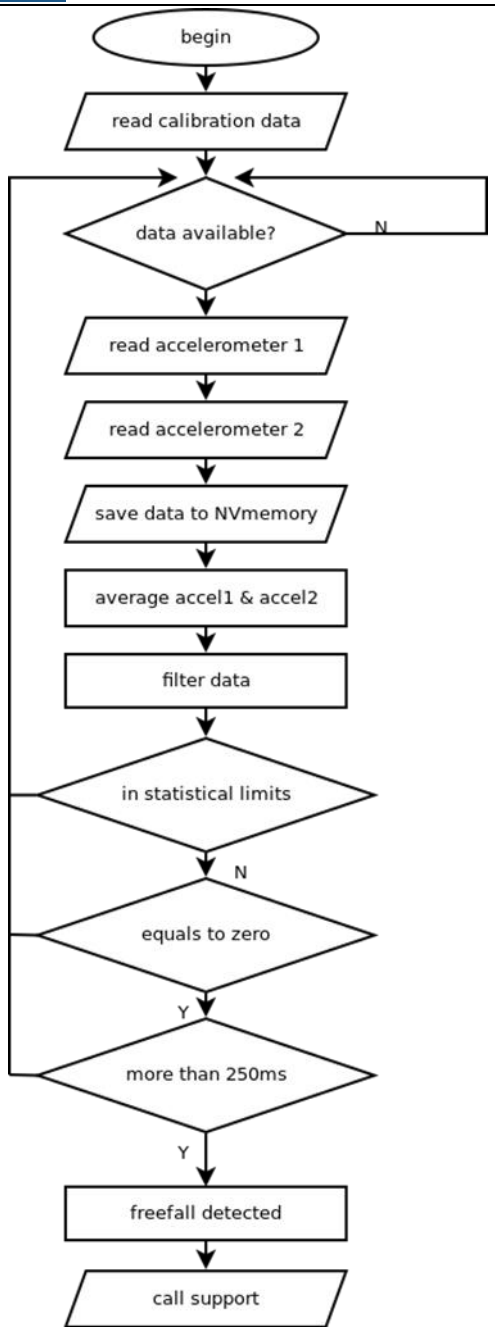


Fig. 5. Free fall data accelerometer MPU6050.

X axis – time; Y axis – acceleration m/s2

REFERENCES

[1] Assoc. Prof. Dr. K. Krastanov, "About the safe operation of lifts", Scientific proceedings XIV international congress "Machines. Technologies. Materials.", 2017 winter session, Web ISSN 2535-003X Print ISSN 2535-0021; www.mtmcongress.com/proceedngs/2017/Winter/1/16.ABOUT%20THE%20SAFE%20OPERATION%20OF%20LIFTS.pdf

[2] State agency for metrological and technical surveillance, Statistical data 2016, www.damtn.government.bg/

[3] The Ordinance on the Essential Requirements and the Assessment of the Compliance of Lifts and their Safety Devices, SG, issue. 23 25.03.2016.

Presenting Follow UP on Implemented pFMEA Methodology into Industrial Entity as a Quality Control Methodology used on a daily base

Ivo Kuzmanov[1] Silvana Angelevska[2] Roberto Pasic[3] and Ilios Vilos[4]

Abstract –**The paper presents only a segment from extensive done follow up on implemented pFMEA into industrial entity. The follow up was done at the beginning of the year 2018 after the implementation process at the industrial entity started at November 2016. The industrial entity is from the production industry - especially fire stoves production and it's the largest one in Macedonia but also one of the key players on the Balkan markets. The basic aim of the paper is to present the aimed results after the implemented FMEA methodology and the achieved benefits after a period of time (almost 2 years).**

Keywords –**FMEA, pFMEA, production system, Quality, Assurance, Quality Control, R. Macedonia.**

I. INTRODUCTION

The basic aim of the paper is to present a second follow up on a previously implemented FMEA into an industrial entity from Macedonia. The same one works into the metal cutting industry, or to be more precise the same one is maybe the largest producer of fireplaces for home use in R. Macedonia, and one of the key players on the Balkan market. Also what is more relevant is to say that the business entity is maybe one of the oldest ones in this area, or to be precise the same one has a constant production more than 60 years. Also the business entity has its unique capabilities – the largest one in this area according two main criteria's (production on a year base and the number of employees) but also has one of the most new and sophisticated and best lines among competitors. They use CNC machines in several production stages which makes the capacity maybe one of few in Macedonia with so many equipment doing precise things. On the other hand the papers presents only a small part from a second follow up on the effectivity of the implemented FMEA methodology and its benefits to the production stages and to the company itself. Also what is relevant to mention is to say that there were previously two published papers which reveals the details from the implementation (started at the year 2016) and the follow up of the same one. The first paper reveals the implementation stages, its problems, the first benefits, the changed mind of the managers after the implementation, the reducements of waste materials in production stages, the financial benefits etc. On the other hand the second published paper reveals the production stages after a while and the benefits they brought to the company. This paper presents an momentarily view, done as a fresh research at the same business entity, which presents the matrixes and it's real benefits after almost 2 years at the same business entity.

From this point of view the initial implementation brought the company a lot of reducements into the non conformities, problem solving technique implemented on a monthly level, some quality improvements, reducements of the production expenses etc. On the other hand the first review done showed up that the company had significant reducements of the waste materials, implemented problem solving technique on a daily base, significant quality improvements, and significant reducements of expenses, also bigger profits and ideas proposed from the internal workers.

So on this paper maybe the first hypothesis is to prove that the same one is still used on a daily or monthly base, than that the matrixes had maybe the same problems but with smaller RPN numbers in the same ones and finally that the level of quality in every stage is significantly improved.

Also what is relevant to say at this stage is to point that this is a third paper which is published, but as overview on the same entity from a different time sequence. At this point also what is relevant to mention is that the initial research and implementation was done by a multidisciplinary team (conducted from different persons – university professor, managers, different shift managers, workers from different work departments, workers from the warehouse and even an customer), but the follow up was done only as a monitoring to the work of the internal (business entity) team members. Also as the follow up, actually in this second follow up only a small part from the initial research is presented and shown up in this paper.

II. SHORT OVERVIEW OF THE FMEA METHODOLOGY

The presented methodology used at the initial research and used after the same one on a daily base (or in some cases used monthly) is the FMEA methodology. It's a worldwide known and recognized by companies as a method which will improve the quality, will reduce the problems, will deal with spotted

[1]Ivo Kuzmanov is with the Faculty of Technical Science at the University St. Kliment Ohridski Bitola - UKLO, Bitola 7000, R. Macedonia, e-mail: ivo.kuzmanov@tfb.uklo.edu.mk

[2]Silvana Angelevska is with the Faculty of Technical Science at the University Kliment Ohridski Bitola - UKLO, Bitola 7000, R. Macedonia

[3]Roberto Pasic is with the Faculty of Technical Science at the University Kliment Ohridski Bitola - UKLO, Bitola 7000, R. Macedonia

[4]Ilios Vilos is with the Faculty of Technical Science at the University Kliment Ohridski Bitola - UKLO, Bitola 7000, R. Macedonia

problems but primarily is used for detection and analyses of potential non conformities. Also the same one is known as a method for systematic detection of potential errors but also a one that creates potential solutions for the spotted errors. Its full name is worldwide known as Failure Mode and Effect Analyses. It is most commonly used for:

- Detection of potential errors which has a significant influence to the system, to the quality, to the work effectiveness and to the system productivity
- Evaluating the potential and spotted effects of each error or non conformity and their influence to the system. But also the same one as a method evaluates the influence over the elements, production stages, functions, sub processes and subsystems.

On the other hand the most competitive thing of the FMEA method among other methodologies is that the same one is build up and based on a team work and that the same one is the most commonly used one for continuous improvements. The improvements could be spotted in all of the production stages from the raw material department try production stages till the final product, but also seen as an improvements spotted by the customers. It's a situation where the entity could spot all of the potential non-conformities, could evaluate the same ones, could divide the non-conformities to a priority or no priority for the system at the moment but also as process of actions which will reduce the influence of the same ones to the system.

The methodological approach of the same one is based on a team work, process of evaluation of the system, and after the same ones created tabular views which actually are a multiplied numbers from three relevant factors. The same ones are the following ones: the severity, the occurrence and the possibility for detection. Actually the multiplication brings the team the RPN number (Risk Priority Number) which could be aimed by the following formula:

$$RPN = S\ (severity) \times O\ (occurrence) \times D\ (detection)$$

Each of the main criteria's (the severity, the occurrence and the detection) could be in a scale from 1 to 10 and could be precisely read from generated tabular views. So that is the reason why the highest RPN number could be 1000. And the final thing worth full for mentioning is the solving approach. Actually every team could find another solution for maybe the same problem, but the priority of the tasks is according to the RPNs. A higher number means a preventive action which should be taken as soon as possible.

The implementation of the FMEA in real industrial entity actually means that the following steps should be taken:

- Team creation
- Defining TIME for implementation
- Defining place for implementation
- Creating a structural, functional and non-conformity analyses
- Defining RPNs for each problem
- Defining potential solutions for each problem
- Realization of the recommended steps for each problem
- Additional monitoring
- Continuous improvements
- Implementing PDCA cycle (plan-do-check-act)
- Monitoring of the process
- Doing thinks from the beginning so they could achieve smaller RPNs

III. PRESENTING THE BUSINESS ENTITY AND THE PRODUCTION STAGES OF THE SAME ONE

At this stage first of all we will present the industrial entity. As previously said the same one is one of the oldest in the Balkan area in the production of fireplaces for home use, but also one of the most competitive one. The same one is on the market more than 60 years, producing and selling in R. Macedonia but also in all of the Balkan countries. Since the beginning of the 90ties the same one from a state property was transformed to a private property and since then continuous improvements to the processes, to the knowledge, to the equipment are symbols of the same one. Now more than 150 people are employed by the same one and are working on automated and semi automated processes. Also the business entity has a lot of CNC machines which are used on a daily base for production. But still there are some older machines, older processes and older buildings that are in a constant use, and could be a thing to improve in the future.

On the other hand having in mind that this is a second follow up of the same industrial entity, researching the benefits from the implemented FMEA matrix, the first think to mention in this stage is to mention that there were two successful publications of papers from the previously done implementation and the first follow up. But just to present only a segment of the benefits the first think is to divide and to present the sub processes into the industrial entity. The same ones are the following:

- Consumption (process of buying) raw materials
- Quality control on the raw materials
- Placing the same ones in a warehouse
- Segmenting the raw materials
- Process of cutting (using small and large scissors)
- Quality control
- Making appropriate holes to the material
- Using hydraulic presses
- Delivering the final product (semi product) to another process

Generally this is only the first process into the industrial entity and according to the production plans the same ones are used for the production in production stages and then as final products are placed into the warehouses for final products before selling the same ones.

On other hand just to use the same approach to the follow up the following production characteristics are also taken under consideration:

- Methodology of work
- All of the documents used while planning the work
- Machines
- Raw materials and other materials used while production
- Human factors (employees)
- Measurement instruments
- Work conditions (in different shifts)
- Customer demands

Having in mind that there was already a follow up done on this project at the business entity, but also that they use the FMEA methodology on some cases on a daily base and on some on a monthly base, the mistake factor should be smaller and smaller. So the expectation was that the monitored processes and the RPNs of the same one from the first follow up should be lower than before. But in this case because still is a process where workers do their daily activities, enrolled on the machines and a process which enrols materials from different producers still there are some mistakes, there are some non-conformities and some problems. But seeing the same ones and comparing the same ones with the past the benefits are more than visible. In this stage we could freely say that the post FMEA era into the company is more successful than the pre FMEA era of the company. In this stage also a detail worth full for mentioning is that the producer especially in the last years has some problems on the market. The main reason is because some producers offer stoves on pallets and they are more and more used from customers and buyers. So in the general strategy of the company some considerations for swathing the productions are already are on.

IV. Presenting the Results

This is the segment where the results are presented. But before showing the same ones we should say that this part is only a segment from the extensive follow up. The same one was intended to see all of the processes, but because this paper is the third one dealing whit the FMEA in this paper only the process of Transferring done pieces to the warehouse is presented. Also this part could be compared to the firs follow up done in June 2017, from which if compared the three potential failures has been reduced even more. For example the first one in June was 18, the second one it was 40 and the third one was 8. And after more than 7 months in February we got the following results presented in the following tabular view.

TABLE I
PRESENTING ONE PROCESS UNDER FMEA

PROCESS	POTENTIAL FAILURE	NUS EFFECT
Transferring done pieces to warehouse	Damaged piece	Replacing time sequences which are long, but compared to previously far more faster
	Long time required for transferring	Production delay and free work force with nothing to do at the moment
	Non appropriate conditions into the warehouse	

And seeing the table the first thought is that we have some improvements into the first field, where the replacement time is shorter than before, but also if you see the third one you could see that there is nothing there. It's because the company now is aware that if they had such problems with the conditions in the warehouse they will have a lot of damaged products (row materials and final products) and that will cost money. So they invested in the warehouse and renovated the same one, so the damaged products now aren't such a problem, and the conditions in which the raw materials and the final products are placed are better. Also it's a situation in which the workers are more happy, because they are now working in a newer building. Still in some cases they have damaged pieces in the warehouse but the conditions are not the reason for the same one. Also in addition another tabular view is given in which the Reasons are given with the appropriate RPN numbers.

TABLE II
PRESENTING REASONS FOR MISTAKES AND RPNS

POTENTIAL FAILURE	NUS EFFECT	REASON	RPN
Damaged piece	Replacing time sequences which are long, but compared to previously far more faster	Mistakes made by workers while transferring the materials	10
Long time required for transferring	Production delay and free work force with nothing to do at the moment	Transport equipment which is old, but some of the same one is replaced with new ones	30
Non appropriate conditions into the warehouse		Old building which was renovated in the year 2017	6

Seeing the RPNs at the moment and comparing with the previous one (48,70,72) the results are visible. Also the results from the first follow up were 18,40,8 so it's more than clear that this method brings results for the processes and for the company as well. But still there are things to do. That is the reason why this method is based on continuous improvements. In this case things that should be done are the following:

- Training for the workers especially for the process of transferring
- Special two week training activities for the new employees in each case
- Quality check by the workers on the machines before they start using the raw material (piece by piece)
- Generating workers which will be the ones who will transfer the materials (to know which worker is the one in charged for such an activity)
- Buying new equipments for a safer and faster transfer of the materials
- Replacing the older transport equipments – the ones that they have at the moment
- Follow up after doing the same ones

In this case, step by step with the predicted actions the company will get benefits in future. There are still some investments that should be made and which will be a financial costs at first but seeing the final result it will be a long term benefit, and the same one will return. Also at this stage the company is considering to start a new project from which with the usage of SPSS method combined with the FMEA they will first get an exact numbers in percent with GANT charts and then they will analyze the problems try a process of FMEA. So in this case they will use two relevant methods which could bring results (visible ones). On the other hand because we previously mentioned the thing, because they have lost some of the market shares in some markets, at the moment there is a consideration to switch the production and to start a production of stoves for pallets as a fuel. That plan means problems at first so in that case those two methods (FMEA and SPSS) could be the right solutions to deal with the potential problems. On the other hand this kind of a situation could be a problem for the intended steps for improvements previously showed. In that case all of the financials will go to the new production lines or the new equipment and some of the previously mentioned activities will be momentarily stopped (training, equipment for raw material transfer etc.) But still we will have to see how thinks are maybe in the next months. That could be a good material for a new paper publication.

V. CONCLUSION

The paper presents only a segment from the done research regarding an industrial entity from Macedonia. The same one is actually a second follow up on an implemented FMEA or to be precise pFMEA in the same business entity, and actually presents the benefits which were aimed into the time frame year 2016 – 2018. Seeing the results the actual benefits could be seen such as: better commitment of the management, renovated building (warehouse), some of the equipment replaced, a multidisciplinary approach etc. But also the future steps such as: training for the employees than plans to buy even newer equipment but also to use another new method as SPSS, are good proofs that they got another way of doing things. Still there are some considerations for the future activities and plans that are interesting to be monitored in future. So, this paper is the third one, but also could be only a good starting point to another papers published in future.

REFERENCES

[1] Ivo Kuzmanov, Roberto Pasic, "Results from Implemented FMEA methodology – Follow up on Implemented pFMEA", TEMEL International Journal., vol. 1, issue 2, pp. 23-27, October 2017.
[2] Ivo Kuzmanov, Roberto Pasic, Oliver Slivoski "Implementing FMEA methodology into industrial capacity from Macedonia", TEMEL International Journal., vol. 1, issue 1, pp. 18-21, May 2017.
[3] Ivo Kuzmanov, pFMEA methodology – follow up activities done into real industrial entity, February 2018
[4] Ivo Kuzmanov, pFMEA methodology – follow up activities done into real industrial entity, June 2017
[5] Ivo Kuzmanov, research conducted into real entity, 2016 -2017
[6] Ivo Kuzmanov, FMEA methodology – internal documents for application into real entities, 2016

Data Analysis and Improvement of Adaptive Weather Station for Work in Antarctic Conditions

Dian Iliev[1]

Abstract – **This paper describes the improvement and initial data analysis of adaptive weather station. The device relies on ultra-low power micro architecture and adaptive power distribution mechanism. It is dynamically reconfigurable for working in high performance, real time transfer mode with direct operator control; and ultra-low power, fully autonomous, self-monitoring, long-term measurement mode. For convenience the collected data of the environmental parameters could be initially analysed and visualized by specialized end-user software tools.**

Keywords – **Weather Station, Ultra-Low Power Management, Self-Monitoring, Data acquisition.**

I. Introduction

In order to support a research of Bulgarian Antarctic Expedition was requested to create an adaptive compact system that ensures all the necessary measurements to support experiments performed in real-time and able to autonomously provide series long-term measurements, by dynamically reconfiguring its working rate. First prototype was used by 24-th Bulgarian Antarctic Expedition.

Purpose of the current paper is to present improvements in the system used by 25-th Bulgarian Antarctic Expedition, and the data received during the active period.

Design and development of the adaptive weather station, calibration process, and initial experimental data are described in series of papers ([1], [2], [3]), thus here is provided just a brief description.

II. Design and Improvements

A. Hardware Changes

Key factors affecting the development of the system with the specified requirements are the working conditions under which the system operates, power consumption, autonomy and cooperativeness. A special feature is the combination of high power consumption real-time monitoring system of the environmental parameters and minimum power consumption requirements for long-term measurements. For the purpose of which was developed a specialized power control system that could isolate from the power source all unnecessary subsystems, for the duration of their passive states, and switch them on again when necessary. The system is equipped with two power sources – rechargeable accumulator to support routine daily and power demanding operations; and back-up battery to be used during long-term measurements, where accumulator capacity may not be enough.

The new version of the device is equipped with enhanced integrated Li-ion linear charger and system power path management device. Initial one was proved to have a design defect. New one allows faster charging of a high-capacity accumulator. In addition to the new charge controller is added companion DC-DC module, allowing battery charging to be done with high-voltage (up to 36V) power source (solar panels, accumulators, generators, etc.).

The control subsystem is based on a microcontroller and includes user interface for direct work with the system, interfaces for data transfer with other systems, real-time clock, and a data memory.

The device has USB interface designed for direct transmission of data to the computer systems, so the operator can monitor real-time measurements.

Real-time clock provides a time stamp required for the synchronization of the measured parameters with a global database.

The sensor subsystem includes sensors and peripherals required for the operation of the station – temperature sensors, pressure sensor, humidity sensor, light sensor, and magnetic field sensor.

Original humidity sensor used by design (HDC1008) is replaced by enhanced version (HDC1080 [5]). New one has high accuracy (±2%) and better power management.

Pressure sensor is also replaced with more enhanced version (BMP180 [6]), for lower altitude noise and fast conversion time.

The sensor subsystem includes interfaces for anemometer [4], GPS module, accelerometer, and four expansion interfaces for adding additional sensor modules.

Fig. 1 shows the updated version of the adaptive weather station

B. Firmware Updates

System management is executed by embedded software, including algorithms for determining the operating modes of the system (autonomous control or subordinate work), powering the sensors, synchronization, error detection, analysis and reconfiguration of the operating modes, the data logging and transmission.

There are three main algorithms that build the firmware:

- Self-diagnostic
- Power Management
- Data collection

[1]Dian Iliev is with the Technical University of Sofia, Department of Electronics and Electronics Technologies, Faculty of Electronic Engineering and Technologies, 1000 Sofia, Bulgaria, 8 Kl. Ohridski Blvd., e-mail: d.m.iliev@gmail.com.

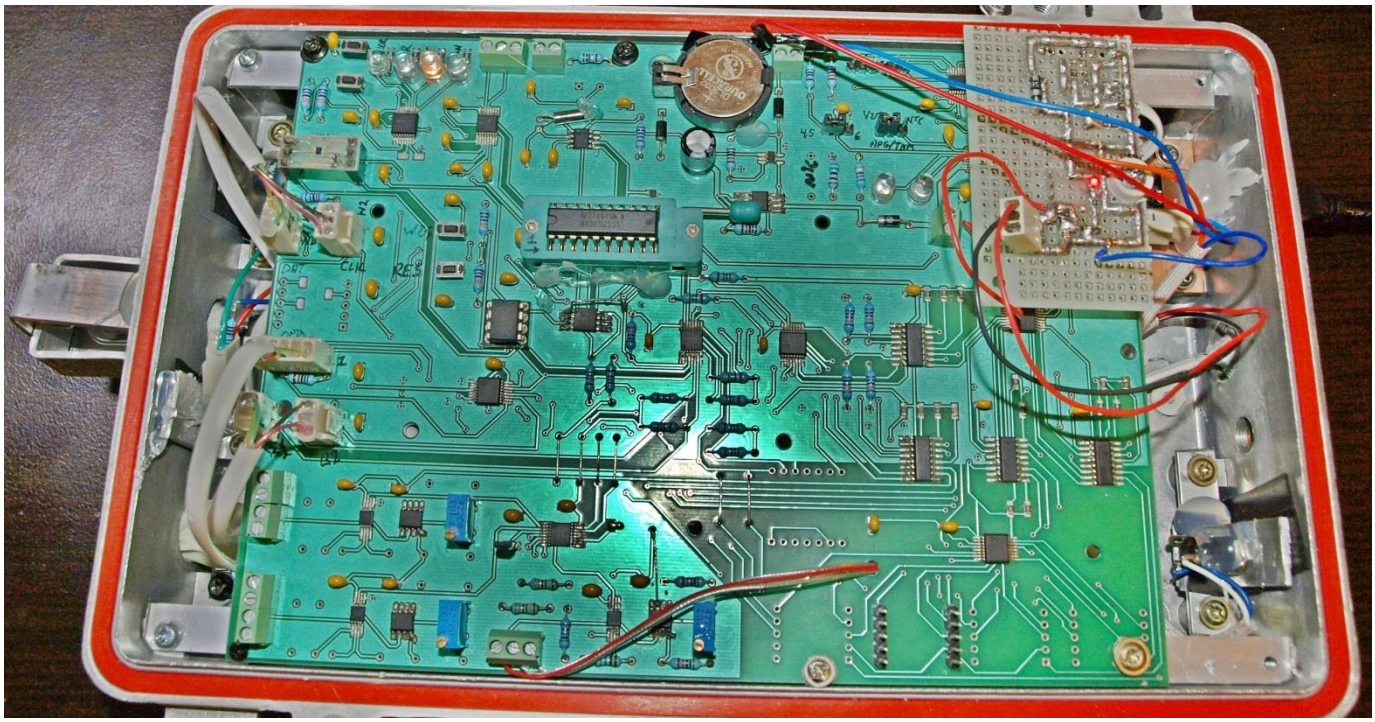


Fig. 1 – Evaluation board of the Adaptive Weather Station

Self-diagnostic algorithm is executed during the first (diagnostic) system scan after restart. Its main task is to check the system health status and to configure appropriate working mode. It is executed in two phases.

Component detection - during this phase firmware is scanning all peripherals within the system. Each scan is performed for a predefined period of time (2x sensor's maximum response time); if a response does not appear during this time frame, the program cancels the scanning process, flag the sensor as unusable, and log an error message. Once detection scanning is completed, watchdog (WDT) is responsible to monitor the system during its normal operation. Component detection is performed only once after restart, as this is high power consuming operation. If sensor fails during normal working mode, WDT will restart the system and the failed sensor will be detected during new diagnostic scan. New firmware version includes extended set of diagnostic codes, which provide more detailed information of diagnostic conditions.

Error detection – this phase is active during the normal operation mode of the system. It logs all errors that could appear – mode change, reset, manual scan requests, inappropriate user configurations (invalid input data, syntactic error, invalid command, etc.), data memory overflow, etc.

Power Management algorithm is executed immediately after initial initialization and have the responsibilities to perform power-up and power-down sequences necessary to prepare the peripheral devices for work, and to switch them off to save the power. This program is in direct service of data collection algorithm. Additionally, this program monitors the traffic upon the user interfaces and can change the working mode from user-controlled (UI Mode) to autonomous if a defined time of user inactivity expires, and vice versus if the user request control. In cooperation with Data collection algorithm, Power Management also monitors the battery status, and can switch between primary and secondary source depending on the battery levels. It can also cut down the power exhaustive sensors, if the secondary source drops below predefined thresholds, and keep the system active for longer periods.

Data collection algorithm is the main program that is running during normal operation mode. Its main task is to collect data from the sensors and record them in the embedded memory. It also provides vital data for Power Management and Self-diagnostic algorithms in order to keep them function properly. It is also responsible for user communication and all data exchange including acceptance of control commands and data conversion for real-time user usage.

New version of the firmware includes update that allows this algorithm to save one additional scan data on every hour in additional data memory chip connected to sensor bus. This operation provides data and system operation back-up in case of main memory bus operation fail or main memory malfunction.

C. Other Improvements

During the development of this version of the device, several mechanical improvements are done.

Pt100 is now inside the box for protection and thermally connected to the box for better thermal stability.

Optical sensor is now integrated in special glass casing and aluminum shield.

USB connector, LED indication, wake-up button, high-power jack, humidity and barometric sensors are now secured inside the box, accessible through isolated on-box openings.

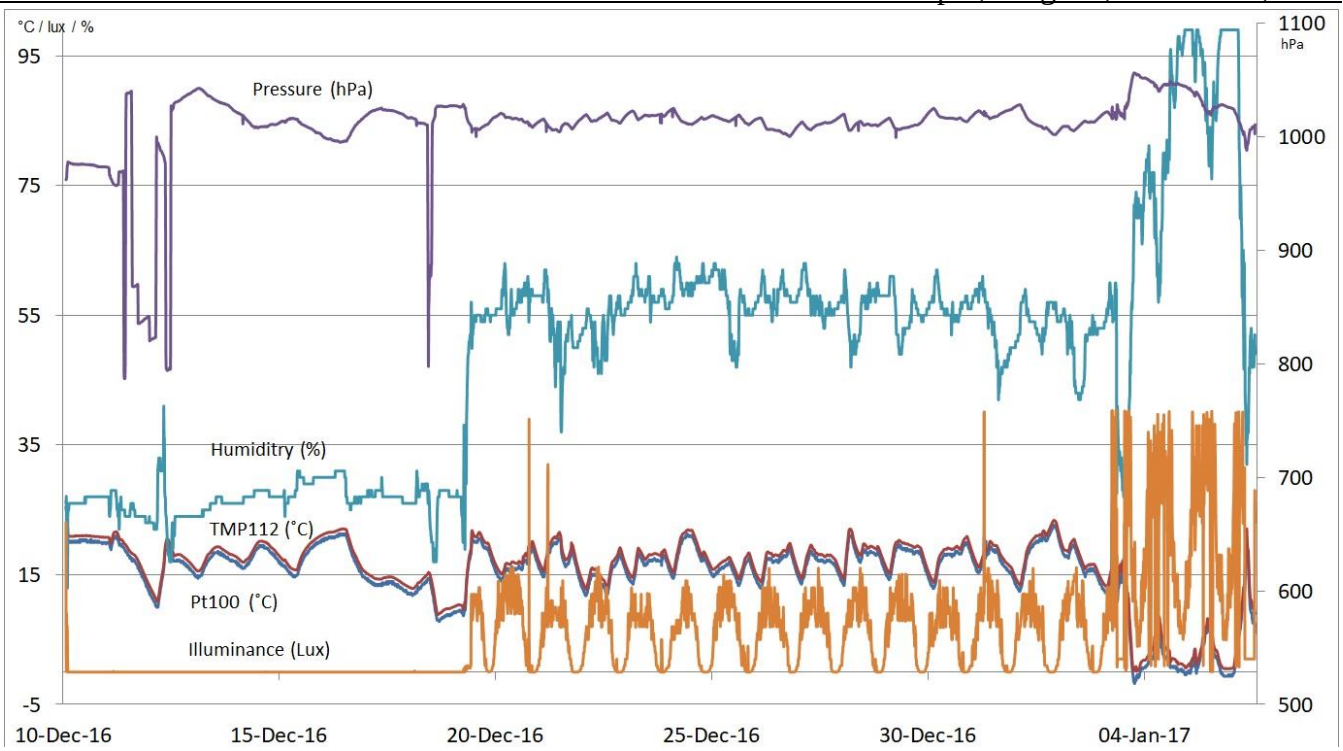


Fig. 2 – Initial Data from 25-th Bulgarian Antarctic Expedition

The computer application, designed to ease the work with the system, is also updated. New version includes several bug-fixes and new automation request mode. The application enables direct recording and storing data obtained during real-time operation of the station. It can be processed immediately and visualized on the display for direct monitoring and / or exported as a file for further analysis. It also provide current device status data, and set of options for configuration setup.

III. Results

Fig. 2 presents the data collected during the active season of from 25-th Bulgarian Antarctic Expedition. It contains three clearly defined regions of data.

First part is the transportation data. It is very easy to define this period as illuminance sensor is covered with a protection shell, thus it's data is pointing to zero. Specific drops in the temperature and pressure clearly define plane flights.

Second part is a period of real-time measurement used to support a research work. System is located near the rest of research equipment and provides reference data for it. Period of this time is about two weeks, which is clearly visible from the daylight oscillations of the illuminance sensor.

The two temperature graphs have identical trend lines, where the TMP112 is around 0.5°C lower than platinum RTD, which is matching the specification of the sensor, [7]. The graph of the platinum RTD is much more stable in comparison with the previous version (see [3]), which is in direct result of the improvements of this version of the system.

Third period corresponds to the preparation of the system for the winter season. It is moved to new location and secured to a wind-turbine mast in order to withstand to the harsh winter conditions. This change is clearly visible from the highest illuminance values. Daily cycles are still good outlined. Fig. **4** provides more detailed view of this period.

This part shows a sudden change of the weather – lower temperature, high humidity, and higher pressure. Data trends of the different parameters show a good correlation, which is expected and confirms data reliability.

There are some disturbances in the illuminance, which are very sharp and presume to be indication of a human interaction with the system. This assumption is also confirmed from the correlation between those disturbances and the intensity of the data log as shown on Fig. 3.

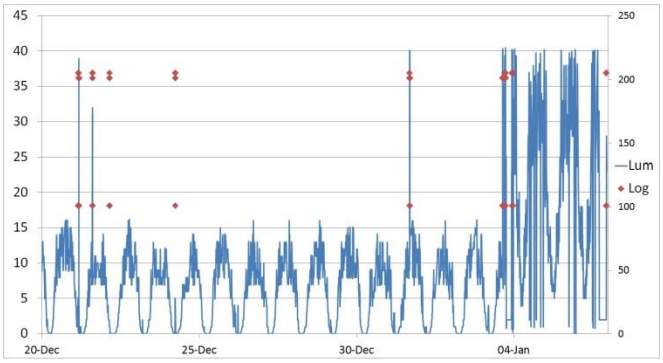


Fig. 3 – Data log vs Illuminance

In addition to the presented data, Fig 5 presents the data from the magnetic field sensor. This sensor is a chopper-stabilized Hall IC with great sensitivity stability over temperature. Its main purpose is to detect strong magnetic fields that may affect the field tests or measurements.

As the diagram shows, there are no such events appeared in this period, except some small deviations during human interaction as seen from the correlation with data log.

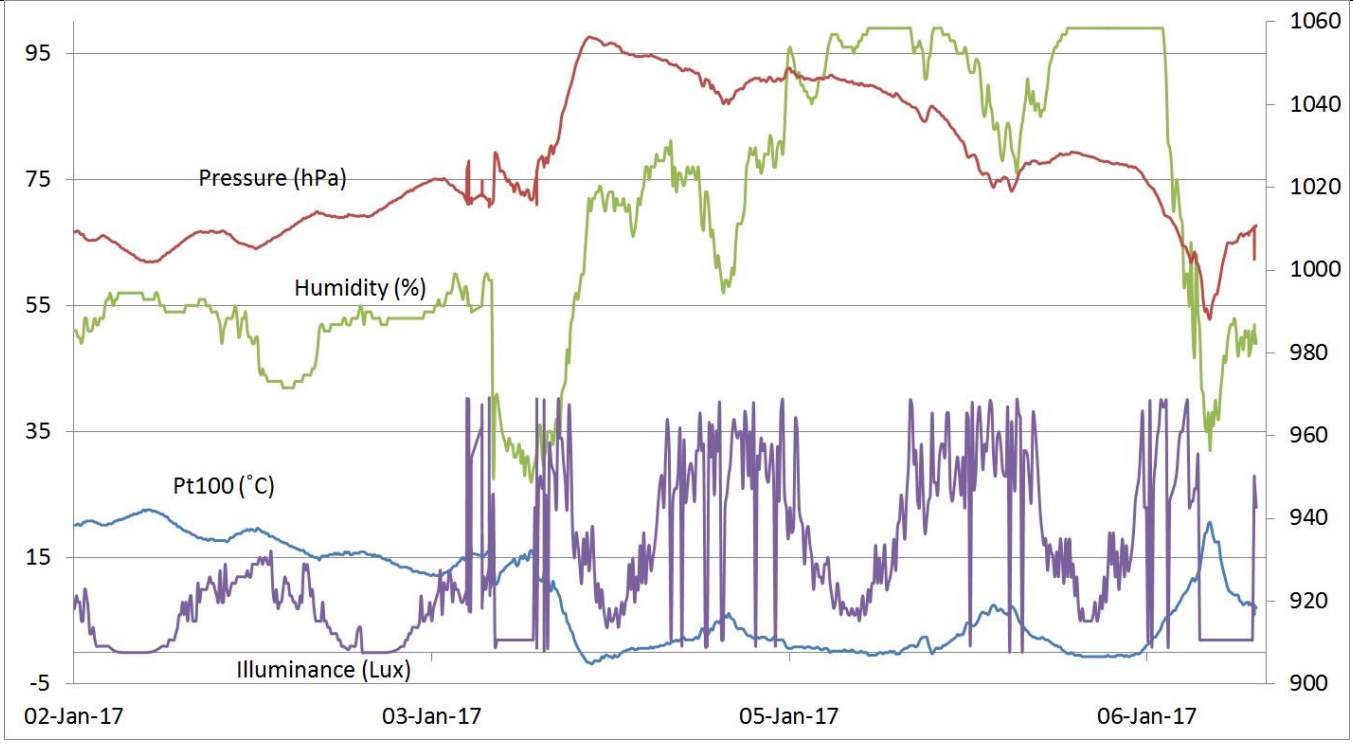


Fig. 4 – Detailed data from the third period

Interesting here is that even thought this sensor is not designed to accurately measure Earth's magnetic field, the data is in good correlation with it, as seen from the trend-line on the diagram.

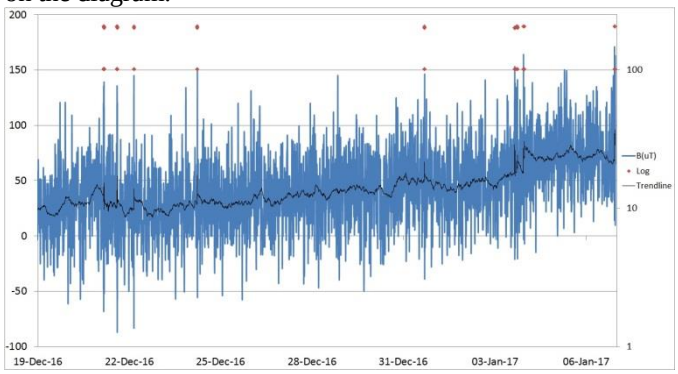


Fig 5 - Magnetic field data

IV. Conclusion

Bulgarian Antarctic Expedition reported there was a massive storm during last winter season, which cause lots of damage to the equipment and buildings, including the Adaptive Weather Station. The remains of the system will be delivered later this year.

Specialized strain gauge anemometric system is under development [4], as an evaluation of the project in aspect of hardware design.

For a future evaluation of the project in aspect of software environment is planned development of specialized online tool that will be able to transfer the data from the system through web application.

Acknowledgement

The present research is supported by the Bulgarian National Science Fund under Contract И02/11/2014.

The author would like to express highest gratitude to Lyuben Aleksiev for his help.

References

[1] D. M. Iliev, V. N. Gourev and M. G. Mitev, "Design of Adaptive Weather Station with Reduced Power consumption" 2015 XXIV International Scientific Conference Electronics (ET), Sozopol, 2015, pp. 171-174. ISSN 1314-0078.

[2] D. M. Iliev, V. N. Gourev and M. G. Mitev, " Development of Software Environment for Adaptive Weather Station with Reduced Power Consumption" 2016 XIII National Scientific Conference "Electronics 2016" 12-13 May 2016.

[3] D. M. Iliev, V. N. Gourev and M. G. Mitev, "Realization of adaptive weather station for work in Antarctic conditions" 2016 XXV International Scientific Conference Electronics (ET), Sozopol, 2016, pp. 1-4. doi: 10.1109/ET.2016.7753487.

[4] D. M. Iliev, M. G. Mitev and V. N. Gourev, "Design of strain gauge anemometer for work in Antarctic conditions" 2017 XXVI International Scientific Conference Electronics (ET), Sozopol, 2017, pp. 1-4. doi: 10.1109/ET.2017.8124351.

[5] HDC1080 Low Power, High Accuracy Digital Humidity Sensor with Temperature Sensor - SNAS672 – www.ti.com

[6] BMP180 Digital Pressure Sensor - www.bosch-sensortec.com

[7] TMP112 High-Accuracy, Low-Power, Digital Temperature Sensor - SLOS887 – www.ti.com

The Elements of Low-Cost DCS for Electricity Consumption Control

Viša Tasić[1], Marijana Pavlov-Kagadejev[2], Radoš Jeremijić[3], Vladimir Despotović[4],

Olivera Tasić[5], Ivana Stojković[6]

Abstract – **The article describes elements of low-cost Distributed Control System (DCS) for electricity consumption control and its application. The main elements of such DCS are power or current transducers, microcontrollers or PLCs with communication modules (Wireless, Ethernet or USB) and PC workstations. Low-cost DCS can be easily applied for almost all types of electricity consumers. Using the proposed concept, it is possible to achieve optimal electricity consumption because it offers the possibility to apply a control algorithm to all consumers that are included in DCS. This hypothesis has been verified in practice by applying the proposed DCS concept to the reduction of electricity consumption of electric motors in the ventilation systems.**

Keywords – **low-cost, control, electricity, consumption**

I. INTRODUCTION

Department of Industrial Informatics, at the Mining and Metallurgy Institute Bor (MMI Bor), Serbia, has a long tradition in designing of the real-time DCS for monitoring and control of the industrial processes and electricity consumption [1-5].

The objective of electricity consumption control is to avoid exceeding one or more specified chargeable demand set-points. This means that energy must be limited to a fixed value. In this way, the optimization system actually limits the amount of imported energy during high demand measuring periods by shutting down certain power consumers, which is then compensated in the subsequent period. In order to determine which power consumers can be shut down at the specified point of time an extensive analysis must be conducted. Also, the constant and reliable communications between the elements of such DCS are of the critical importance. That is the reason why network nodes and segment building have to be carefully planned.

Most of our previously realized systems for electricity consumption control are based on the transformer substations control level (cells level). In order to make the DCS system cheaper, but also to be applicable to almost all consumers, we decided to install cheaper components (transducers, controllers, communication modules, etc.). The new DCS should satisfy the requirement that the measurement accuracy remains unchanged in relation to previously implemented systems (accuracy class 1). Such DCS use low-cost devices for the measurements, acquisition and data transfer. The proposed low-cost DCS can be easily applied for the control of electricity consumption for almost all types of electricity consumers. The elements of realized and implemented low-cost DCS will be presented in the sequel.

II. DCS HARDWARE

A. Power and Current Transducers

The basic elements of each DCS for electricity consumption control are power and current transducers that adjust electrical signals for further processing and transmission. Different types of measuring transducers have been recently developed for measurement of the electrical power and current of consumers.

Examples of the realized printed circuit boards of such devices are shown in Fig 1 and Fig 2. The power transducer (shown in Fig. 1) gives the standard voltage (0-5V DC) or current signals (4-20 mA) at the outputs, proportional to the active and reactive electric power at input, with less than 1% nonlinearity (using Aron measuring methods).

The current transducer (Fig. 2) gives the standard voltage (0-5V DC) signal as an output, proportional to the input electric current, with less than 1% nonlinearity.

B. PLCs and Microcontrollers as DCS Network Nodes

Output signals from the power or current transducers are connected to the inputs of the PLC or microcontroller that serves both as data logger and DCS network node.

[1] Viša Tasić is with the Mining and Metallurgy Institute Bor, Department of Science, Zeleni bulevar 35, 19210 Bor, Serbia, e-mail: visa.tasic@irmbor.co.rs,

[2] Marijana Pavlov is with the Mining and Metallurgy Institute Bor, Department of Industrial Informatics, Zeleni bulevar 35, 19210 Bor, Serbia, e-mail: marijana.pavlov@irmbor.co.rs,

[3] Radoš Jeremijić is with the Dielectric DOO, Dr Milovanovića 15, 19210 Bor, Serbia, e-mail: radosh2009@gmail.com,

[4] Vladimir Despotović is with the University of Belgrade, Technical Faculty in Bor, Vojske Jugoslavije 12, 19210 Bor, Serbia, e-mail: vdespotovic@tf.bor.ac.rs,

[5] Olivera Tasić is with the Mechanical and Electrical High School Bor, Zeleni bulevar 24, 19210 Bor, Serbia, e-mail: oliverina.posta@gmail.com,

[6] Ivana Stojković is with the University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, e-mail: ivana.stojkovic@elfak.ni.ac.rs

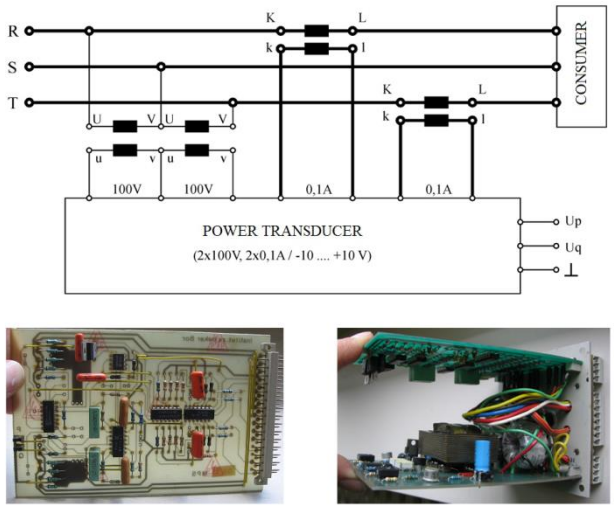


Fig. 1. The principle of power transducer connection to the electric grid with the realized printed circuit boards of transducer

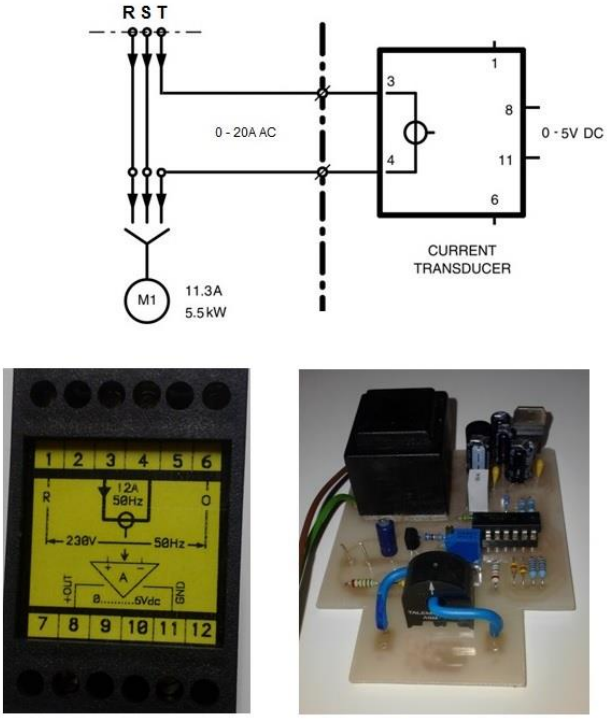


Fig. 2. The principle of current transducer connection to the electric grid with the realized printed circuit boards of transducer

In the case of local control, as a result of processing the input signals the controller will generate the appropriate control signal based on the control algorithm. When it comes to implementing more complex management algorithms, i.e. when the controller is a DCS node, then the measurement results are passed to the decision-making DCS node. From the control level of DCS the commands are transferred back to the DCS nodes. So that, reliable communications between the DCS nodes are very important for it's functioning.

The Portable Measuring Station (PMS) and Universal Measuring Station (UMS) devices has been used as a core DCS nodes in the earlier realizations of DCS [3, 5].

In order to add new functionality to already existed DCS nodes, the interface module with the Arduino Mega 2560 microcontroller was realized. This board is shown in Fig 3. We chose the Arduino Mega because the language for its programming is similar to the C language. Further, because Arduino Software IDE is an open source environment and free. Also, the existing experiences we had in working with the Arduino platform indicated that this is an extremely stable platform. Finally, the Aduino Mega microcontroller was chosen because it's cheap, and has enough analogue and digital inputs and outputs for planned applications.

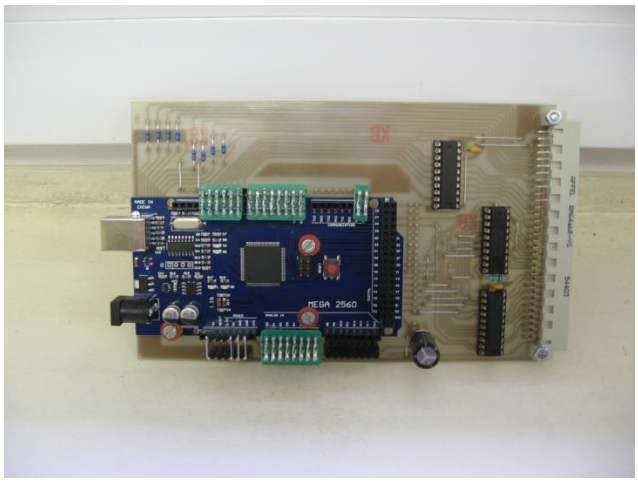


Fig. 3. Arduino Mega 2560 interface board

Arduino Mega 2560 microcontroller has 54 digital input/output pins (of which 15 can be used as PWM outputs), 16 analog inputs, 4 UARTs (hardware serial ports), a 16 MHz crystal oscillator, a USB connection, a power jack, an ICSP header, and a reset button [6]. When Arduino board serve as an UMS CPU (shown in Fig 4.), then a DCS network node is provided that can receive up to 64 analog inputs (0-5V DC), 64 digital inputs, and 32 digital outputs.

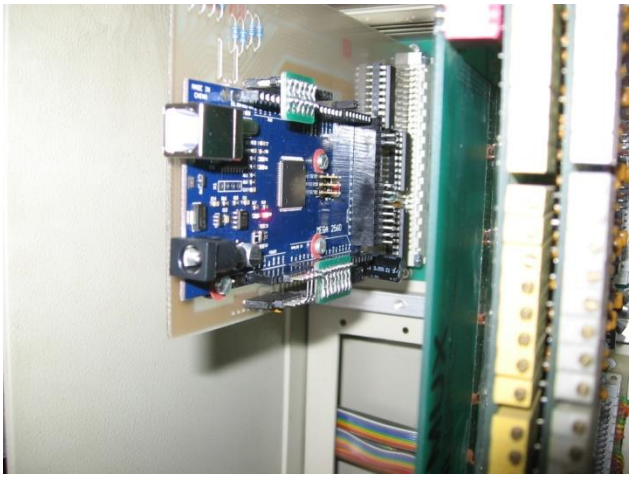


Fig. 4. DCS node with an Arduino Mega board used as UMS CPU

The Arduino Mega interface board can be used as a DCS network node. In order to ensure efficient connection of the Arduino Mega board with other DCS nodes, additional modules, so-called shields, can be used. Fig 5. shows the Arduino Mega board with the GSM module attached (Arduino SIM800C Module).

The Arduino GSM shield allows an Arduino based board to connect to the Internet, and send and receive SMS, using the GSM library.

Fig. 5. Arduino Mega interface board with the GSM module attached (Arduino SIM800C Module)

Expanding the functionality of DCS system demanded the installation of new transducers and controllers. All output signals from the different type of transducers can be connected either to the UMS inputs or to the Arduino Mega interface board inputs. Also, since the most of DCS nodes and other elements were dislocated, they have to be connected into the single DCS network.

C. DCS Control and Interaction Nodes

DCS network can be consisted of several dislocated segments. PC workstations are DCS control and interaction nodes. Depending on the control algorithms, such PC workstations can manage parts of the DCS system, or be subordinated to a single central workstation.

III. DCS SOFTWARE

A. DCS Node Software

As it was described in the previous session, in order to obtain a low-cost DCS node, an Arduino Mega board was used, either as UMS CPU or as stand-alone DCS node. Arduino Software (Integrated Development Environment - IDE) is an open source environment that was used for writing the control programs that provides data acquisition, communication (USB, Ethernet, and wireless) and control actions [7].

Arduino platform is chosen as a node of DCS because it has shown good stability over time, so that, it has been frequently used in a many Wireless Sensor Network (WSN) projects.

B. DCS PC Workstation Software

Supervising Control and Data Acquisition (SCADA) real-time application, named Process Control Program (PCP) [8], is upgraded in order to support novel DCS elements and control principles (shown in Fig 6.).

PCP is developed using Microsoft Visual C++ [9]. It is based on a client/server architecture running on PC workstations. PCP enables integration of the DCS network nodes in a complex DCS network. It has a complex structure and consists of several modules.

Main program modules are designed for communication with DCS nodes, real-time data processing and presentation, interaction with DCS nodes according to the appropriate algorithm, reporting, data archiving, off-line data processing and database management. Interactions with the electricity consumers are performed according to the control algorithms considering the actual values of power consumption.

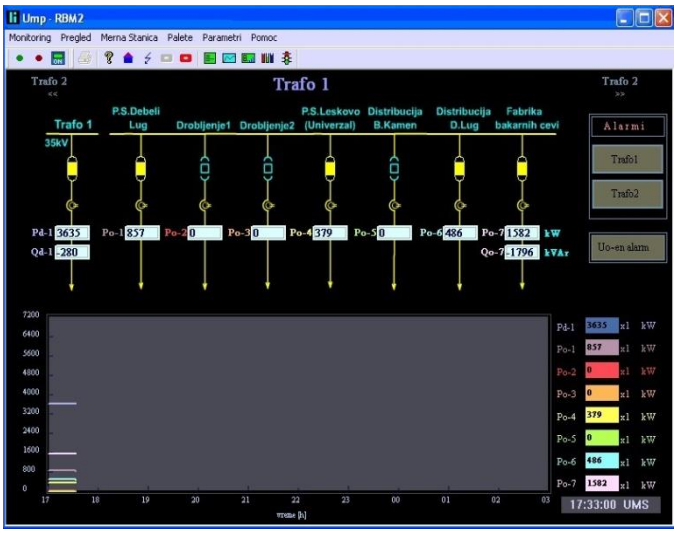


Fig. 6. PCP application screenshot

Furthermore, in order to present data about actual electricity consumption PC workstation is used to host web server. It works under the Linux OS with the SQLite database and the Apache web server installed. A Python script that runs every 3 minutes (another time can be set as needed) was used for the data collecting from the Arduino DCS nodes and to records them into the SQLite database.

Python script also generates files with the data needed for the graphical presentations on the web page. Web pages are created by using PHP and JavaScript. Arduino DCS nodes also work as web servers wait for HTTP requests from the host and respond to them by sending measured values from the transducers connected to them.

IV. An example of the realization of a low-cost DCS

As an example of the implemented low-cost DCS system, on/off control principle of a three-phase low voltage electric motor by measuring the current in one phase will be shown. This electric motor is part of the ventilation system of the factory. The ventilation system of the factory hall consists of several similar electric motors. Those are consumers who can be stopped, if necessary, for a few minutes in order to avoid the peak power, or exceeding the allowed value of the electricity consumption of the factory. The DCS node is added into the existing overload protection system of this electric motor, as shown in Fig 7. This enables the measurement of current in the single phase of electric motor, or indirectly, assuming a symmetric load at all phases, provide information about actual electric power of this consumer. This information is forwarded via wireless network to a DCS control level (PC workstation). Based on actual electricity consumption of the factory, and other elements contained in the control algorithm, DCS decides whether to turn off the electric motor or not.

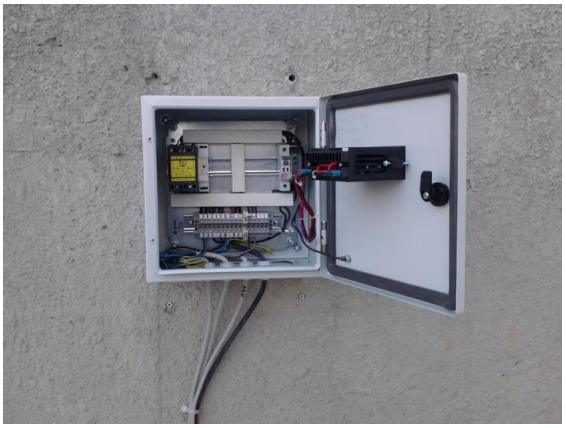


Fig. 7. The cabinet with elements of the electric motor overloads protection system

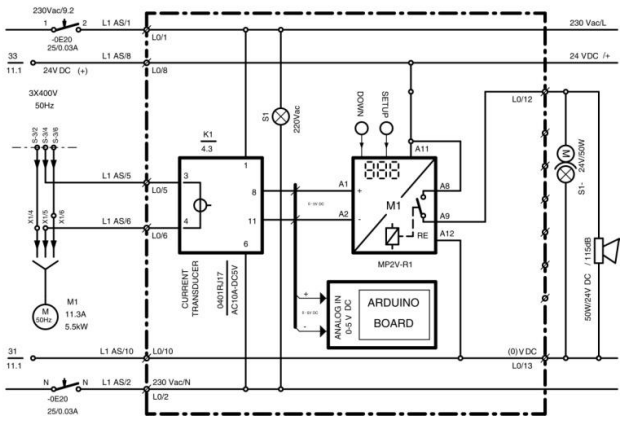


Fig. 7. Electric scheme of the electric motor overload protection system with the addition of DCS node

Conclusion

Development of production technology of microcontrollers and measuring devices leads to the development of low-cost devices that are commercially available. The aim of this paper was to present new low-cost DCS system elements for monitoring and control of the electricity consumption. Unlike the existing DCS, which are based on the consumption control at the transformer stations level, this paper presents the concept of a low-cost DCS, which can be applied to almost all consumers.

In this way, the possibility of optimizing the electricity costs is increased, as is shown in the example of the on/off control of the electric motor. In the upcoming period, we will work on the implementation of the proposed DCS concept, in order to confirm its efficiency in practice in comparison with the existing DCS.

Acknowledgement

This work was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia, project no. TR33037: “Development and Application of the Distributed System for Monitoring and Control of Electrical Energy Consumption for Large Consumers."

References

[1] D.Milivojević, V.Tasić, M.Pavlov and V.Despotović, "Synthesis of DCS in Copper Metallurgy", ICEST 2007, Conference Proceedings, Book 2, pp. 629-631, Ohrid, FYR Macedonia, 2007.

[2] V. Tasic, D. R. Milivojevic, M. Pavlov and V. Despotovic, "Reducing the Costs of Electricity in the Copper Mining and Smelting Complex RTB Bor (Serbia) by Using of the Distributed Control System", MIPRO 2010, Conference Proceedings, pp. 664-667, Opatija, Croatia, 2010.

[3] V.Tasić, D.Milivojević, M.Pavlov, V.Despotović, D. Brodić, "Microcontroller Based Systems for Peak Load Reduction, " Proceedings of 35th International Convention, MIPRO 2012, pp. 919-923, Opatija, Croatia, 21.5.-25.5.2012.

[4] V.Tasić, D.Milivojević, V.Despotović, D.Brodić, M.Pavlov and I.Stojković, "Communications in Realized Industrial networks", ICEST 2013, Conference Proceedings, pp. 619-622, Ohrid, FYR Macedonia, 2013.

[5] V.Tasić, V.Despotović, D. Brodić, M.Pavlov, D.Milivojević, "Twenty years of monitoring and control of electricity consumption in RTB Bor, Serbia", MIPRO 2013, Conference Proceedings, pp. 1134-1138, Opatija, Croatia, 2013.

[6] https://store.arduino.cc/usa/arduino-mega-2560-rev3

[7] https://www.arduino.cc/en/Main/Software

[8] D.R.Milivojevic, V. Despotovic, V.Tasic, M. Pavlov, "Process Control Program as an Element of Distributed Control System", Information Technology and Control, vol. 39, no. 2, pp. 152-158, 2010.

[9] B.Stroustrup, *The C++ Programming Language*, Addison-Wesley Pub Co, 3rd edition, 2000.

ENERGY SYSTEMS AND EFFICIENCY, AND ELECTRONICS – POSTER SESSION

Remote Platform for Photovoltaic Panel Characterization

Boyanka Nikolova[1], Georgi Nikolov[2] and Marin Marinov[3]

Abstract – **An important factor for the successful installation and maintenance of solar panels is the good training of staff. Good training involves carrying out appropriate practical activities. The present paper introduces a platform for remote measurement of base photovoltaic parameters, based on a source-measure unit, graphical programming environment and technology for remote access to hardware via a standard web-browser.**

Keywords – **Active teaching methodologies, LabVIEW programming, Photovoltaic characterization, Remote panels, Source-measure units.**

I. Introduction

Solar energy systems have become an appropriate source of renewable energy, and are now widely used for a variety of domestic and industrial applications. In particular, photovoltaic (PV) power is an established technology and has grown rapidly over the last ten years. Photovoltaics offer many advantages such as: negligible environmental pollution, high reliability, absence of noise, low maintenance cost, etc. The characteristics and parameters of photovoltaic systems are variable because the amount of sunlight energy received by any surface depends on many factors including season, geographical location, time of the day, local weather, and local landscape. For this reason, installing and maintaining such systems requires highly trained staff. The training of this staff requires implementation of new pedagogical methodologies, which integrate information and communication technologies. Those modern learning methods allow students to train at anytime and anywhere. The new teaching methodologies are called "active methodologies", and can be divided into blended learning, flipped learning, ubiquitous learning, work-based learning, problem-based learning and challenge-based learning. All above listed methodologies require practical work. On the other hand, contemporary practical exercises require appropriate measuring equipment and remote access to it.

In order to analyse the important PV parameters, specialized measurement devices are needed. There are a number of commercial measurement systems developed for the semiconductor industry by Keysight Technologies, Tektronix, Inc., and others. These units generate and measure voltage or current and can be controlled by a computer with a wide range of features and capabilities including four-quadrant operation. Of course, these so-called Source-Measure Units (SMU) require software to download the data, save it, and calculate relevant photovoltaic parameters [2]. A very popular and appropriate software platform for control of measuring instruments is the graphical programming environment LabVIEW. This environment also has built-in capabilities to create remote experiments. A key feature of LabVIEW for distance learning is the support of technology called Remote Panels. With this feature, it is possible to create an intuitive user interface for instrumentation and publish it on the World Wide Web. This internet-based LabVIEW technology comes with a web server that automatically deploys the application, and on the client side the user is able to see and manage the user interface by standard web-browser.

The present study proposes a virtual platform for remote characterisation of low power photovoltaic panel. This platform is designed for educational purposes and consists of source-measure unit U2722A, graphical programming environments LabVIEW and technology Remote Panels.

II. Base Parameters of Photovoltaic Module

Characterization of photovoltaic panel includes different measurements at the solar cell level. Although measurement accuracies, speeds, and parameters may differ in importance across different levels of the industry and across space and terrestrial use, there are a number of key parameters that are typically measured in any testing environment [1, 2]:

Open-circuit voltage (V_{OC}) – the cell voltage at which there is zero current.

Short-circuit current (I_{SC}) – the current flowing out of the cell when the load resistance is zero.

Maximum power output of the cell (P_{max}) – the voltage and current point where the cell is generating its maximum power. The P_{max} point on an I-V curve is often referred to as the maximum power point (MPP).

$$P_{max} = I_{max} V_{max} . \quad (1)$$

Voltage at P_{max} (V_{max}) – the cell's voltage level at P_{max}.

Current at P_{max} (I_{max}) – the cell's current level at P_{max}.

Fill factor (FF) – the ratio of the maximum power point, P_{max}, divided by the V_{OC} and the I_{SC}:

$$FF = \frac{P_{max}}{V_{OC} . I_{SC}} . \quad (2)$$

Conversion efficiency of the device (η) is the percentage of power converted (from absorbed light to electrical energy)

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

[2]Georgi Nikolov is with the Faculty of Electronic Engineering and Technologies at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: gnikolov@tu-sofia.bg.

[3]Marin Marinov is with the Faculty of Electronic Engineering and Technologies at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: mbm@tu-sofia.bg.

and collected when a solar cell is connected to an electrical circuit. This value is calculated using the ratio of the maximum power point, P_{max}, divided by the input light irradiance (G, in W/m²) under standard test conditions (STC G_0 = 1000 W/m² with an air mass of 1.5 and cell temperature of 25°C) and the surface area of the solar cell (A_C in m²):

$$\eta = \frac{P_{max}}{G.A_C} = \frac{V_{OC}.I_{SC}.FF}{G.A_C}. \quad (3)$$

Generally, a photovoltaic panel, as an elementary PV electrical energy source, is described by equivalent electrical circuit - a five-parameter model. Parameters of the equivalent circuits of PV cell are cell diode properties (like ideality factor N and diode saturation current or dark current I_0), cell series resistance (R_S) and cell shunt resistance (R_{SH}).

Most of the parameters given in data sheets of the solar cell can be taken from the I-V characteristics. Figure 1 shows the I-V characteristics of a typical PV cell. Note that the amounts of current and voltage available from the cell depend upon the cell illumination level. In the ideal case, the I-V characteristic equation is

$$I = I_{PH} - I_0\left(e^{\frac{V}{V_T}} - 1\right), \quad (4)$$

where I_{PH} is photo-current and V_T is the diode thermal voltage.

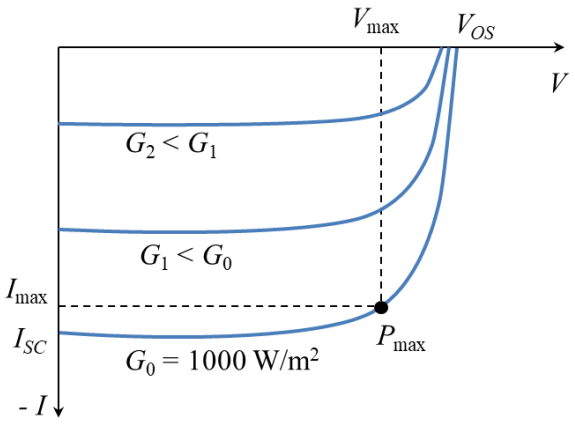


Fig. 1. I-V characteristics of a typical PV cell

Because the measured I-V characteristics of actual PV cells differ from this ideal version, including the parameters of the equivalent circuit, the equation (4) can be expressed as:

$$I = I_{PH} - I_0\left[exp\left(\frac{V + I.R_S}{N.V_T}\right) - 1\right] - \frac{V + I.R_S}{R_{SH}}. \quad (5)$$

As it can be seen from Fig. 1, parameters V_{OC}, I_{SC}, I_{max} and V_{max} can be defined directly from I-V characteristic, measured under STC. For example, to determine the short circuit current of a PV cell, simply set V = 0. Then parameters P_{max}, FF and η can be easily calculated using the equations from (1) to (3). The parameters of the equivalent circuits of PV cell can be estimated from forwarded and reversed biased I-V characteristics at different solar irradiances as described in reference [3].

One of the primary problems with PV cell characterization is the measurement of solar irradiances G. The simplest approach is to allow for a very good approximation, that the cell current is directly proportional to the cell irradiance. Thus, if the photo current under STC I_{PH0} is known, then the cell current at any other irradiance, G, is given by

$$I_{PH} = \frac{I_{PH0}}{G_0}G = k_G G. \quad (6)$$

Regarding temperature, when all parameters are constant, the higher the temperature the lower the voltage. This is considered a power loss. The drop in V_{OC} with temperature is mainly related to the increase in the leakage current of the PV cell I_0. The magnitude of voltage reduction varies inversely with V_{OC}. This means that cells with higher V_{OC} are less affected by the temperature than cells with lower V_{OC}.

The temperature coefficient is defined as the rate of change of a parameter with respect to the change in temperature. It can be current, voltage, or power temperature coefficient. For example, the temperature coefficient of voltage is the rate of change of the voltage with temperature change. A typical datasheet of a commercial PV module specifies temperature coefficients for the power, V_{OC} and I_{SC} under STC. Therefore, the values of these parameters under current cell temperature T_{Cell} can be calculated as:

$$\begin{aligned} P_{max} &= P_{STC} + TCP_{max}(T_{Cell} - 25°C), \\ V_{OC} &= V_{STC} + TCV_{OC}(T_{Cell} - 25°C), \\ I_{SC} &= I_{STC} + TCI_{SC}(T_{Cell} - 25°C), \end{aligned} \quad (7)$$

where P_{STC}, V_{STC} and I_{STC} are the values of the parameters under STC, and TCP_{max}, TCV_{OC} and TCI_{SC} are the corresponding temperature coefficients.

III. Remote Platform Design

A. Hardware Design

The block diagram of the remote virtual platform for photovoltaic panel characterization is shown in Fig. 2. The hardware part of the system consists of SMU U2722A and Test Fixture U2941A for I-V measurement of solar cell, PV reference device for irradiance measurement and temperature sensor for measuring current cell temperature T_{Cell}.

Modular SMU U2722A from Keysight Technologies has three channels and can operate in four-quadrants of I-V characteristic. Each SMU can force voltage or current and simultaneously measure voltage or current, as shown on the equivalent circuit in the Fig. 3. The figure shows the three basic modes of operation of every SMU – voltage source, current source and common. In addition, SMUs have the ability to specify a compliance setting. The setting is current compliance when the SMU is in voltage source mode and

voltage compliance when it is in current source mode. When the SMU reaches compliance, it acts appropriately as a constant current source or voltage source.

The modular device U2722A can operate entirely with the Keysight U2941A parametric test fixture which is designed for the end connection between SMU and electronic components and semiconductors, including solar cells.

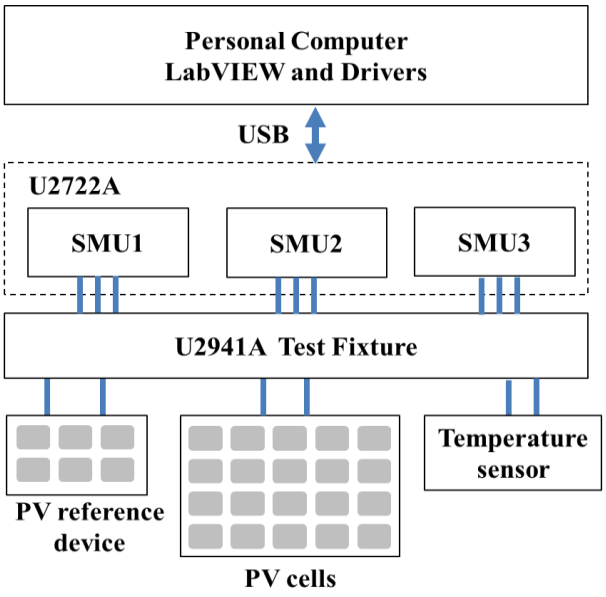


Fig. 2. The block diagram of the remote virtual platform for photovoltaic panel characterization

For PV reference device in this platform is used low-power, low cost photovoltaic panel SilverCrest SLS 2200 calibrated for irradiance as described in reference [3]. As can be seen in Fig.2. the PV reference device is connected to SMU1. The SMU1 is programmed in zero voltage source and current measure mode. Thus, a short-circuit current of reference photovoltaic is measured and irradiance is obtained according to the calibration equation. Since the SMU has capabilities of current/voltage source and measure, the temperature can easily be measured with voltage output temperature sensor (like integrated temperature sensor AD22100 by Analog Devices), resistance temperature detector (Pt100, Pt1000) or thermistor. In the present project the temperature is measured with Pt100 by SMU3, programmed in current source/voltage measuring mode of operation. The base electrical parameters of SMU U2722A are summarized in Table 1 [4].

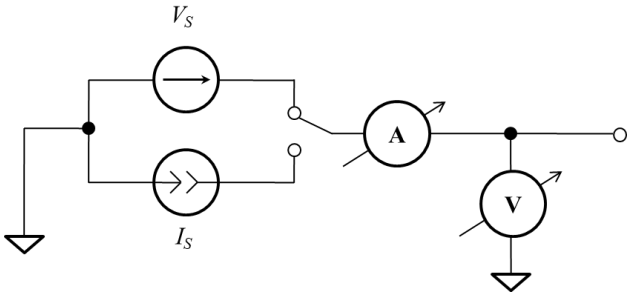


Fig. 3. The simplified equivalent circuit of SMU

TABLE I
THE BASE ELECTRICAL PARAMETERS OF SMU U2722A

Parameter	Range	Accuracy	Resolution
Voltage programming or measure	±2 V	0.075% + 1.5 mV	0.1 mV
	±20 V	0.05% + 10 mV	1 mV
Current programming or measure	±1 µA	0.085% + 0.85 nA	100 pA
	±10 µA	0.085% + 8.5 nA	1 nA
	±100 µA	0.075% + 75 nA	10 nA
	±1 mA	0.075% + 750 nA	100 nA
	±10 mA	0.075% + 7.5 µA	1 µA
	±120 mA	0.1% + 100 µA	20 µA

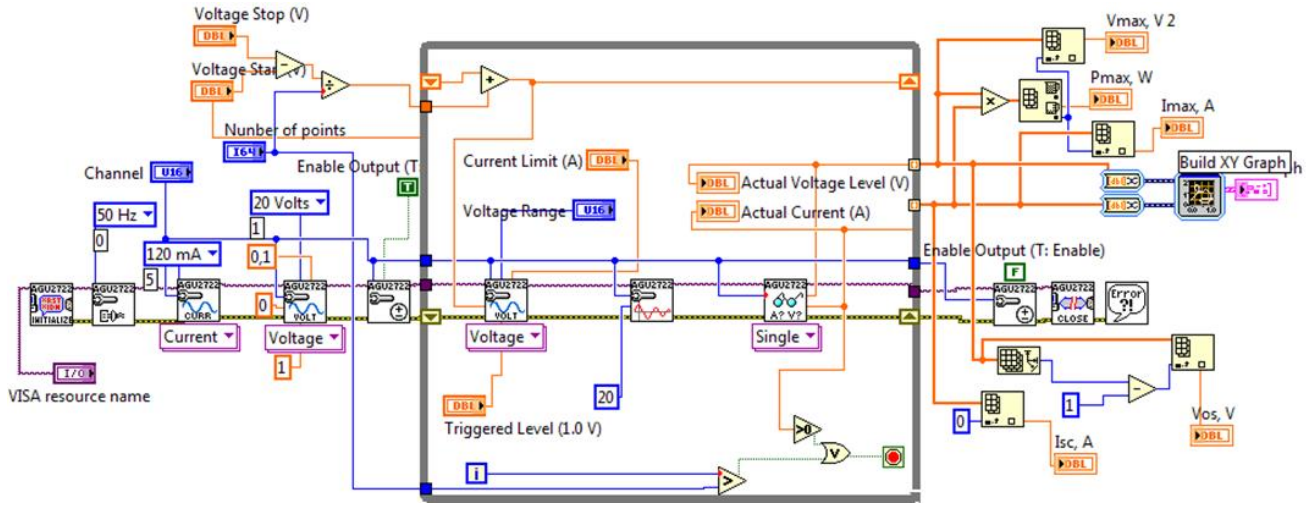


Fig. 4. The LabVIEW block diagram of developed remote virtual platform

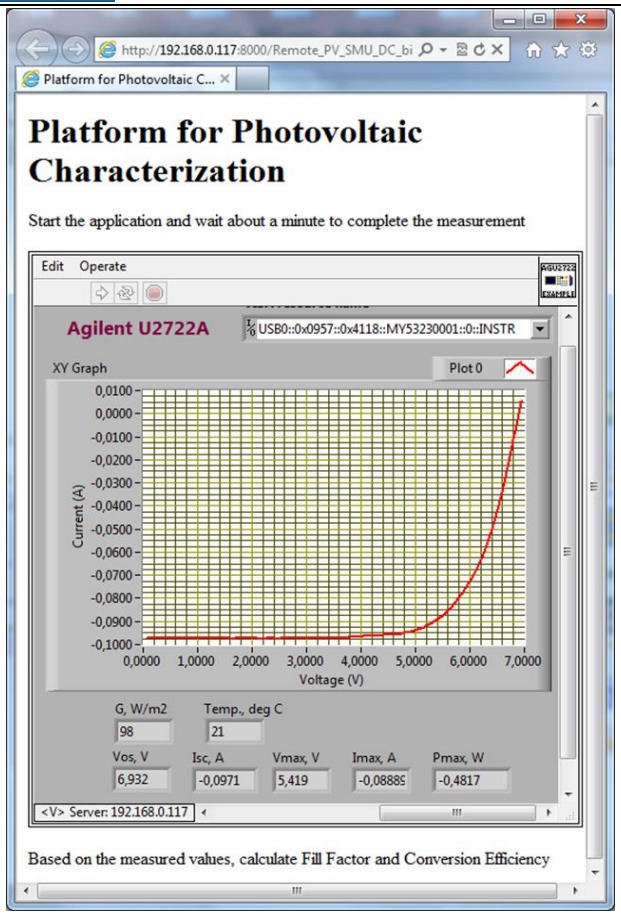


Fig. 5. The front panel of developed remote virtual platform for photovoltaic panel characterization

B. Software Design

The graphical block diagram of the developed virtual platform for photovoltaic characterisation is shown in the Fig. 4. This graphical code consists of LabVIEW Plug and Play instrument driver for U2722A and conventional programming functions and components. On the left are the functions for configuring source-measure units, mode of operation, sweep type, start value, stop value and number of points. The sweep of voltage is accomplished in *While loop*. The loop has two stop conditions - a count reaches a programmed number of steps or the measured current becomes positive. On the right of the block diagram are the functions for manipulation of arrays by which the necessary parameters of the photovoltaic are obtained. It should be noted that the remote user is not given the opportunity to change the operating mode of the measurement experiment, as can be seen from Fig. 5

IV. Distance Measurement of Base Photovoltaic Parameters

After the hardware and software parts of the system are completed, it is necessary to provide remote access to the system. With the Remote Panels technology, it is possible to automatically publish the same user interface to a web browser. The process is simple, because LabVIEW comes with its own web server. The developer only selects to publish the desired application and any remote user will see and operate with the front panel as if they were in front of the local computer. The front panel of the developed remote platform for photovoltaic characterisation is shown in Fig. 5. In the figure the measured DC characteristic of low power photovoltaic and determination of base parameters can be seen. The learner can determine the other parameters using the equations (2) to (7).

V. Conclusion

In this paper, the design and implementation of a remote platform for photovoltaic panel characterization has been presented. In the project for measuring the electrical properties of solar cells are used modern solutions based on SMU and virtual techniques. Different parameters are extracted automatically using measured I-V curve by developed virtual system and they played an important role for evaluating the PV modules. The technology that provides the implementation and administration of a remote access to the platform is LabVIEW Remote Panels.

Knowing in details the electrical characteristics of a solar cell is critical for determining the device's output performance and efficiency. Therefore, the accomplished remote platform can be used to establish a LabVIEW based distance learning remote laboratory.

Acknowledgement

This paper has been produced within the framework of the ERASMUS + project Geothermal & Solar Skills Vocational Education and Training (GSS-VET).

References

[1] An. Luque and St. Hegedus, *Handbook of Photovoltaic Science and Engineering*, John Wiley & Sons Ltd, England, 2013.

[2] Keysight Technologies, "IV and CV Characterizations of Solar/Photovoltaic Cells Using the B1500A", Application Note, 2014.

[3] G. Nikolov, B. Nikolova, E. Gieva and I. Ruskova, " Extraction of Solar Cell Parameters using Source Measure Unit", 40th International Spring Seminar on Electronics Technology (ISSE), pp. 1-6, Sofia, Bulgaria, 2017.

[4] Agilent Technologies, "Agilent U2722A/U2723A USB Modular Source Measure Units", User's Guide, 2013.

[5] R. Tamrakar1, A. Gupta, "A Review: extraction of solar cell modelling parameters", International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering, Vol. 3, Issue 1, ISSN 2321-5526, January 2015.

[6] P. Bauer, R. Ionel, "LabVIEW Remote Panels and Web Services in Solar Energy Experiment – A Comparative Evaluation", IEEE SACI, 23-25 May 2013.

A. Rohit, A. Tomar, A. Kumar, S. Rangnekar, "Virtual lab based real-time data acquisition, measurement and monitoring platform for solar photovoltaic module", Resource-Efficient Technologies, pp.446-451, Elsevier, 2017.

Design and Realization of a Synchronous Buck Converter with eGaN FETs

Zoran B. Zivanovic[1] and Vladimir J. Smiljakovic[2]

Abstract – **This paper presents the straightforward design of a synchronous buck converter using the Gallium Nitride transistors. The operating principles are briefly explained, including the functional block diagram. The prototype has been built and tested through lab measurements to verify the design. The goal was to demonstrate the capabilities of GaN transistors in order to show their significant advantages over silicon MOSFETs.**

Keywords – **Buck, efficiency, eGaN, Figure of Merit, synchronous.**

I. Introduction

Rapid commercialization of switch mode power supplies was made possible by mass production of silicon power MOSFETs in the late seventies. Constant demand for more efficient power converters, with higher efficiency and smaller size, forced the power supply engineers to reduce losses and increase the switching frequency. After 30 years, silicon MOSFET development has approached its theoretical limits. In contrast, the production of the enhancement mode gallium nitride FETs started in 2009, and is less investigated in literature [1]-[5]. A major factor in eGaN FETs performance is die (Fig. 1) that minimizes parasitic elements and allows footprint that helps to make optimal layout design. Interleaved drain and source bars generate small loops with opposing current flow, resulting into magnetic field cancellation and smaller common source inductance.

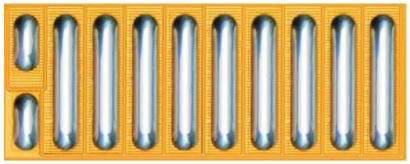


Fig. 1. eGaN EPC2015C passivated die (size 4.6 x 1.6 mm)

The objective of this paper is to contribute to the analysis of eGaN FETs characteristics by measuring the efficiency and capturing the waveforms of the synchronous buck converter with eGaN FETs over a wide range of operating conditions.

The paper is organized as follows. In Section II, synchronous buck converter basics are given. Two state of the art MOSFETs are compared with two eGaN FETS with the same voltage rating in section III with calculations of critical parameters and components using design equations. Experimental results of a synchronous buck converter with eGaN FETs are presented in section IV to validate design. At the end, the conclusions are given in section V.

[1]Zoran Zivanovic is with the IMTEL KOMUNIKACIJE AD, Bul. Mihajla Pupina 165b, 11070 Belgrade, Serbia, E-mail: zoki@insimtel.com.

[2]Vladimir Smiljakovic is with the IMTEL KOMUNIKACIJE AD, Bul. Mihajla Pupina 165b, 11070 Belgrade, Serbia, E-mail: smiljac@insimtel.com.

II. Synchronous Buck Converter Basics

The synchronous buck converter is given in Fig. 2.

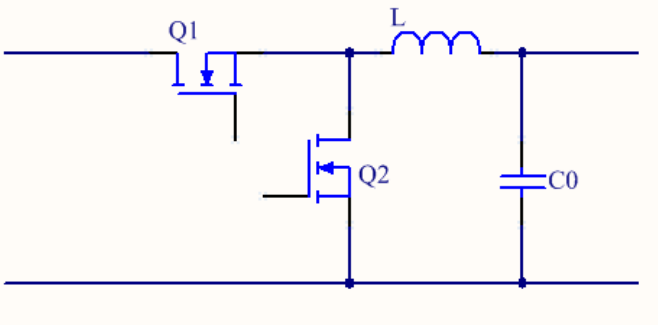


Fig. 2. Synchronous buck converter

When the switch Q_1 is ON during T_{ON} the energy is transferred to the output as well as to the inductor L. The current flows through the switch Q_1 and the inductor. The difference of voltages between V_{IN} and V_{OUT} is applied to the inductor and the inductor and switch current I_{D1} rises linearly. During the OFF period of time, when the switch Q_1 is OFF, the inductor current continues to supply load current. The switch Q_2 is ON providing freewheeling path for the inductor current. The output voltage is applied to the inductor, so the inductor current decreases linearly (Fig. 3).

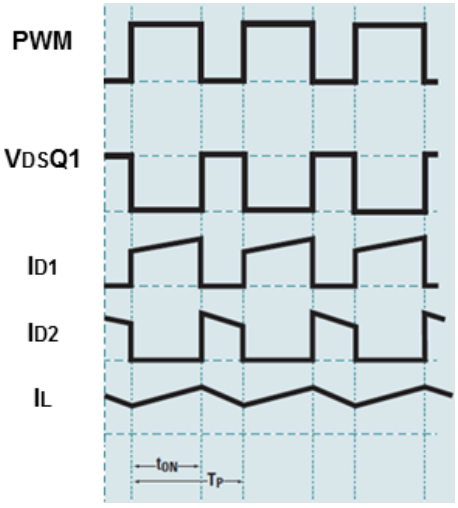


Fig. 3. Synchronous buck converter waveforms

The inductor current flows to the capacitor and the load. The capacitor C_O acts as a reservoir and holds the output voltage nearly constant.

III. Design and Analysis

To design synchronous buck converter we will start from the design specifications given in Table I.

TABLE I
DESIGN SPECIFICATIONS

		Min	Typ	Max	
Input voltage	V_{IN}	8	12	18	V
Output voltage	V_O		5		V
Output current	I_O	0.1	7		A
Output current limit	I_{OCL}		8.4		A
Full load efficiency	η		> 90		%
Switching frequency	f_{SW}		440		KHz

In order to dampen sub-harmonic oscillation, the minimum inductor value should be calculated as

$$L = \frac{V_O}{0.3 f_{SW} I_O} \tag{1}$$

The maximum and minimum duty cycle can be calculated as

$$D_{MAX} = \frac{V_O}{V_{INMIN}} \tag{2}$$

$$D_{MIN} = \frac{V_O}{V_{INMAX}} \tag{3}$$

The peak to peak inductor current is

$$\Delta I = \frac{V_{INM} - V_O}{L} \frac{D_{MIN}}{f_{SW}} \tag{4}$$

Knowing that the peak inductor current is

$$I_{pk} = I_O + \frac{\Delta I}{2} \tag{5}$$

The output capacitor must be large enough to limit the voltage overshoots and undershoots during the load change transients. The minimum value of output capacitor is

$$C_{O\min} = \frac{L I_{step}^2}{2 \Delta V_O D_{MAX} (V_{INM} - V_O)} \tag{6}$$

The power switches must be chosen with voltage rating to withstand the maximum input voltage and also they must be capable of delivering the load current. Knowing that, our choice is two MOSFETs (IPC50N04S5L-5R5 and IPZ40N04S5L-7R4) and two eGaNs (EPC2014C and EPC2015C). We will use figure of merit (FOM) expressed by the equation

$$FOM = (Q_{GD} + Q_{GS}) R_{DS} \tag{7}$$

as a selection criteria. Results are given in Table II. Although both MOSFETs are state of the art power switches, the eGaN FETs are better choice because of superior switching characteristics and zero reverse recovery time.

TABLE II
POWER SWITCH CHARACTERISTICS

	Q_{GD}	Q_{GS}	R_{DS}	FOM	Q_{rr}
IPC50N04S5L	3.6nC	2.7nC	5.7mΩ	35.91	22ns
IPZ40N04S5L	3.0nC	2.0nC	7.9mΩ	39.5	26ns
EPC2014C	0.3nC	0.7nC	12mΩ	12	0ns
EPC2015C	1.2nC	2.7nC	3.2mΩ	12.48	0ns

The high side FET losses can be calculated as

$$P_{DQ1} = I_O^2 R_{DS} D_{MAX} + \frac{1}{2} V_{IN} (t_r + t_f) I_O f_{SW} \tag{8}$$

The losses in the low side FET can be expressed as

$$P_{DQ2} = I_O^2 R_{DS} (1 - D_{MIN}) + I_O (t_{dr} + t_{df}) f_{SW} V_D + Q_{rr} f_{SW} V_{IN} \tag{9}$$

Calculated values are given in Table III.

TABLE III
CALCULATED VALUES

Inductance calculated	L	5.4	µH
Inductance adopted	L	6.8	µH
Max duty cycle	D_{MAX}	0.625	
Min duty cycle	D_{MIN}	0.277	
Peak to peak current	ΔI	1.2	A
Peak current	I_{pk}	7.6	A
Min capacitance	C_{Omin}	290	µF

We will also adjust passive components around the error amplifier in order to have 25 kHz loop bandwidth with 60° phase margin.

IV. Realization

The synchronous buck converter has been realized on a four layer FR-4 substrate, with a thickness of 1.6mm and 50µm copper on outer layers and 35 µm on inner layers. The size of

the PCB is 50x42 mm. For the high-side and low-side switches we have used eGaN FETs EPC2015C. The eGaN FETs have a maximum gate voltage of +6V. The buck controller LM5141, require a logic-level FETs, so it is a great choice for this design. The output filter is made of four pcs of multilayer ceramic capacitors 47µF, 10V, X7R, to reduce high frequency noise and two 100 µF, 10V tantalum capacitors for energy storage.

Using lab power supply and electronic load, we have recorded the waveforms of interest and measured efficiency at various loads and input voltages.

Drain voltage of the high side GaN FET is given in Fig. 4. His gate voltage is given in Fig. 5. Turn-on waveform of high-side GaN FET with parasitic 125 MHz ringing is given in Fig. 6. Drain voltage of the low-side side GaN FET is given in Fig. 7 and his gate voltage is given in Fig. 8. The 125 MHz ringing on the drain of the low-side GaN FET at turn-off is given in Fig. 9. Measured efficiency of the converter is given in Fig. 10.

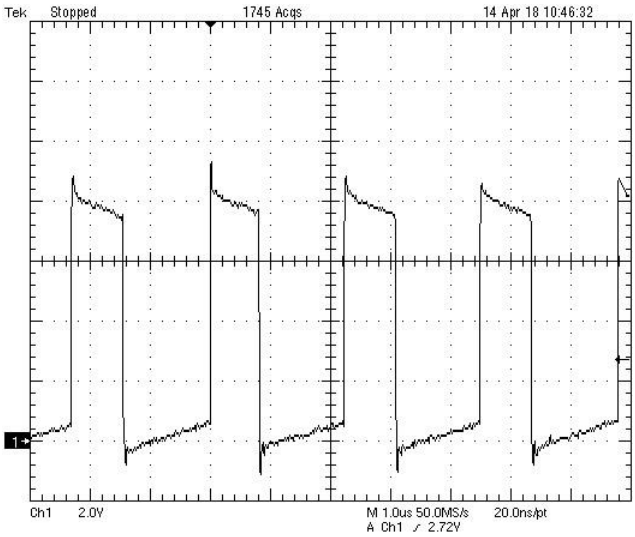


Fig. 4. Drain voltage of high-side GaN FET

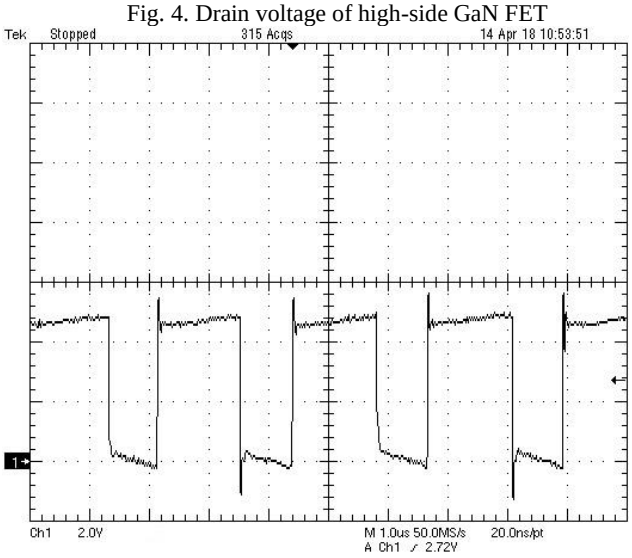


Fig. 5. Gate voltage of high-side GaN FET

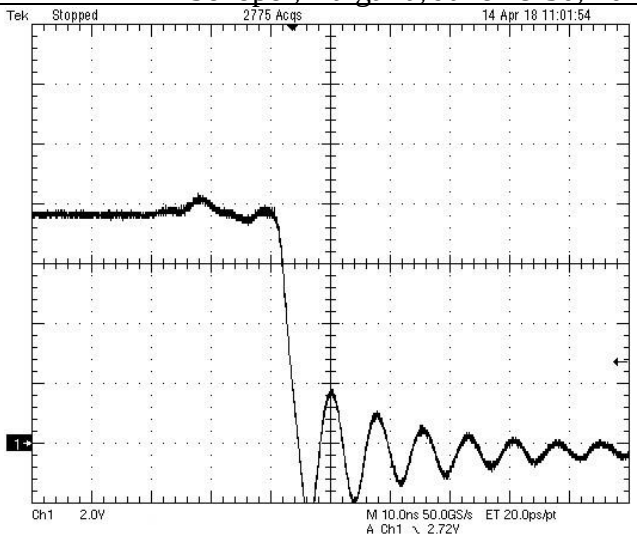


Fig. 6. Turn-on waveform of high-side GaN FET

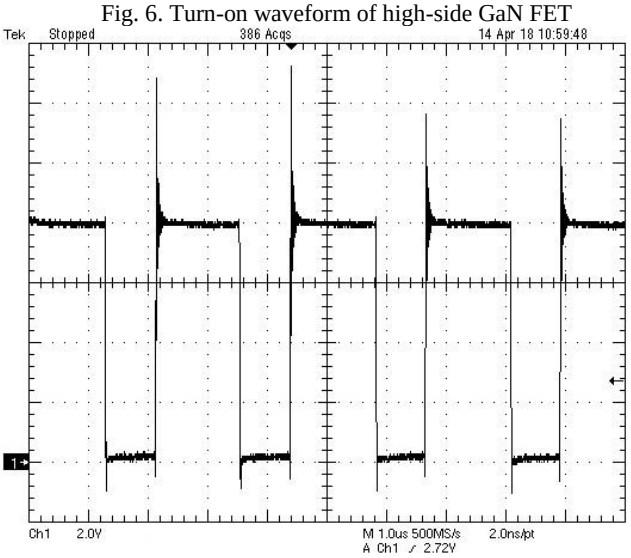


Fig. 7 Drain voltage of low-side GaN FET

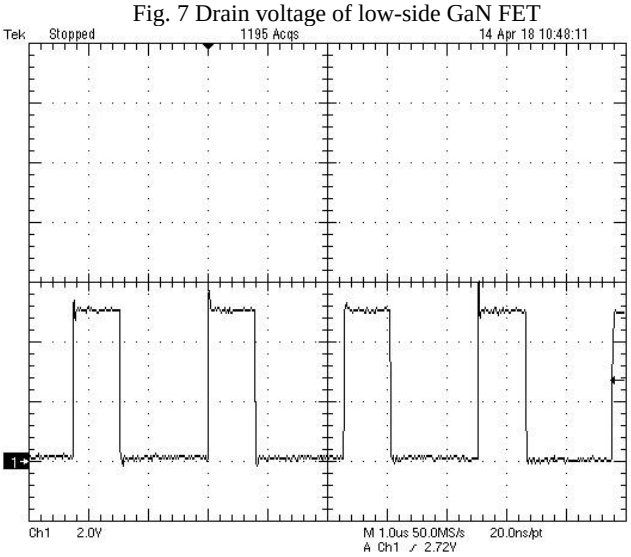


Fig. 8. Gate voltage of low-side GaN FET

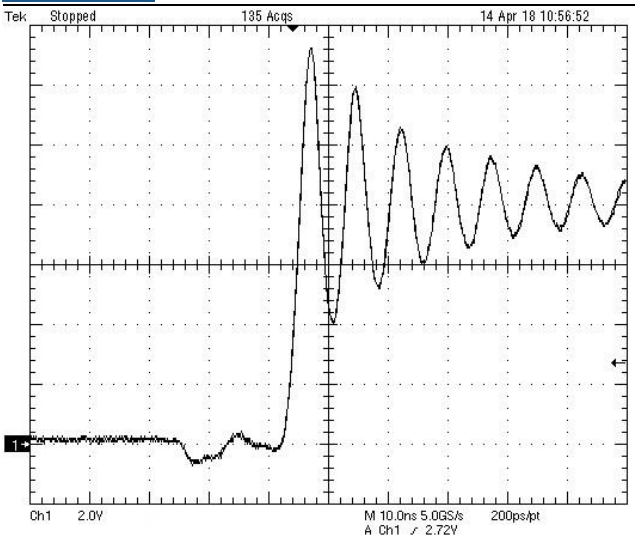


Fig. 9. The ringing on the drain of low-side GaN FET at turn-off

Fig. 12. Experimental setup in the lab

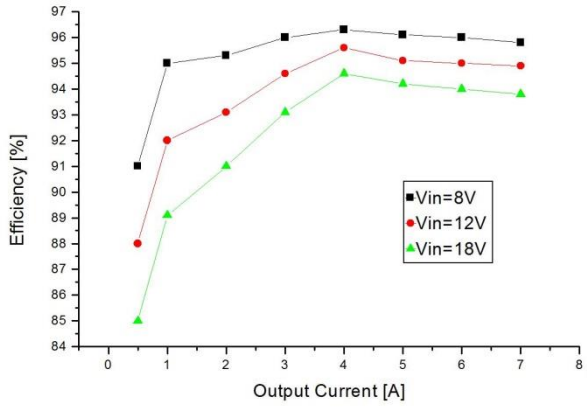


Fig. 10. Measured efficiency vs. output current

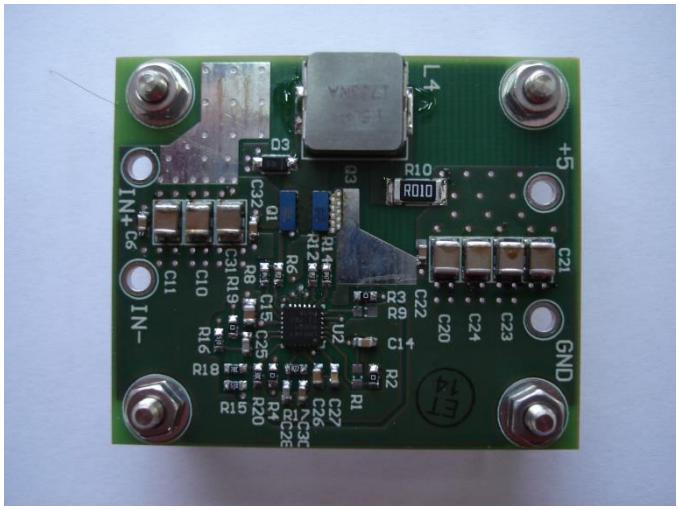


Fig. 11. The converter prototype

V. Conclusion

In this paper the design and realization of a synchronous buck converter with GaN FETs is presented. Calculations and experimental results are presented. The prototype was built and tested. The results verified that the efficiency can go as high as 96%.

Acknowledgement

The work is partially supported by the Serbian Ministry of Education and Science (Project III-45016). The authors would like to thank Texas Instruments for providing buck controller samples.

References

[1] Zoran Zivanovic, Vladimir Smiljakovic, "Design and Realization of a eGaN FET DCDC Converter", TELFOR2014 Proceedings, pp. 637-640

[2] Zoran Zivanovic, Vladimir Smiljakovic, Sinisa Jovanovic "Analysis and Design of a Resonant Reset GaN Forward Converter with Self-Driven Synchronous Rectifiers", ICEST2015 Proceedings, pp. 275-278

[3] Zoran Zivanovic, Vladimir Smiljakovic, "Design and Realization of a Interleaved Boost Converter with GaN FETs and SiC Diodes", IcETRAN2017, pp. ELI2.3.1-4

[4] Zhi-Wei Xu, Zhi-Liang Zhang, Ke Xu, Zhou Dong and Xiaoyong Ren, "2-MHz GaN PWM Isolated SEPIC Converters", 2017 IEEE Applied Power Electronics Conference and Exposition (APEC), pp. 149-156

[5] Johan Delaine, Pierre-Olivier Jeannin, David Frey, Kevin Guepratte, "High Frequency DC-DC Converter Using GaN Device", 2012 Twenty-Seventh Annual IEEE Applied Power Electronics Conference and Exposition (APEC), pp. 1757-1761

Behavioral Model of Copper Indium Gallium Selenide Solar Module

Rostislav Rusev

Abstract – **A behavioural model of Copper Indium Gallium Selenide photovoltaic modules is elaborated applying semi-empirical approach. Model equations are coded in Matlab. Model simulation results are verified against experimental data for four module data sheets from two producers. The result shows that the model very well predicts the current, voltage and power at high irradiance and it lessens the predicted values of voltage and generic power for low irradiance values. The model can be successfully used for fast estimation of generated power in real conditions – ambient temperature and irradiance.**

Keywords – **Copper indium gallium selenide (CIGS) photovoltaic modules, behavioural model**

I. Introduction

Copper indium gallium selenide (CIGS) solar cells are extensively introduced to photovoltaic market. Their advantages include relatively high efficiency (12-15%) low fabrication costs (e.g. small number of materials used and small number of low temperature processes). The solar cells efficiency is determined by their composition. The I-III-VI CIGS is an alloy of copper indium diselenide (CIS) and copper gallium diselenide (CGS). This means that the semiconductor band gap varies from 1.0 eV for CIS to about 1.65 eV for CGS. The alloy has high absorption coefficient of solar irradiation, hence a thinner film is required (1-2 μm) compared to other semiconductor materials. CIGS is one of the three main thin film photovoltaic technologies, and the other two are cadmium telluride (CdTe) and amorphous silicon. It is possible to fabricate thin films of CIGS to make them flexible and ready for deposition onto flexible substrates. In contrast to amorphous silicon technology, that is high temperature, the CIGS technology is low temperature and it is constantly improved. This will lead to higher production efficiency and lower price for the panels. For this reason, many manufacturers adopted this technology. In addition, many customers prefer this type of solar panels for mounting on roofs and building facades.

Shortcomings of CIGS cells include their lower efficiency (12 – 15%) compared with Si-based cells. With technology improvement, the efficiency is expected to increase up to 23% [1]. This is approximately 1/3 of the theoretic maximum of 28-30% efficiency for CIGS solar cells. At module/panel level the CIGS efficiency is ~ 17% [2], which can be circumvented by better design of cells and modules.

Theoretical analysis of CIGS cells relies on modified models of CdTe solar cells [3] because the structure of CIGS and CdTe cells is similar. One diode and two diode equivalent circuits could be used for modeling the dependence of *I-V* on temperature and illumination [4].

Since the one diode or two diode models do not describe well the efficiency of CIGS modules, behavioral model for quick estimation of efficiency and energy performance of a given module is developed. In practice, we should quickly and easily estimate the behavior of commercial CIGS solar module in real conditions (ambient temperature and irradiance).

Rostislav Rusev is with the Faculty of Telecommunications, Technical University of Sofia, Kl. Ohridski 8, 1797 Sofia, Bulgaria, E-mail: rusev@ecad.tu-sofia.bg.

II. Model and Equations

A simplified behavioral model for quick estimation of CIGS solar module in real conditions is presented. The model consists of semi-empirical equations coded in Matlab [5]. Block-diagrams of solar module are shown in Figure 1.

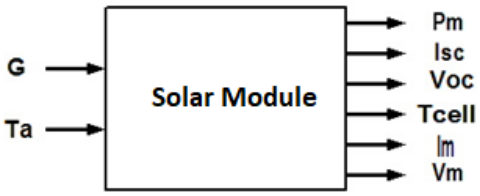


Fig. 1. Solar module block-diagram.

The model data input is solar irradiation (*G*) and ambient temperature (*Ta*). We are also using data from manufacturer references such as open circuit voltage (*Vocr*), short circuit current (*Iscr*), under standard testing conditions (STC), etc.

The objective is to determine the solar module parameters for short circuit current (*Isc*), open circuit voltage (*Voc*), power (*Pm*) at maximum power point (MPP), cell temperature (*Tcell*), current at MPP (*Im*), voltage at MPP (*Vm*). The short circuit current equation is derived from the standard diode-model taking into account the radiation effect and temperature effect:

$$Isc = (Jscr*A*G/1000) + (DIFJscT)*(Ta - Tr) \quad (1)$$

where *Jscr* – current density, *A* – area of the module, *DIFJscT* – temperature coefficient of short circuit current, *Tr* – reference temperature.

$$Tcell = ((NOCT - 20)*G/800) + Ta \quad (2)$$

where *NOCT* – nominal operating conditions temperature.

Different equations can be used for *Voc*. Most of them show that *Voc* is proportional to the thermal potential *Vt*, but

TABLE I
MEASURED DATA (PROVIDED BY THE MANUFACTURER) VS. SIMULATED DATA OF SOLIBRO MODULES WITH 115, 125 AND 135 WP .

PERFORMANCE AT STANDARD TEST CONDITIONS (1000 W/m2, 25˚C, AM 1.5 G SPECTRUM)							
SOLIBRO		Data	model	Data	model	Data	model
Minimum Power *Pm*	[W]	115	115,02	125	124,95	135	134,96
Short Circuit Current *Isc*	[A]	1,75	1,75	1,78	1,78	1,81	1,81
Open Circuit Voltage *Voc*	[V]	97,3	97,3	100,2	100,2	102,6	102,8
Current at MPP *Im*	[A]	1,49	1,47	1,56	1,57	1,63	1,61
Voltage at MPP *Vm*	[V]	77,2	77,2	80,1	80,1	82,8	82,8
PERFORMANCE AT NOMINAL MODULE OPERATING TEMPERATURE (800 W/m2, NMOT, AM 1.5 G SPECTRUM)							
Minimum Power *Pm*	[W]	85,6	85,1	92,8	92,7	100,6	100,4
Short Circuit Current *Isc*	[A]	1,4	1,42	1,43	1,44	1,45	1,44
Open Circuit Voltage *Voc*	[V]	91,3	91,8	94,1	94,7	96,5	97,1
Current at MPP *Im*	[A]	1,19	1,2	1,24	1,27	1,3	1,29
Voltage at MPP *Vm*	[V]	71,9	71,3	74,8	74,3	77,4	77,3

non-ideality factor (*n*) need to be accounted for as well as the voltage temperature coefficient [V/°C] (*DIFVocT*):

$$Voc \sim n*(DIFVocT)*Vt \quad (3)$$

The output results do not fulfil the standard deviation requirements. Hence, the next semi-empirical equation is derived:

$$Vocc = \log((Isc)/(Iscr)) \quad (4)$$

$$Voc = Vocr + n*(DIFVocT)*Vocc*(Vt*(Ta + 273)) \quad (5)$$

The *vo* parameter below is the normalized value of the open circuit voltage to the thermal potential *Vt*. The value of the fill factor for the ideal solar cell without resistive effects *FFo* is

$$FFo = (vo - \log(vo + 0.72))/(1 + vo) \quad (6)$$

The series resistance *Rs* can be calculated from the fill factor:

$$Rss = (Vocr/Iscr) \quad (7)$$

$$rs\_znam = Iscr.\hat{}2 \quad (8)$$

$$Rs = Rss - (Pmr/(FFo*rs\_znam)) \quad (9)$$

In reality, the series resistance is also affected by irradiation and temperature. The current at MPP is calculated in the same manner as *Isc*:

$$Im = Imr*(G/Gr) + DIFJscT*(Ta - Tr) \quad (10)$$

For the voltage at MPP is used:

$$Vmro = \log(Im/(Imr)) \quad (11)$$

$$Vm = Vmr + Vmro.*(Vt*(Tcell + 273)) \quad (12)$$

From here for fill factor – *FF*: is obtained the following

$$Fm = Vm.*Im \quad (13)$$

$$Fsoc = Voc.*Isc \quad (14)$$

$$FF = Fm/Fsoc \quad (15)$$

Finally the maximum output power *Pm* is

$$Pm = Vm.*Im \quad (16)$$

III. RESULTS AND DISCUSSION

The model with data from SOLIBRO – SL2 CIGS THIN-FILM MODULE, Generation 2.1 | 115-135 Wp [6] is tested first. The model is verified for three panels with power at MPP for 115, 125 and 135 Wp respectively; these panels operate at standard test conditions (1000 W/m^2, 25˚C, AM 1.5 G SPECTRUM) and at nominal module operating temperature (800 W/m^2, NMOT, AM 1.5 G SPECTRUM). The area of the three panels is equal to 0.94 m^2. The results are in Table 1.

The results from Table 1 clearly show that the model well describes the behavior at STC and NOCT. The accuracy for current, voltage, and power is under 1%. Unfortunately, the manufacturer has given insufficient information for module behavior at different temperatures and irradiation – only *IV*-characteristic for irradiation of 1000, 500 and 200 W/m^2 and 25 and 50 °C. We tested the model at these conditions and obtained good description for the current at different conditions but poor description for voltages at low irradiation.

In Table 2 the variation of voltage at different temperatures and irradiations is shown.

TABLE II
VOC VS. IRRADIATION OF SOLIBRO 135 WP MODULE.

SOLIBRO 135Wp	Tcell= 25			Tcell = 50		
G [W/m^2]	data [V]	model [V]	error %	data [V]	model [V]	error %
1000	102,6	102,8	0,19	96	102	6,25
500	99	88	-11,11	91	86	-5,49
200	93	67	-27,96	85	64	-24,71
average			-12,96			-7,98

In Table 2 is seen that the model underestimates the voltage at low irradiation but the average error at different temperatures is between 8–13%., which is close to the tolerance given by the manufacturer (10%) [6]. As expected, the power is also underestimated.

For better verification of the model, other data sheets are used – SOLAR FRONTIER [7] (SF170-S 170 W Module with area of 1.23 m^2). Experimentally obtained characteristics for the current, voltage, and power at different irradiation are shown in Fig. 2. In Fig. 3 we show the simulated dependence of Pm on irradiation.

In Table 3 are psoted the results of manufacturer provided data, the simulated model data and the accuracy of simulation. Similarly to the previous simulations, the model well describes the current, voltage and power for high values of irradiation but the accuracy for the voltage and power worsens for low values of irradiation.

The model well describes the values for the current at 1000 W/m^2 irradiation – the error is 0%. A small increase of the error at low irradiation values is observed – 2.22%. The averaged error is under 1%, which is much less than the 10% tolerance given by the manufacturer.

In the simulation results for the voltage vs. irradiation, it is observed the same dependence as in the case of SOLIBRO 135Wp module – the error increases with decreasing the irradiation. At 400 W/m^2 the irradiation reaches maximum deviation for the voltages – 5%. Nevertheless, we still remain in the range of the manufacturer tolerance of 10%.

It is reasonable to expect that the accuracy of the power vs. experimental data will be maximal. At irradiation of 400 W/m^2 the accuracy is 15%. The average accuracy for the power is 7.23%, which is in the range of manufacture provided tolerance of 10% [7]. The same model was applied for simulation of polycrystalline PV modules and the calculated accuracy was in the same range [8].

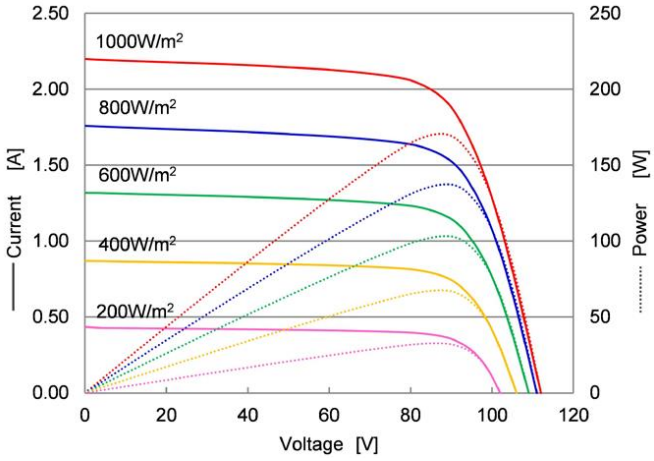


Fig. 2. Experimental data (manufacturer data) *I-V* and *Pm-V* characteristics dependent on irradiation *G*.

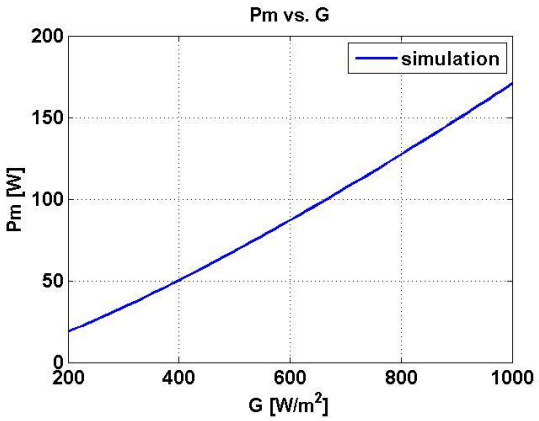


Fig. 3. Modeled relation of power (*Pm*) vs irradiation (*G*).

TABLE III
MEASURED DATA (PROVIDED BY THE MANUFACTURER) VS. SIMULATED DATA OF SOLAR FRONTIER SF170-S 170 W MODULE.

SOLAR FRONTIER SF170-S 170 W Module									
G [V/m2]	Isc data [A]	Isc model [A]	error %	Voc data [V]	Voc model [V]	error %	Pm data [W]	Pm model [W]	error %
1000	2,2	2,2	0,00	112	112	0,00	170	170,7	0,41
800	1,76	1,75	-0,57	110	110,5	0,45	126	125,5	-0,40
600	1,29	1,3	0,78	107	108	0,93	101	92	-8,91
400	0,92	0,9	-2,17	105,5	100,5	-4,74	61	52	-14,75
200	0,45	0,44	-2,22	102	99,5	-2,45	24	21	-12,50
average			-0,84			-1,16			-7,23

IV. CONCLUSION

The developed simplified behavioral model for quick estimation of CIGS module operation proves its applicability in real conditions. The test show that the model very well predicts the current, voltage and power at the big values of solar irradiance and decrease the values of voltage and power as decreasing of irradiance but the average error from model is in a good similarity with the tolerance of the experimental data producers.

ACKNOWLEDGEMENT

The research described in this paper was carried out within the framework of Contract No. BG05M20P001-1.001-0008-C01/28.02.2018.

REFERENCES

[1] "Solar Frontier Achieves World Record Thin-Film Solar Cell Efficiency of 22.9%," 2017. [Online]. Available: http://www.solar-frontier.com/eng/news/2017/1220_press.html

[2] M. A. Green, K. Emery, Y. Hishikawa, W. Warta, and E. D. Dunlop, "Solar cell efficiency tables (version 47)," Prog. Photovoltaics Res. Appl., vol. 24, no. 1, pp. 3–11, Jan. 2016.

[3] Kosyachenko L.A., Mathew X., Roshko V.Ya., Grushko E.V. "Optical absorptivity and recombination losses: The limitations imposed by the thickness of absorber layer in CdS/CdTe solar cells", Solar Energy Materials and Solar Cells 114: 179-185, 2013.

[4] Sun, X., Silverman, T.J., Garris, R.L., Deline, C., & Alam, M.A. "An Illumination- and Temperature-Dependent Analytical Model for Copper Indium Gallium Diselenide (CIGS) Solar Cells", IEEE Journal of Photovoltaics, Vol. 6, pp. 1298-1307, 2016.

[5] Matlab website: http://www.mathworks.com

[6] Solibro website: http://solibro-solar.com/fileadmin/image/05_News_Downloads/Downloads/Data_sheets/G2.1/Solibro_datasheet_SL2_G2-1_2017-04_Rev03_EN.pdf

[7] Solar Frontier website: http://www.solar-frontier.com/eng/solutions/products/pdf/datesheat_170.pdf

[8] Angelov, G., Rusev, R., Andonova, A., Hristov, M., "Behavioral modeling of polycrystalline module ET-P660230WW", Proceedings of the International Spring Seminar on Electronics Technology, art. no. 6648246, pp. 221-226, 2013.

Mathematical model and Calculation of operating characteristics of Three-Phase Asynchronous Motor with Double Squirrel Cage

Blagoja Arapinoski[1], Mirka P. Radevska[1], Metodija Atanasovski[1] and Mile Spirovski[1]

Abstract – **The paper is presented part of a complete and complex research over the three phase asynchronous motor with double squirrel cage rotor. The section that follows shows the determination of the operating characteristics of much easier way, using specialized software package for the design and analysis of electrical machines. Given equations that receive operating characteristics, and on the module of the program that is used, Rmxprt - Ansoft Maxwell 14. Simulations were made under different loads and the results are presented in tables and graphics.**

Keywords – **Operating characteristics, three phase asynchronous motor with double squirrel cage rotor.**

I. INTRODUCTION

The software package Ansoft - Maxwell 14 is specialized software designed for design and simulation analysis of electrical machines.[1]-[5] The concept of the program is so designed to enable continuous communication between the user and the computer at all stages of design, and it also enables the input of own creations and scientific knowledge by the user-designer. The program itself basically consists of three modules, interconnected:

The first module can design any electrical machine, calculate the parameters of the equivalent scheme that represents it, and determine all the performance characteristics in different modes of operation. In the database of this module, a large number of empirical equations from classical theory and part of the modern theory of electric machines are entered.

Designing and receiving a large number of results for the machine parameters, as well as the performance characteristics of different modes of operation in this module is very simple and fast.

As research subjects from which a portion is presented in the labor below, is selected three-phase asynchronous motor with double squirrel cage and with the following nominal data:

$$P_n = 3.5kW, U_n = 240V,\ f = 50Hz$$

$2p = 4, \cos\varphi = 0.85,\ \eta = 84\%,$ and Δ winding connection.

II. MATHEMATICAL MODEL AND EQUIVALENT REPLACEMENT SCHEME

The subject of the research is three phase asynchronous motor with double squirrel cage. As compared to the construction of the stator coil, the coil is identical to the standard three-phase asynchronous motors. The difference in this type of motor is in the rotor circuit, which are set two interconnected cages made of material with different conductance. The upper cage is known as start cage (used to run the electric motor when sliding is equal to 1), and is made of phosphor bronze alloy that has a lower conductivity compared to the bottom cage. The bottom cage is made of copper and has the role of the working cage when the speed of the rotor has a large enough value that scrolling is close to nominal. Currents redistribution from the top in the bottom cage is completely automatic and is dependent on the rotor speed and load.

The theory of asynchronous motors with double cage can be traced to the theory of three-phase three-winding transformers. That means double cages asynchronous motors can be considered as three separate electrical circuits that are magnetically coupled. The circuit indicated by I, represents the stator and circuit II, and III, representing the upper and lower cage rotor respectively. Each of these circuits respond appropriately active resistance R_I, R_{II}, R_{III} and inductance, respectively corresponding total inductive winding resistance X_I, X_{II}, X_{III} .

In asynchronous motors with double cage rotor, when the load changes, and changes in engine speed and thus the frequency of the current in the rotor-conductors, which causes a change of inductance and resistance in the rotor circuit. This phenomenon can be expressed mathematically in a way that the active component of the resistance of the rotor circuit is divided by sliding s, or simply if the rotor circuit of the upper and bottom cage add extra value to the active resistance $R\frac{1-s}{s}$. In that case we can write the following expressions:

$$\begin{aligned} \bar{U}_2' &= \bar{I}_2' R_2' \frac{1-s}{s} \\ \bar{U}_3' &= \bar{I}_3' R_3' \frac{1-s}{s} \end{aligned} \quad (1)$$

[1]Blagoja Arapinoski,Mirka Popnikolova Radevska, Metodija Atanasovski and Mile Spirovski are with the Faculty of Technical sciences at University of Bitola, st. Makedonska Falanga 33, Bitola 7000, Macedonia.

Voltage equations for this type of motor received form given by:

$$\begin{aligned}\bar{U}_1 &= \bar{I}_1(R_1 + jX_{\sigma 1}) - \bar{I}_2'\left(R_2' + jX_{\sigma 2}'\right) - j\bar{I}_3' X_\sigma' - \bar{I}_2' R_2' \frac{1-s}{s} \\ \bar{U}_1 &= \bar{I}_1(R_1 + jX_{\sigma 1}) - \bar{I}_3'\left(R_3' + jX_{\sigma 3}'\right) - j\bar{I}_2' X_\sigma' - \bar{I}_3' R_3' \frac{1-s}{s}\end{aligned} \quad (2)$$

Based on the previous expressions can be compiled equivalent electric scheme of asynchronous motor with double cage and is given on Figure 1.

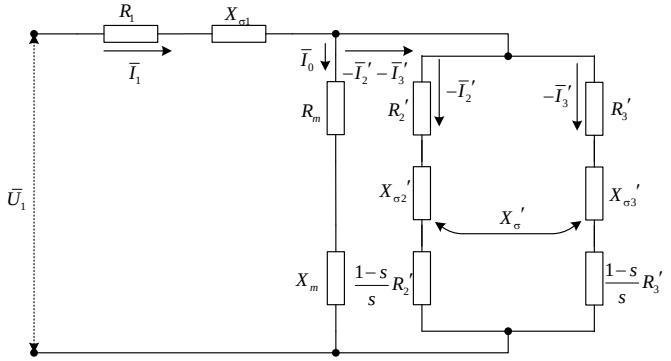


Fig. 1. Asynchronous motor with double cage - equivalent electric scheme.

III. WORKING CHARACTERISTICS OF DOUBLE – CAGE THREE PHASE ASYNCHRONOUS MOTORS

The performance characteristics of the two-cages asynchronous motor are dependent on the output power P_2, the stator current I_1, the power factor $\cos\varphi$, the efficiency coefficient η, the slip s, the moment M, the speed of rotation of the rotor n, and the currents in the upper I_g and lower rotor cages I_d, and function of the consumed electric power P_1. The performance characteristics of the asynchronous two-phase motor can be obtained in several ways: Numeric with a computer program, which basically has an algorithm in which the equations connected with the aforementioned dependencies are added, Experimental obtaining of the working characteristics by direct method by reading the sizes and by an indirect method using a circular diagram. The software used for reconstruction of the calculation of the two-phase asynchronous motor provides the possibility of obtaining the operating characteristics in the numerical way mentioned above. Namely, the Ansoft-Maxwell 14 Rmxprt module calculates and displays the performance characteristics of different motor operation modes, using a model programmed based on the equivalent dual cage-like asynchronous motor pattern.

Equations with which the working characteristics are obtained:

The current I_2'' that flows into the rotor circuit of the motor reduced to a stator is given by the expression:

$$\bar{I}_2' = \bar{I}_2'' \cdot C_1 \quad (3)$$

The power factor of this current is given by:

$$\cos\varphi_2 = \frac{\left(R_1'' + \frac{R_{2e}}{s}\right)}{\bar{Z}_e}; \quad \sin\varphi_2 = \frac{\left(X_{\sigma 1}'' + X_\sigma'' + X_{\sigma 2e}\right)}{\bar{Z}_e} \quad (4)$$

For the active and reactive current in the rotor circuit, the following expressions are valid:

$$\begin{aligned} I_{2a}'' &= I_2 \cos\varphi_2 \\ I_{2r}'' &= I_2 \sin\varphi_2 \end{aligned} \quad (5)$$

The impedance in the branch of the magnetization is:

$$\bar{Z}_0 = \sqrt{(R_1 + R_m)^2 + (X_{\sigma 1} + X_m)^2} \quad (6)$$

The current of an ideal Idling, active and reactive component is calculated by:

$$\bar{I}_{0i} = \frac{\bar{U}_n}{\bar{Z}_o}; \quad I_{0a} = I_{0i}\cos\varphi_0; \quad I_{0r} = I_{0i}\sin\varphi_0 \quad (7)$$

Where the power factor is given by:

$$\cos\varphi_0 = \frac{R_1 + R_m}{Z_0}; \quad \sin\varphi_0 = \frac{X_{\sigma 1} + X_m}{Z_0} \quad (8)$$

The current that the motor takes from the network, i.e., the current flowing through the three-phase distributed stator winding is the vector sum of the currents from the ideal idling and the current flowing into the rotor circuit of the motor:

$$\bar{I}_1 = \sqrt{\left(I_{0a} + I_{2a}''\right)^2 + \left(I_{0r} + I_{2r}''\right)^2} \quad (9)$$

The actual value of the current flowing through the rotor circuit of the machine is:

$$\bar{I}_2 = \bar{I}_2' k_i \quad (10)$$

For the currents in the upper and lower cage of the rotor, the expressions are:

$$\bar{I}_g = \frac{\bar{I}_2}{\bar{Z}_g\left(1 + \frac{\bar{Z}_d}{\bar{Z}_g}\right)}; \qquad \bar{I}_d = \frac{\bar{I}_2}{\left(1 + \frac{\bar{Z}_d}{\bar{Z}_g}\right)} \quad (11)$$

The power factor of the motor is given by:

$$\cos\varphi_1 = \frac{I_{1a}}{I_1}; \quad \bar{I}_{1a} = \bar{I}_{0a} + \bar{I}_{2a}'' \quad (12)$$

For the energy balance, that is, the power that the motor takes from the network and the losses that occur in it, the

following expressions apply:

$$P_1 = m_1 U_n I_{1a}; \quad P_0 = m_1 (R_1 + R_m) I_0{}^2$$
$$P_{cu1} = m_1 R_1 I_1{}^2; \quad P_{cu2} = m_1 R_{2e} I_2{}''^2 \qquad (13)$$
$$\sum P_g = P_0 + P_{cu1} + P_{cu2} + P_{men}$$

The useful power of the motor axle is obtained when the total power losses in the motor are deducted from the input power:

$$P_2 = P_1 - \sum P_g \qquad (14)$$

The coefficient of efficiency , the speed at which the rotor turns and the electromagnetic torque are calculated with the following expressions:

$$\eta = \frac{P_2}{P_1}; \quad n = (1-s) n_1 \qquad (15)$$

$$M_{em} = \frac{p m_1 U_n{}^2 \frac{R_{2e}}{s}}{2\pi f_1 \left[\left(R_1'' + \frac{R_{2e}}{s} \right)^2 + \left(X_{\sigma 1}'' + X_\sigma'' + X_{\sigma 2e} \right)^2 \right]} \qquad (16)$$

IV. Obtained results from the simulation

The performance display in Ansoft Maxwell 14's Rmxprt can be tabular and graphical. Simulations were made to obtain the performance characteristics at different loads, i.e. 20 different operating modes were analyzed with the step of changing the sliding $\Delta s = 0.2\%$, and the results of the simulations are given in Table 1.

Table 1 - Results of the calculations for obtaining the performance characteristics

s	P_1[W]	P_2[W]	I_1[A]	cosφ	η	n[min^{-1}]	M[Nm]	I_a[A]	I_g[A]	I_2[A]
0.001	416.3	32.8	2.12	0.27	0.0789	1498	0.846	4.95	1.85	6.75
0.003	681.3	295.2	2.26	0.42	0.4332	1495	2.523	14.65	5.52	20.17
0.005	943.9	552.5	2.45	0.53	0.585	1492	4.175	24.33	9.177	33.59
0.007	1203.9	804.6	2.69	0.62	0.668	1489	5.799	33.92	12.81	46.72
0.009	1460.7	1051.1	2.96	0.68	0.719	1486	7.395	43.42	16.39	59.81
0.011	1714.1	1291.6	3.25	0.73	0.753	1483	8.958	52.81	19.95	72.77
0.013	1963.5	1526.1	3.55	0.77	0.777	1480	10.48	62.19	23.48	85.56
0.015	2208.8	1754.1	3.87	0.79	0.794	1477	11.98	71.27	26.98	93.25
0.017	2449.7	1975.5	4.19	0.81	0.806	1474	13.44	80.31	30.44	110.76
0.019	2685.4	2190.2	4.51	0.83	0.815	1471	14.86	89.23	33.86	123.09
0.021	2917.1	2398.1	4.84	0.84	0.822	1468	16.25	98.81	37.25	135.23
0.023	3143.1	2598.8	5.16	0.845	0.826	1465	17.58	106.64	40.59	147.27
0.025	3363.7	2792.6	5.49	0.851	0.83	1462	18.88	115.13	43.91	159.03
0.027	3578.8	2979.3	5.81	0.856	0.832	1459	20.15	123.46	47.17	176.63
0.029	3788.4	3158.8	6.12	0.859	0.833	1456	21.36	131.62	50.41	182.84
0.031	3992.2	3331.3	6.44	0.861	0.834	1453	22.54	139.66	53.58	193.25
0.033	4190.3	3496.7	6.75	0.862	0.8345	1450	23.67	147.52	56.78	204.26
0.035	4382.4	3635.1	7.06	0.8626	0.835	1447	24.77	155.23	59.88	215.06
0.037	4568.8	3806.1	7.36	0.863	0.833	1444	25.82	162.76	62.91	223.65

The results of the calculated currents in the stator and the rotor rings are significant because they can serve as input data for the load in the three-dimensional electromagnetic analysis with finite element method.

For good visibility, the performance characteristics are calculated and displayed graphically on one diagram in the relative units in Figure 2:

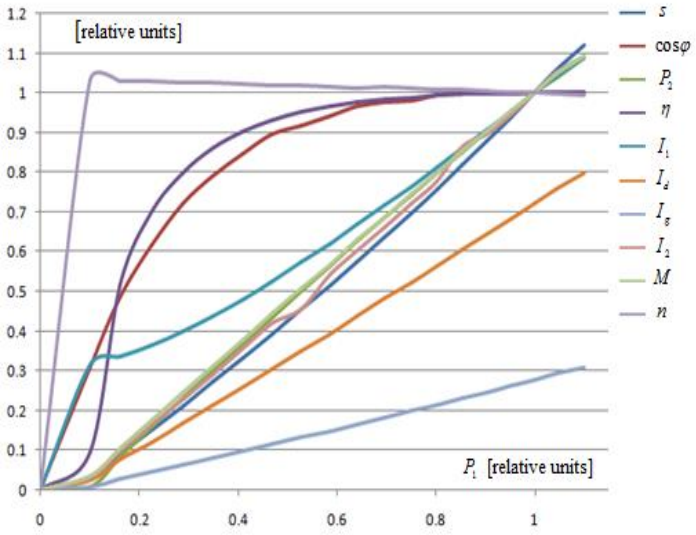


Fig. 2. Operating characteristics of an asynchronous motor with a double cage

V. Conclusion

Designing with the (Rmxprt) Ansoft Maxwell 14 module is simple and at the same time gives excellent results when calculating the parameters of the equivalent electrical scheme and obtaining the operating characteristics for different operating modes. By comparing the calculated parameters with the help of analytical methods and numerical calculated parameters of the asynchronous motor with a double cage with the software used, the reliability of this kind of design is confirmed. Comparing the working characteristics obtained with the software and those of the experimental testing and in this respect, it confirms the accuracy and the possibilities.

References

[1] 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.

[2] M. Popnikolova Radevska: "Calculation of Electromechanical Characteristics on Overband Magnetic Separator with Finite Elements", ICEST 2006, p.p. 367-370, Sofia, Bulgaria 2006.

[3] B. Arapinoski, M. Popnikolova Radevska, "*Electromagnetic and thermal analysis of power distribution transformer with FEM*" ICEST 2010, Ohrid, R.Macedonia 2010.

[4] Mirka Popnikolova Radevska, Blagoja Arapinoski, *Computation of solid salient poles synchronous motor electromagnetic characteristic,*10th international conference of applied electromagnetic ПЕС 2011, Nis, Serbia, 2011.

[5] 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.

Integrated DC-DC Converter for IoT Applications

Tihomir Brusev[1], Georgi Kunov[2] and Elissaveta Gadjeva[3]

Abstract –**A low-power monolithic DC-DC converter for Internet of things (IoT) applications is presented in this paper. The maximum efficiency η of the circuit is 79.76%, when the input voltage is 3.6V while output voltage is regulated to be equal to 1.8V. The DC-DC converter is designed with Cadence on CMOS technology. The efficiency η of the circuit is investigated as a function of average output current. The reaction of the system is evaluated when the input voltage is changed in the range between 2.5V and 4.2V and the temperature varies from -50°C to 120°C.**

Keywords – **Cadence, CMOS technology, DC-DC converters, Integrated circuits, Internet of things (IoT).**

I. INTRODUCTION

The new fifth generation (5G) mobile network communication technology allows different type of sensors, electrical cars, electronic, automation and industrial devices to exchange data. Internet of thing (IoT) enables large number of embedded systems to be interconnected through the web. Using smartphones, tablets or computers the users can monitor and control different type of electronic devices in their offices and homes. The industry predicts that billions of components and systems will transmit data and receive instruction through the IoT [1].

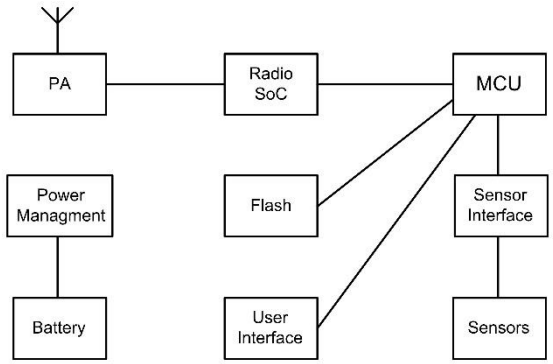


Fig. 1. Block circuit diagram of IoT wireless sensors network.

Power energy in industrial and home systems could be saved if IoT network technology is applied [2]. Appropriate operation time for electronic devices can be set in such way, that they could work after the peak energy demand. For this purpose different sensing and control techniques have to be chosen. Many devices connected in 5G network, like wireless motion and light sensors are powered through energy harvesting or battery power sources [3]. Some of them have to work without replacing the battery for years. The battery lifetime could be increased if the energy is used more efficiently.

The block circuit diagram of IoT wireless sensor network is shown in Fig. 1 [4]. The microcontrollers (MCUs) used for sensor monitoring and control have to work with low power consumption in order to save battery energy. Design of high efficiency DC-DC converters is the key for minimizing of power losses.

The power management system of wireless sensor devices connected in IoT is discussed in Section II*A* of this paper. The basic circuit's topologies of linear regulators and synchronous switching-mode DC-DC converters are considered in Section II*B* and Section II*C* respectively. Low-power monolithic BUCK DC-DC converter system designed on CMOS technology for IoT application is presented in Section III.

II. THEORY

A. Battery powered wireless sensor networks

Wireless sensors connected in IoT networks have to be low energy consuming devices in order to be increased battery lifetime. They have to operate in most power efficient mode. Two different power management strategies for wireless sensors networks are shown in Fig. 2 [3]. Integrated BUCK DC-DC converter can help for decreasing of overall power losses [4].

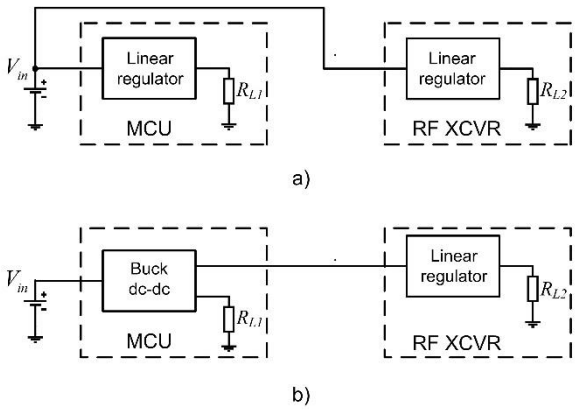


Fig. 2. Block circuit's diagrams for battery powered wireless sensors connected in IoT networks.

As it can be seen from the picture shown in Fig. 2a, linear regulators convert battery output voltage and ensure the desired input voltages of microcontroller (MCU) and high

[1]Tihomir Brusev is with the Faculty of Telecommunications, Technical University of Sofia, Kl. Ohridski 8, 1797 Sofia, Bulgaria, E-mail: brusev@ecad.tu-sofia.bg.

[2]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.

[3]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.

frequency transceiver (RF XCVR). In the second power management strategy presented in Fig. 2b, monolithic BUCK DC-DC converter is used to deliver the input voltage MCU. Thus the first linear regulator from the block diagram shown in Fig. 2a is replaced with more power efficient circuit. The switching-mode converter also ensures the input voltage of second linear regulator. Therefore overall power losses of the wireless sensors, connected in IoT networks, are minimized.

B. Linear regulators

Linear regulators are simple electronic circuits which convert higher input to smaller output dc voltage [5]. These devices use small number of electronic components. The schematic of a linear regulator is shown in Fig. 3. Their advantage in integrated circuits applications is that they occupy small silicon area.

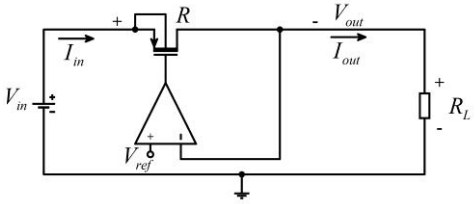


Fig. 3. Schematic of linear regulator.

Usually linear regulators use at least one transistor to control the desire dc level of output voltage. Their theoretical maximum possible efficiency is equal to:

$$\eta_{\max} = \frac{V_{out}}{V_{in}}, \tag{1}$$

where V_{out} is the dc output voltage, while V_{in} is the dc input voltage of linear regulator. In practice this maximum value cannot be achieved, because there are power losses in the series connected switch and in the parasitic impedances of the feedback circuit. A lot of waste heat is generated in linear regulators, which is a disadvantage of these circuits. This energy must be dissipated using radiators. Also when the difference between the input and the output voltage is large, linear regulators indicate low efficiency results.

C. Integrated BUCK DC-DC converters

The real synchronous BUCK DC-DC converters are circuit which could indicate high efficiency results η of about 90 %. Compared to the basic regulator's architectures the voltage drop of the diode in their structure is eliminated. Therefore they are suitable choice for low-power integrated circuit application. The efficiency η of the BUCK converter is equal to:

$$\eta = \frac{P_{out}}{P_{in}}, \tag{2}$$

where P_{out} is the average output power and P_{in} is the average input power of the DC-DC converter. The schematic of synchronous BUCK DC-DC converter is shown in Fig. 4. The energy losses are equal to the difference between the input and output average powers. They are distributed mainly in the output transistors and in the filters of the circuit.

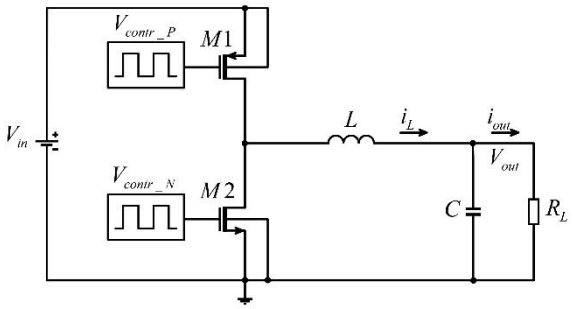


Fig. 4. Synchronous BUCK DC-DC converter.

Power losses in the transistors *M1* and *M2* dominate and have a major impact over the BUCK DC-DC converter's efficiency. They depend strongly on electrical circuit's parameters and are composed respectively of switching and conduction power losses [6]. The total power losses in MOS transistor can be further expressed by [7]:

$$P_{tot,MOS} = a\sqrt{\left(I^2 + \frac{\Delta i_L^2}{3}\right) f_s}, \tag{3}$$

where Δi_L is the inductor current ripple, f_s is the switching frequency, 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 output MOS transistors and power supply voltage. The BUCK switching-mode DC-DC converters are more efficient circuits compared to the linear regulators. [8].

III. LOW-POWER BUCK DC-DC CONVERTER DESIGNED ON CMOS TECHNOLOGY

The low-power integrated BUCK DC-DC converter is designed on CMOS technology with Cadence [9]. The block circuit's diagram is shown in Fig. 5.

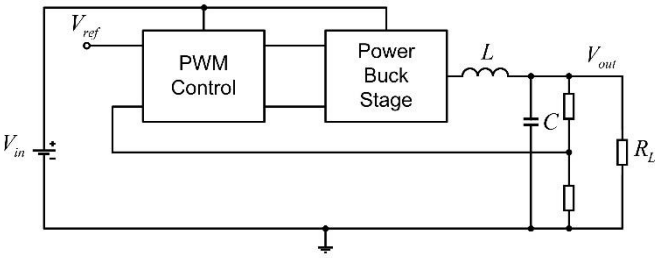


Fig. 5. Block circuit's diagram of PWM controlled BUCK DC-DC converter.

PWM control is used for regulation of the output voltage of the system. The input voltage of the BUCK converter is V_{in}=3.6V_{dc} and the average output voltage is regulated to be equal to $V_{out\text{-}avg}$=1.8V [4]. The operating switching frequency

f_s of the system is 77MHz, while the corner frequency f_c of the output low-pass LC filter is selected to be 4.5MHz. High values of f_s and f_c are chosen in order to decrease the sizes of output inductor and capacitor of the circuit. Thus low-pass filter can be integrated on the chip and the occupied silicon area is minimized. The received results for the efficiency η as a function of average load current $I_{out\text{-}avg}$ of the designed BUCK converter are presented in Table I. The maximum efficiency η of the circuit equal to 79.76% is achieved when the load current $I_{out\text{-}avg}$ is 50mA.

TABLE I
EFFICIENCY OF BUCK CONVERTER AS A FUNCTION OF $I_{OUT\text{-}AVG}$

R_L [Ω]	BUCK DC-DC Converter	
	$I_{out\text{-}avg}$ [mA]	Efficiency [%]
10	180	63.3
15	120	72.8
25	72	78.75
36	50	79.76
50	36	78.7
100	18	70.72
180	10	59
250	7	51
360	5	42.5
900	2	23
1800	1	13

As it can be seen from the results presented in Table I, the efficiency of BUCK converter is decreased dramatically when the load current is smaller than 10mA. The efficiency of the BUCK DC-DC converter as a function of the average output current $I_{out\text{-}avg}$ is presented graphically in Fig. 6.

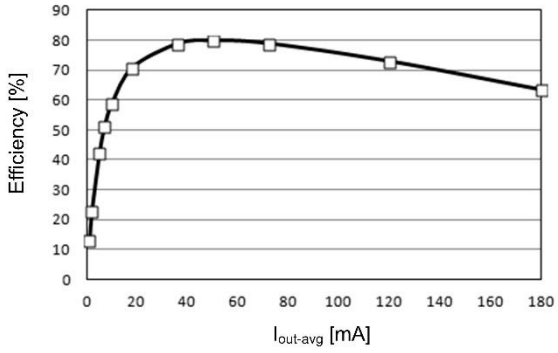


Fig. 6. Efficiency of the BUCK DC-DC converter as a function of the $I_{out\text{-}avg}$.

The detailed block diagram of PWM control, formed by error amplifier, ramp generator, comparator and buffer, plus power BUCK stage is presented in Fig. 7. The BUCK DC-DC converter systems which use PWM control have worse efficiency at light-load conditions. From energy point of view the designed circuit indicate better efficiency at high load currents. The BUCK DC-DC converter has to ensure the desired input voltages for microcontroller (MCU) and the high frequency transceiver (RF XCVR), when the operating conditions are changed.

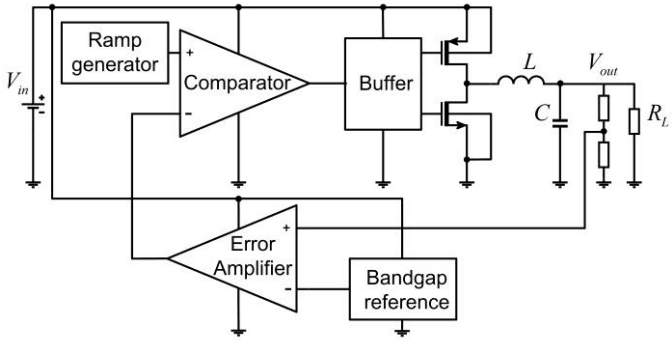


Fig. 7. PWM control circuit plus power BUCK DC-DC converter stage.

The waveform of DC-DC converter's output voltage V_{out} as a function of time is shown in Fig. 8. In this specific case the average value of the load current $I_{out\text{-}avg}$ is equal to 72mA.

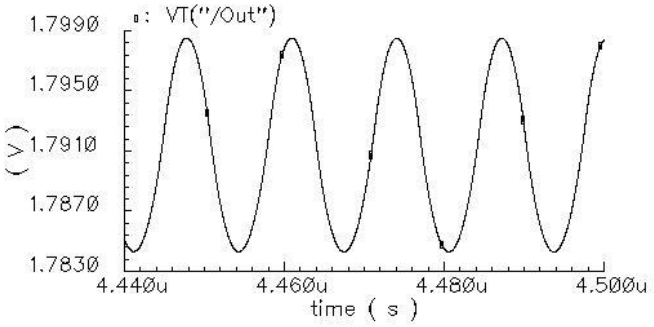


Fig. 8. The waveform of BUCK converter output voltage V_{out} as a function of time, when $I_{out\text{-}avg}$=72 mA.

The reaction of BUCK converter is estimated when input voltage varies from 2.5V to 4.2V. All the investigations are achieved at output current $I_{out\text{-}avg}$ equal to 50mA. The dependence of $V_{out\text{-}avg}$ as a function of V_{in} is illustrated in Fig. 9.

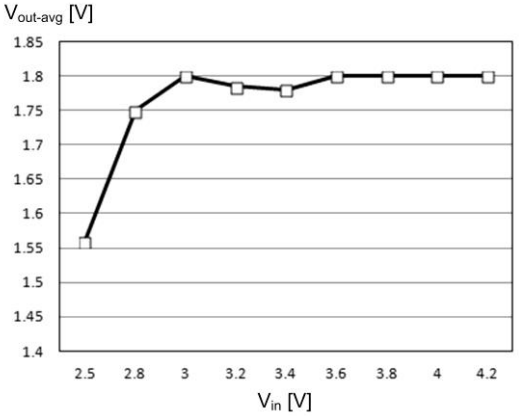


Fig. 9. Average output voltage $V_{out\text{-}avg}$ as a function of V_{in}.

As it can be seen from the picture, output voltage of BUCK converter $V_{out\text{-}avg}$ remains stable, when V_{in} varies between 2.8V and 4.2V. If the input voltage is changed in the range from 2.8V to 2.5V, $V_{out\text{-}avg}$ drops down with almost 0.2V.

Wireless sensors connected in IoT networks should display stable performance in wide temperature range. The schematic of the bandgap voltage reference is shown in Fig. 10. This circuit's topology provides a stable voltage for the whole converter system that is independent from the power supply, load current, and temperature variations [10].

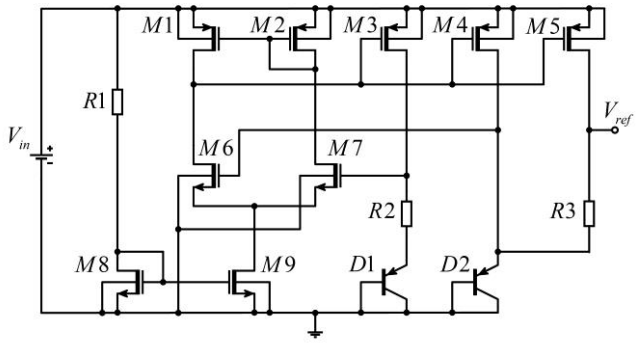


Fig. 10. Schematic of bandgap voltage reference.

The influence of the temperature over the circuit's parameters is investigated. The load current $I_{out\text{-}avg}$ is equal to 50mA, when BUCK DC-DC converter displays highest efficiency values. The received results of converter's average output voltage $V_{out\text{-}avg}$ and efficiency η at different temperatures are presented in Table II.

TABLE II

$V_{OUT\text{-}AVG}$ OF BUCK CONVERTER AS A FUNCTION OF THE TEMPERATURE

T [°C]	BUCK DC-DC Converter	
	$V_{out\text{-}avg}$ [V]	Efficiency [%]
-50	1.818	82.69
-25	1.817	81.59
0	1.813	79.93
25	1.8	79.76
40	1.796	78.72
60	1.787	77.73
80	1.775	80.05
100	1.764	76.13
120	1.76	76

The investigations show that when the temperature varies from -50°C to 120°C output voltage remain stable. As it can be seen from Table II $V_{out\text{-}avg}$ drops down from 1.82V to 1.76V with almost 0.06V.

The block circuit diagram illustrated in Fig. 2b is considered and the advantages of using the monolithic BUCK converter instead of linear regulator are estimated. The overall efficiency of the voltage regulators, shown in Fig. 2a and Fig. 2b respectively, are calculated and compared. In the calculation it is assumed that the linear regulators will work with maximum theoretical possible efficiency. The achieved investigation data, of the designed monolithic switching-mode circuit presented in this paper, are taken into account. The received results show that the integrated BUCK DC-DC converter increases the overall efficiency of voltage regulators presented in Fig. 2b with 35%, compared to the block circuit's diagram shown in Fig. 2a.

IV. CONCLUSION

Integrated BUCK DC-DC converter designed with Cadence on CMOS technology for IoT applications is presented in this paper. The circuit indicates better efficiency results at high load currents. The average output voltage $V_{out\text{-}avg}$ remains stable, when the input voltage V_{in} varies between 2.8V and 4.2V. The results show that when the temperature varies from -50°C to 120°C, the output voltage of the BUCK DC-DC converter is changed in the range between 1.82V and 1.76V. The maximum efficiency η of the circuit is 79.76%, when V_{in} is equal to 3.6V and $V_{out\text{-}avg}$ is regulated to be 1.8V. The advantages of using the integrated BUCK converter instead of linear regulator, in battery powered wireless sensors connected in IoT networks, are evaluated and estimated.

ACKNOWLEDGEMENT

The research described in this paper was carried out within the framework of BG05M20P001-1.001-0008-C01/28.02.2018.

REFERENCES

[1] M. Alhawari, D. Kilani, B. Mohammad, H. Saleh and M. Ismail, "An efficient thermal energy harvesting and power management for μWatt wearable BioChips", 2016 IEEE International Symposium on Circuits and Systems (ISCAS), Montreal, QC, pp. 2258-2261, 2016.

[2] www.silabs.com, *Overcoming Challenges of Connecting Intelligent Nodes to the Internet of Things*, 2012, Silicon Laboratories Inc.

[3] R. L. B. Zamparette, H. D. Klimach and S. Bampi, "A 90% efficiency 60 mW MPPT switched capacitor DC — DC converter for photovoltaic energy harvesting aiming for IoT applications", 2017 IEEE 8th Latin American Symposium on Circuits & Systems (LASCAS), pp. 1-4, Bariloche, 2017.

[4] www.keysight.com, *The big five IoT challenges*, issue 3, 2017.

[5] P. Hazucha, S. T. Moon, G. Schrom, F. Paillet, D. Gardner, S. Rajapandian, T. Karnik, "High Voltage Tolerant Linear Regulator With Fast Digital Control for Biasing of Integrated DC-DC Converters", IEEE Journal of Solid-State Circuits, vol.42, no. 1, pp.66-73, Jan. 2007.

[6] 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, April 2014, pp. 235-245.

[7] V. Kurson, *Supply and Threshold Voltage Scalind Techniques in CMOS Circuits*, University of Rochester, NY, 2004.

[8] M. Hassan, L. Larson, V. Leung, and P. Asbeck, "A Combined Series-Parallel Hybrid Envelope Amplifier for Envelope Tracking Mobile Terminal RF Power Amplifier Applications", IEEE Journal of Solid-State Circuits, vol. 47, no.5, May 2012, pp. 1185-1198.

[9] www.cadence.com

[10] Y. Wu and V. Aparin, "A monolithic low phase noise 1.7GHz CMOS VCO for zero-IF cellular CDMA receivers", IEEE International Solid-State Circuits Conference (IEEE Cat. No.04CH37519), vol.1, pp. 396-535, 2004.

TELECOMMUNICATION SYSTEMS AND TECHNOLOGY, COMPUTER SYSTEMS AND INTERNET TECHNOLOGIES, AND INFORMATICS AND COMPUTER SCIENCE – POSTER SESSION

Cloud structure development for embedded systems

Neven Nikolov[1] and Svetlin Antonov[2]

Abstract – **This article describes the development of a cloud structure for communicating with IoT embedded systems, data synchronization between an embedded system that reads data from sensors and the cloud structure visualization. Steps are considered in designing the IoT cloud with Node Red IBM.**

Keywords – **Embedded Systems, IoT, Protocol, Cloud, Node Red.**

I. Introduction

The Internet of Things IoT are embedded systems that connect to Cloud (Server) by exchanging data between the cloud and other embedded systems [1-8]. Their role is to read sensors and manage different objects, depending on their purpose [9]. IoT are widely used in smart homes, medicine, electricity, industry 4.0 and other spheres. They need to communicate with Cloud Structures, where they publish data from sensors and parameters of managed objects, and these data must be processed and decisions made. The role of the Cloud is to collect data, make a decision based on pledged algorithms, and provide parameters to the embedded systems associated with it. For example, to manage the heating of smart houses, it is necessary to take into account the room temperature, the outside temperature and on the basis of other set parameters to activate the heating in the given room in the house [10]. The main idea is the automation of the whole process as well as the reduction of energy consumption, facilitating the user to monitor and manage his smart home. For this purpose, it is necessary to design the Cloud as intended.

This article describes building a private Cloud for IoT embedded systems, the technology behind Node Red IBM and the embedded NodeMcu Esp8266 system.

II. Experimental IoT cloud

They have different varieties of the IoT Cloud. Depending on the purpose and type of use, it may be a private and public cloud. The public is convenient if we do not have confidential data and the system is scalable. For example, the number of connected devices increases continuously, resulting in increased server load and more and more computational capability. This is not a suitable solution for a private cloud because more system administrators and technicians need to support the cloud structure to add more and more servers. But the public disadvantages are that there is a vulnerability to data protection, as well as some platform-dependent features.

For the construction of a private cloud, a Dell Power edge R510 server was used, with the Centos 7 operating system installed. For server technology, Node Red, a flow-based development tool developed by IBM, is used to connect together hardware devices, APIs and Online services that are part of the Internet of Things. Node red is an editor working in the browser and is useful for creating JavaScript functions. The application can be stored or shared as well as reused. To run Node Red, you need to install the Node.js - runtime server technology. The project described in Node Red is stored using the JSON type format, both for communication to the server being used and TLS encrypted connection.

For the built-in system that connects to the Cloud realized with Node Red, the NodeMcu Esp8266 embedded system, to which the DHT 22 temperature and humidity sensor is connected, is used. The Cloud Architecture is shown in Fig. 1 a of the embedded system of Fig.2 .

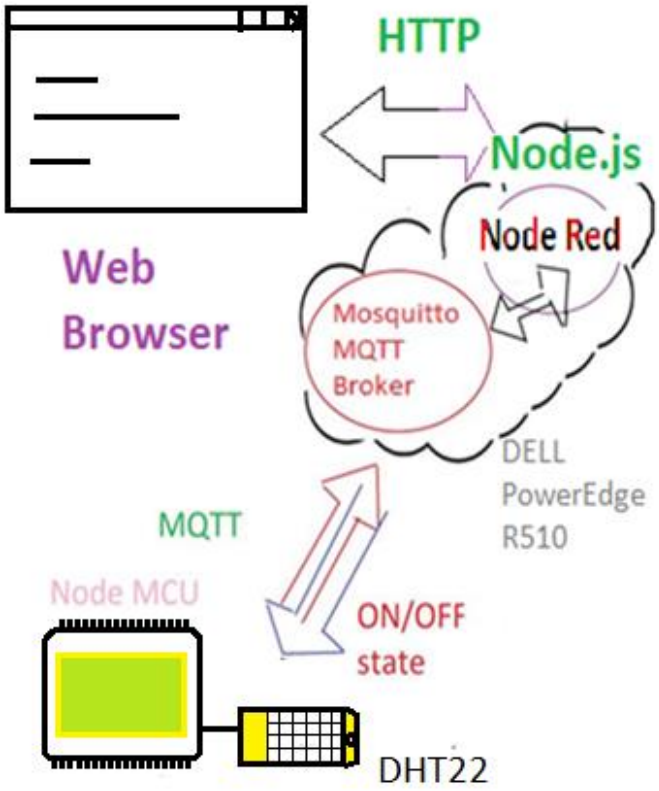


Fig. 1. Cloud structure architecture built with Node Red

III. Software realization and architecture

With regard to Cloud Building, the following technologies were used - Node Red, Node.js, Mosquitto MQTT broker installed on the Centos 7 operating system on a Dell Power

[1]Neven Nikolov is with the Faculty of Computer Systems and Technologies at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: n.nikolov@tu-sofia.bg, tel. 0886 238 266

[2]Svetlin Antonov is with the Faculty of Telecommunications at the Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria. e-mail: svantonov@yahoo.com

edge R510 server. A web based Node Red environment is used to build the backend. It is invoked by running the server by the command in the terminal:

> node-red

after which the system reports - Server now running at http://127.0.0.1:1880. By entering the IP address and port of the Node Red process into the Web browser, the graphical development environment is loaded Fig. 3.

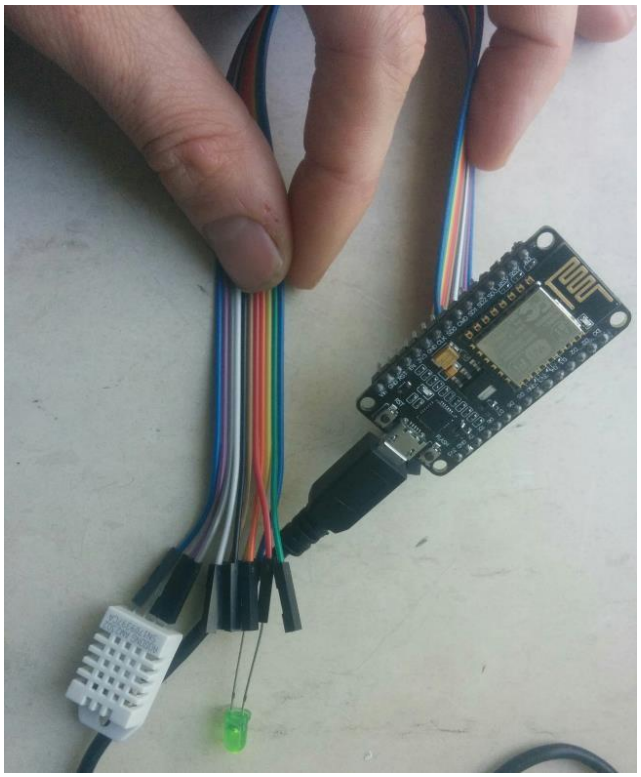


Fig. 2. Experimental staging of NodeMcu Eps8266 embedded system

The embedded system NodeMcu Esp8266 should connect to MQTT Mosquitto Broker first and after Node Red will receive message. For this purpose, in the Node Red functional module search box, mqtt input nodes are selected for temperature and humidity and they are set to connect to MQTT Mosquitto broker. For this purpose, the following settings are shown in Fig. 4, localhost: 1883 which is the address and gateway of the MQTT broker server and the name of Topic - temperature. On Fig. 5 are shown Edit switch node configuration for LED control of embedded system and setting the MQTT input nodes configuration for humidity in Fig. 6.

The parameters with humidity settings are the same, only the difference is the Topic - humidity name is set. To the temperature and humidity input nodes are connected a chart node for temperature readings and a gauge node for humidity. To the embedded system as shown in Fig. 1 is an LED connected, which will be managed via a Web browser. For this purpose, a switch node in Node Red is used, which is a button with two defined states - On/Off. The MQTT output node is linked to the button and its configuration is a connection to the MQTT broker server and Topic name - switchLed.

Regarding the software implementation of the embedded system Esp8266, an MQTT library is used, and the device send to the MQTT Broker server status of sensor. The embedded system sends the values read from the DHT22 sensor for temperature and humidity periodically within 10 seconds to the broker. It monitors continuously for commands from MQTT broker, in the case of activating the LED through the user's web browser Fig.8 . Esp8266 works as Publisher / Subscriber mode, because he receive command for turn of and off the LED diode and send data to Cloud..

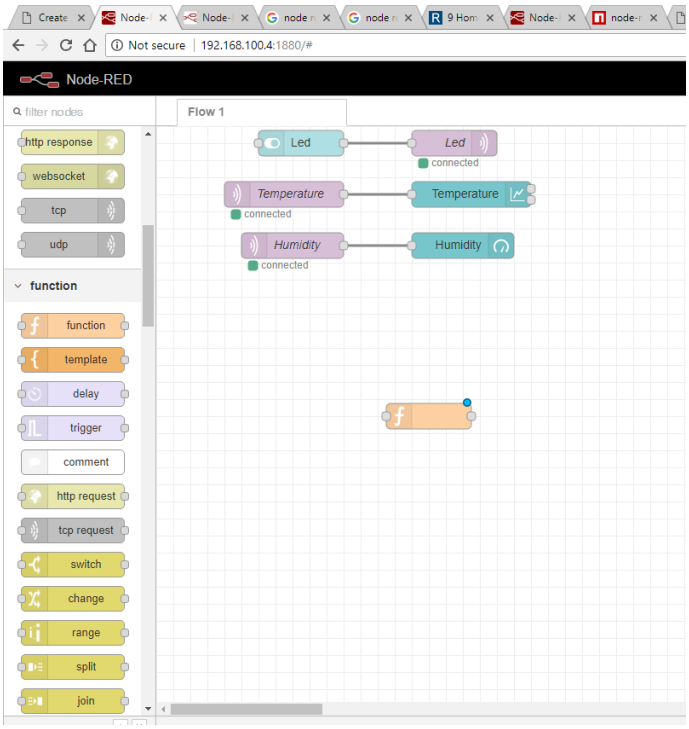


Fig. 3. Development environment for Node Red IBM

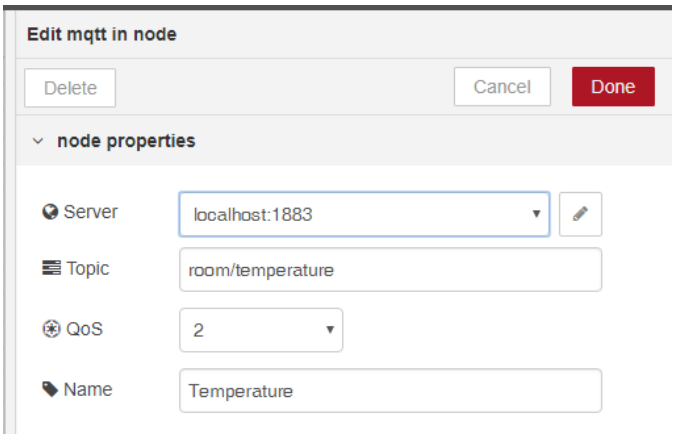


Fig. 4. MQTT input node configuration for temperature from embedded system

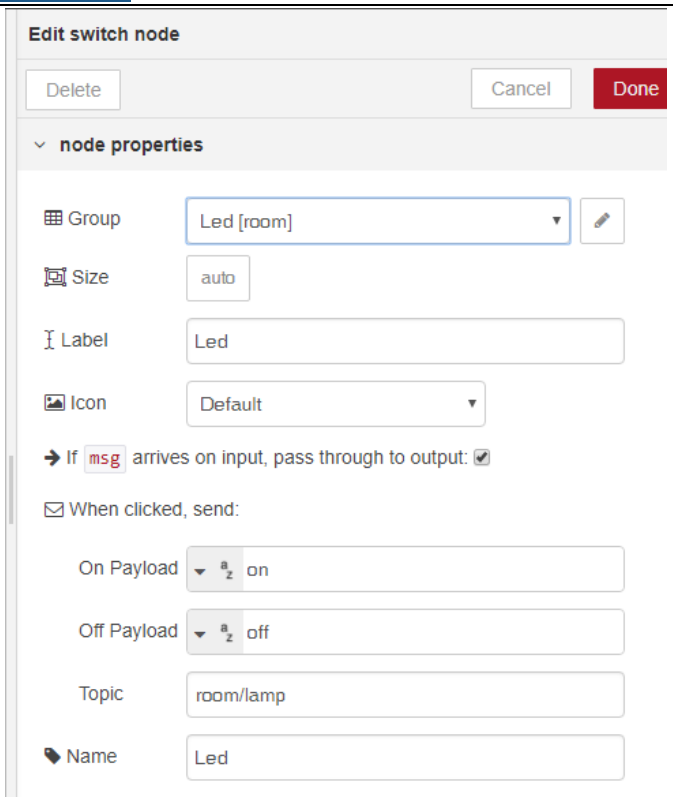


Fig. 5. Edit switch node configuration for LED control of embedded system

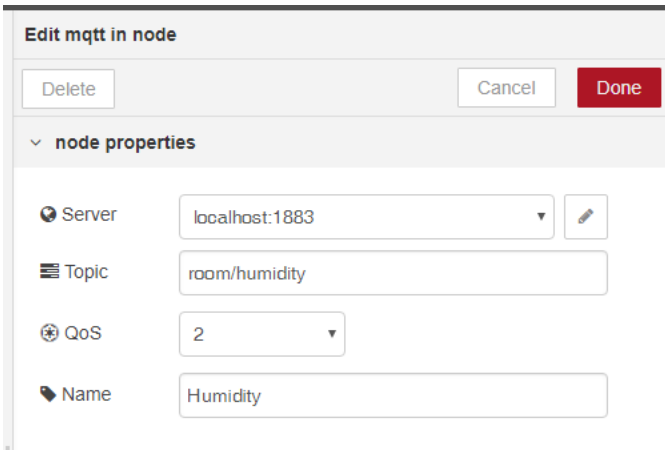


Fig. 6. Setting the MQTT input nodes configuration for humidity

IV. Experimental Results

In Fig. 7 are show a graphical user environment. The values of the measured humidity and temperature parameters are presented, and the graphs named ”Temperature” show the changes in temperature at certain hours. In Fig. 7 are shown pushed Led button for controlling the LED diode to turn ON. In Fig. 7 are shown the Humidity graph.

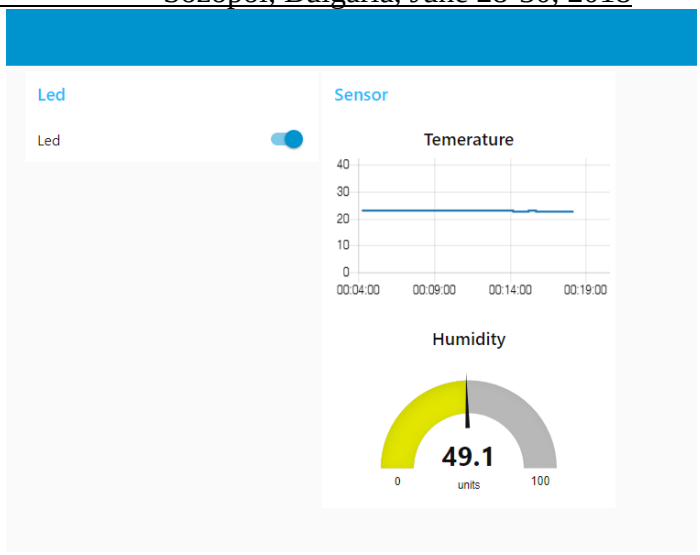


Fig. 7. Temperature and humidity reading in Web interface

In Fig. 9,10,11 are shown communication between Cloud and IoT embedded device and there are shown MQTT Packet anatomy for message topics. For this measurement are used program tool named WireShark, with whom we can measure the anatomy of packages. With program we can see packet parameters like Frame length, Data length and more parameters. On figures we can see the messages in ASCII format in MQTT protocol, because MQTT used port 1883 to Mosquito Broker and that port isn’t encrypt. In table I are shown parameters Frame Length bytes, Data length bytes and Values of sensors.

TABLE I
PARAMETERS OF MQTT MESSAGES BETWEEN IOT DEVICE AND CLOUD

MQTT Topic	Frame Length Bytes	Data Length Bytes	Value of sensor	Sensor
room/lamp	56	2	off	LED
room/temperature	80	26	27.36	DHT22
room/humidity	77	23	40.69	DHT22

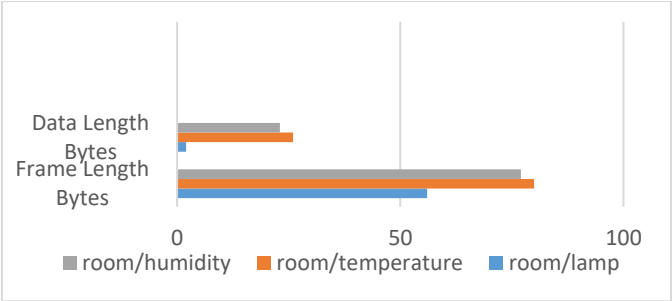


Fig. 8. Size of bytes in MQTT messages

In Fig. 8 are shown the parameters of MQTT messages between IoT device and Cloud. From table can calculate the header part of MQTT message with Eq 1:

Frame length – Data length = Header MQTT Packet (1)

and header values of MQTT protocol in messages between Cloud and IoT device are shown in Table II.

TABLE II
HEADER VALUE OF MQTT PROTOCOL IN MESSAGES BETWEEN CLOUD AND IOT DEVICE

MQTT Topic	Frame Length	Data Length	Header Frame – Data =
room/lamp	56	2	54
room/temperature	80	26	54
room/humidity	77	23	54

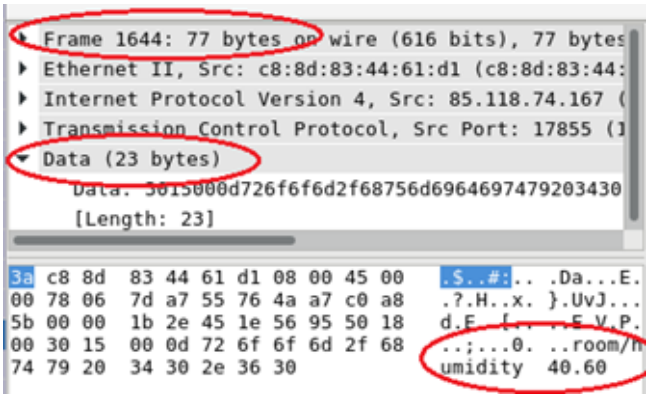


Fig. 9. MQTT packet anatomy for message topic "room/humidity"

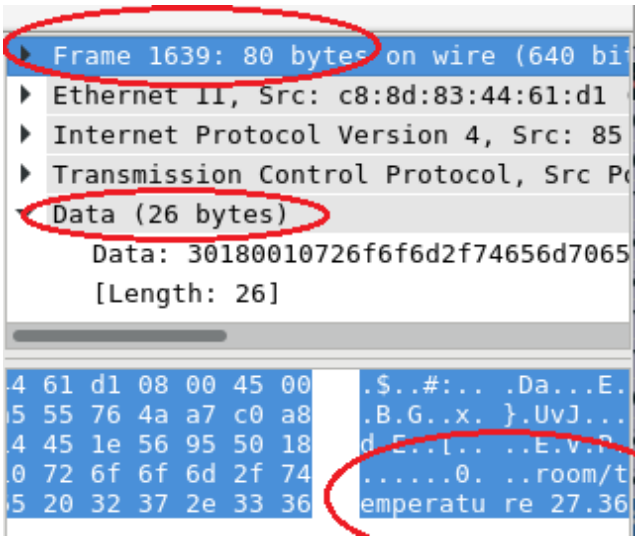


Fig . 10. MQTT Packet anatomy for message topic "room/temperature"

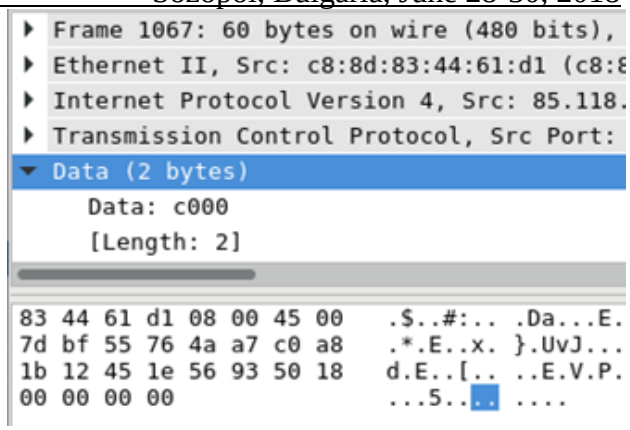


Fig. 11. MQTT Packet anatomy for message topic "room/lamp"

V. CONCLUSION

In this article is described how is released Cloud system for controlling an LED diode and reading data from temperature and humidity sensor. The management and control are realized via Web interface. It is achieved a real-time measurement of parameters such as temperature and humidity. A concept for building a IoT Cloud is shown using Node Red technology. In section IV are measured the protocol between Cloud and IoT embedded system. In Table 2 are shown packet payload Data and Frame Length. From all these studies, it see the MQTT protocol parameters for all messages. The system has big advantage over analog systems – Cloud or IoT devices can be anywhere and far away without losing any information.

REFERENCES

[1] J. Weinman, "The Strategic Value of the Cloud", IEEE Cloud Computing, vol.2, pp 66-70, 2015
[2] W., Peng Xu, L. T. Yang, *Secure Data Collection, Storage and Access in Cloud-Assisted IoT*, IEEE Cloud Computing, 2018
[3] A.Alshehri, R. Sandhu, "Access Control Models for Virtual Object Communication in Cloud-Enabled IoT", IEEE Computer society, pp:16-25, 2017
[4] H.Truong, S.Dustdar, "Principles for Engineering IoT Cloud Systems", vol.2, pp.68-76, 2015.
[5] S.Nastic, H. Truong, S. Dustdar, "A programming model for resource-constrained iot cloud edge devices", Banff, IEEE international Conference on systems, 2017
[6] E. Lingg, G. Leone, K. Spaulding, R. BFar, "Cloud based employee health and wellness integrated wellness application with a wearable device and the HCM data store", Internet of Things (WF-IoT), Seoul, 2014
[7] N. Fujii, N. Koike, *A FPGA-based Remote Laboratory in the Hybrid Cloud*, Chester, Cyberworlds , 2017
[8] G.Merlino, D.Bruneo, F.Longo, A,Puliafito, S.Distedano, "A Novel Paradigm for Smart Cities through IoT Clouds", IEEE press, 2015
[9] M. Lekic, G. Garadasevic, "IoT sensor integration to Node-RED platform", East Sarajevo, Infoteh-Jahorina, 2018
[10] Y. Obuchi, T. Yamasaki, K. Aizawa, S.Toriumi, M.Hayashi, "Measurement and evapuation of comfort levels of apartments using IoT sensors", Las Vegas, Consumer Electronics (ICCE), 2018

Intercept probability evaluation of wireless sensor networks over composite fading environment

Jelena A. Anastasov, Aleksandra M. Cvetković, Daniela M. Milović, Dejan N. Milić and Goran T. Đorđević

Abstract –**In this paper, we have determined the physical layer security of system with an arbitrary number of sensors which send the sensed data to the sink. An eavesdropper tries to intercept the communication of each sensor-sink channel under composite Generalized *K* fading conditions. Looking for the optimal sensors' scheduling scheme, in order to minimize eavesdropper's overhearing, the expressions for evaluating intercept probability in the case of round-robin and best-node sensors' scheduling are derived. Obtained analytical results are in term of Meijer's *G* functions. Based on analytics, the impact of fading/shadowing condition parameters, the number of sensors, elected sensors' scheduling as well as the impact of various the average main signal-to-eavesdropper's signal ratios on intercept probability is analyzed, and presented through numerical results.**

Keywords **–composite fading channel, physical layer security, probability of intercept, wireless sensor network**

I. INTRODUCTION

Initially, wireless sensor networks (WSNs) were used for variety of purposes in military, industry, and today they are challenging in applications such as Internet of Things, smart grid, smart home, etc [1]. Incorporation of WSN provides data sensing, monitoring and communication controlling [2]. The main issue to be taken into account when implementing WSN is for certain system security throughput. Namely, it is very hard to keep simultaneously the WSN's reliability and security on some enviable level, due to the fact that sensors' communicate over an open radio channel medium [3].

A lot of papers deal with traditional cryptographic techniques in securing wireless communications. Still cryptography utilization is not quite suitable in WSN because of high hardware complexity requirements and large energy consumption. Additionally, an eavesdropper as authorized or unauthorized WSN user usually owns unlimited computing power and thus can easily break down confidential keys using brute-force attack.

In this context, physical layer security is an alternative in securing WSN based on exploiting the wireless channel propagation characteristics [3]. The physical layer security works were established by developing higher secrecy rates for typical wiretap channel so-called Wyner's channel consisting of a source, a destination node and an eavesdropper [4]-[6]. The secrecy capacity of such a wiretap model over non-small scale fading channels was investigated in [4]. The security system enhancement over generalized Gamma fading channels was presented in [5]. The security performance for classic Wyner's model over generalized *K* (GK) fading channels was studied in [6]. Relying on mixture gamma distribution in modeling the signal-to-noise ratio (SNR) over generalized *K* channels, novel analytical representations of secrecy capacity and secure outage probability were given in [7]. Again, considering approximate modeling of composite Generalized *K* fading channels, the security of a single-input-multiple-output system model was analyzed in [7]. Both, the destination node and an eavesdropper were equipped with multiple antennas, and both active and passive eavesdropper's overhearing was considered.

Since wireless sensors are usually powered by limited batteries sensor scheduling was proposed as a less energy intensive scheme for WSN security [8]. Authors in [8] have proposed optimal sensor scheduling scheme to maximize the secrecy capacity of an industrial WSN over Nakagami-*m* fading channels.

In this paper, we analyze WSN security performance in scenario with multiple sensors and a single sink. The communication is performed in the presence of an eavesdropper over generalized *K* fading channels. We picked the sensor scheduling analyzing method in order to outperform the conventional relay selection [9] or artificial noise method [10], also avoiding high implementation complexity and saving sensors' battery life. Thus, the intercept probability expression for optimal scheduling scheme is derived. Also, the probability of intercept for round-robin scheduling scheme, as a benchmark, is presented. The influence of various systems' parameters on intercept occurrence is analyzed and discussed in the section Numerical results.

II. SYSTEM MODEL AND PROBLEM FORMULATION

We consider a WSN that contains *N* sensors and one sink. The set of sensors communicate with the sink using the orthogonal multiple access methods such as the time division multiple access or orthogonal frequency division multiple access. An unfavorable licensed or unlicensed WSN node, marked as an eavesdropper attempts to intercept the data transmitted from the scheduled sensor to the sink. Typically, in an orthogonal channel, a sensor with the highest data throughput is scheduled to communicate with the sink.

In the system under consideration, we rely on the physical layer security aiming sensor scheduling schemes which differs from the traditional scheduling method. Namely, we assume

[1]Jelena A. Anastasov, Aleksandra M. Cvetković, Daniela M. Milović, Dejan N. Milić and Goran T. Đorđević, are with the University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mails: {jelena.anastasov, aleksandra.cvetkovic, daniela.milovic, dejan.milic, goran.t.djordjevic }@elfak.ni.ac.rs.

that not only the channel state information (CSI) of the main link is known at the sink but either the wiretap channel CSI is also available. This is a justifiable assumption because the eavesdropper could be a legitimate user in WSN who can be interested in tapping of some secrecy data.

Let us express the received SNR from the *i*th main (sensor-sink) link as

$$\gamma_{si} = \frac{|h_{si}|^2 P_i}{\sigma_{si}^2}, \quad i = 1, \ldots, N, \tag{1}$$

where h_{si} is a fading coefficient on the channel between the *i*th sensor and the sink, P_i is the transmission power and σ_{si}^2 is a variance of zero-mean additive white Gaussian noise (AWGN). According to the Shannon capacity formula [11], we can evaluate the channel capacity of the *i*th main link as

$$R_s(i) = \log_2(1 + \gamma_{si}) \tag{2}$$

We have already assumed a possible presence of an eavesdropper that attempts to intercept transmission on the *i*th path. The attacker has a perfect knowledge of legitimate transmissions from each main link, except of the signals that are confidential [8]. So, the SNR tapped by the eavesdropper can be written as

$$\gamma_{ei} = \frac{|h_{ei}|^2 P_i}{\sigma_{ei}^2}, \quad i = 1, \ldots, N, \tag{3}$$

with h_{ei} being a fading coefficient of the wiretap channel between *i*th sensor and eavesdropper and σ_{ei}^2 being the variance of AWGN. Further, the *i*th wiretap channel capacity can be calculated as

$$R_e(i) = \log_2(1 + \gamma_{ei}). \tag{4}$$

Therewith, the secrecy capacity of specified *i*th sensor can be defined as a difference between the channel capacity of the main link and sensor-eavesdropper link [11]

$$C_{\text{secrecy}}(i) = R_s(i) - R_e(i). \tag{5}$$

III. Intercept Probability Evaluation

A. Round-robin scheduling intercept probability

When *N* sensors, all by random, access a given transmission channel with equal chance for sending its sensed data, the scheduling scheme corresponds to the conventional round-robin scheduling scheme.

We consider that the *i*th sensor is scheduled to transmit confidential signal with a rate $R_s(i)$ which is specified as the maximum achievable rate. The probability of intercept is then the probability that secrecy capacity of the i*th* link becomes non-positive which yields to [8]

$$P_{\text{int}}^i = \Pr\left[C_{\text{secrecy}}(i) < 0\right] = \Pr\left[R_s(i) < R_e(i)\right]. \tag{6}$$

By substituting (2) and (4) in (6), and after some mathematical manipulations we get

$$P_{\text{int}}^i = \Pr\left[\gamma_{si} < \gamma_{ei}\right] = \int_0^{\infty}\left(\int_0^{\gamma_{ei}} p_{\gamma_{si}}(\gamma_{si})\, d\gamma_{si}\right) p_{\gamma_{ei}}(\gamma_{ei})\, d\gamma_{ei}. \tag{7}$$

It was earlier pointed out that the wireless channels between neighboring nodes, are modeled by Generalized *K* fading model. Thus, the probability density function (PDF) of SNR over the *i*th main link has the following form [12]

$$p_{\gamma_{si}}(\gamma_{si}) = \frac{2}{\Gamma(m_{si})\Gamma(k_{si})}\left(\frac{m_{si}k_{si}}{\bar{\gamma}_{si}}\right)^{\frac{m_{si}+k_{si}}{2}} \gamma_{si}^{\frac{m_{si}+k_{si}-2}{2}} \times K_{k_{si}-m_{si}}\left[2\left(\frac{m_{si}k_{si}\gamma_{si}}{\bar{\gamma}_{si}}\right)^{1/2}\right], \tag{8}$$

where $\Gamma(\cdot)$ denotes the Gamma function [13, eq. (8.310)], $K_{\beta}(\cdot)$ is the βth order modified Bessel function of the second kind [13, eq. (8.432.3)], while m_{si} and k_{si} denote the multipath fading and shadowing parameters, respectively. The parameter $\bar{\gamma}_{si} = \mathrm{E}[\gamma_{si}]$ is the *i*th main link average SNR ($\mathrm{E}[.]$ is the expectation operator).

Similarly, the PDF that describes SNR on the *i*th wiretap link is

$$p_{\gamma_{ei}}(\gamma_{ei}) = \frac{2}{\Gamma(m_{ei})\Gamma(k_{ei})}\left(\frac{m_{ei}k_{ei}}{\bar{\gamma}_{ei}}\right)^{\frac{m_{ei}+k_{ei}}{2}} \gamma_{ei}^{\frac{m_{ei}+k_{ei}-2}{2}} \times K_{k_{ei}-m_{ei}}\left[2\left(\frac{m_{ei}k_{ei}\gamma_{ei}}{\bar{\gamma}_{ei}}\right)^{1/2}\right], \tag{9}$$

with m_{ei} and k_{ei} being the multipath fading and shadowing shaping parameters over *i*th wiretap link, respectively and $\bar{\gamma}_{ei} = \mathrm{E}[\gamma_{ei}]$.

So by substituting (8) in (7), then transforming the Bessel *K* function into Meijer's *G* function according to [14, eq. (8.4.23.1)], and relying on [15, eq. (26)] we solved the first integral in (7) and we get

$$P_{\text{int}}^i = \int_0^{\infty} \frac{G_{1,3}^{2,1}\left(\frac{m_{si}k_{si}\gamma_{ei}}{\bar{\gamma}_{si}}\middle| \begin{matrix} 1 \\ k_{si}, m_{si}, 0 \end{matrix}\right)}{\Gamma(m_{si})\Gamma(k_{si})} p_{\gamma_{ei}}(\gamma_{ei})\, d\gamma_{ei}, \tag{10}$$

where $G_{p,q}^{m,n}(\cdot)$ denotes the Meijer's *G* function [13, eq. (9.301)]. Further, by substituting (9) in (10) and using [14, eq. (2.24.3.1)], we derive the probability of intercept of that overheard sensor-to-sink link, as

$$P_{\text{int}}^{i}=\frac{G_{3,3}^{2,3}\left(\frac{m_{si}k_{si}}{m_{ei}k_{ei}\lambda_i}\middle|\begin{matrix}1,1-k_{ei},1-m_{ei}\\ k_{si},m_{si},0\end{matrix}\right)}{\Gamma(m_{si})\Gamma(k_{si})\Gamma(m_{ei})\Gamma(k_{ei})}, \tag{11}$$

with $\lambda_i=\bar{\gamma}_{si}/\bar{\gamma}_{ei}$ being the ith average main signal-to-eavesdropper's signal ratio (MER).

The round-robin scheduling intercept probability is the mean of all N intercept probabilities, leading to

$$P_{\text{int}}^{\text{round}}=\frac{1}{N}\sum_{i=1}^{N}P_{\text{int}}^{i}. \tag{12}$$

B. Best-node sensor scheduling intercept probability

Relying on (5), the best-node scheduling criterion by which the optimal sensor is scheduled to transmit confidential signal to the sink, can be expressed as [8]

$$\begin{aligned}\text{Optimal User}&=\arg\max_{i\in S}C_{\text{secrecy}}(i)\\&=\arg\max_{i\in S}\log_2\left(\frac{1+\gamma_{si}}{1+\gamma_{ei}}\right),\end{aligned} \tag{13}$$

where S denotes the set of N sensors. We assume that each sensor estimates its own CSI and sends it to the sink. The sink collects all the sensors' CSI and determines the optimal one for communication. So, the secrecy capacity for this scenario can be obtained as [8]

$$C_{\text{secrecy}}^{\text{best-node}}=\max_{i\in S}\log_2\left(\frac{1+\gamma_{si}}{1+\gamma_{ei}}\right). \tag{14}$$

Moreover, the expression for intercept probability of the best-node scheduling scheme becomes

$$\begin{aligned}P_{\text{int}}^{\text{best-node}}&=\Pr\left[C_{\text{secrecy}}^{\text{best-node}}<0\right]\\&=\Pr\left[\max_{i\in S}\log_2\left(\frac{1+\gamma_{si}}{1+\gamma_{ei}}\right)<0\right].\end{aligned} \tag{15}$$

For different sensors, random variables γ_{si} and γ_{ei} are independent of each other, so the previous equation can be rewritten as

$$\begin{aligned}P_{\text{int}}^{\text{best-node}}&=\prod_{i=1}^{N}\Pr\left[\log_2\left(\frac{1+\gamma_{si}}{1+\gamma_{ei}}\right)<0\right]\\&=\prod_{i=1}^{N}\Pr\left[\gamma_{si}<\gamma_{ei}\right]\\&=\prod_{i=1}^{N}P_{\text{int}}^{i}\end{aligned}. \tag{16}$$

IV. NUMERICAL RESULTS

Numerical results are obtained according to derived expressions (11), (12) and (16) in *Mathematica* software package. The Meijer's G functions are special built-in functions in aforementioned software.

Fig. 1 shows the intercept probability dependence on the average MER of each ith path ($\lambda_i=\lambda$, $i=1,\ldots,N$). The both conventional round-robin and the best-node scheduling schemes are analyzed when $N=5$ sensors are active in considered WSN. From the figure, we can notice that the optimal scheduling scheme sufficiently outperforms traditional round-robin scheduling scheme. For λ=14dB, P_{int} values for optimal scheduling scheme are low, even lower than 10^{-6} in lighter fading/shadowing channel conditions, while P_{int} value for round robin scheduling are barely $5\text{x}10^{-3}$ in the presence of the most favorable channel conditions ($m_{si}=m_{ei}=3.1$, $k_{si}=k_{ei}=3.7$). Bad channel conditions over the main and wiretap links are fertile for eavesdropper's overhearing. Differently interpreted, the deep faded and severely shadowed channels decrease system's secrecy capacity.

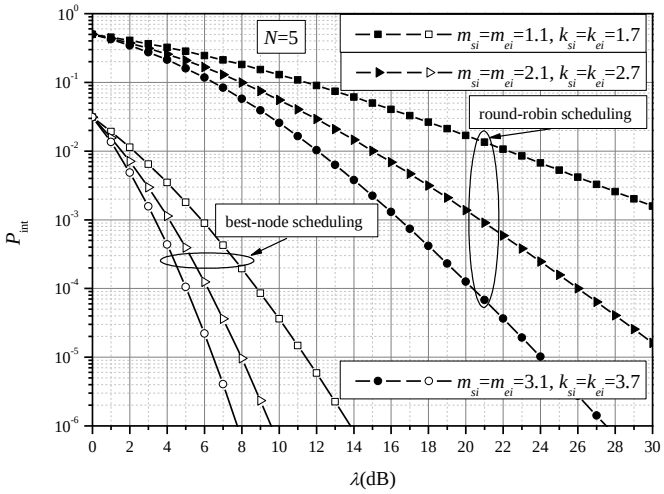


Fig. 1. Intercept probability versus the average MER over different fading/shadowing channel conditions

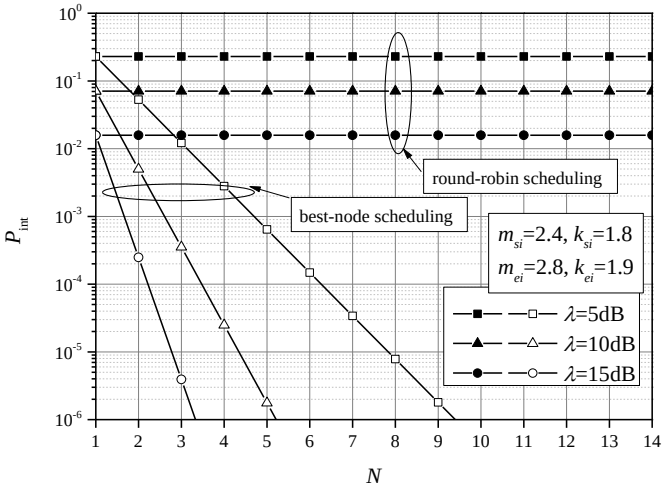


Fig. 2. Intercept probability in the function of different number of active sensors for both scheduling schemes

Fig. 2 illustrates the intercept probability versus number of sensors in WSN. We assume that the distances between

neighboring nodes are small which refer to the scenario with similarly composite fading conditions over links. This can explain the constant value of P_{int} for round robin scheduling scheme versus number of sensors. Namely, by averaging all P_{int}^{i} over N, for identical fading/shadowing parameters, it is obvious that the final P_{int} is only dependent on the MER value. On the other hand, by increasing the number of WSN users, the best-node scheduling intercept probability decreases, even in the scenario under consideration, especially when the average MER increases. For example, when number of sensors increases from N=5 to N=7, the P_{int} decrease for an order of magnitude when λ=5dB, while the decrease of the P_{int} is almost two orders of magnitude when λ=10dB, for the same increase of number of users.

In overall, Fig. 1 and Fig. 2 confirm the advantage of exploiting the best-node scheduling scheme in defending against the eavesdropper's attack.

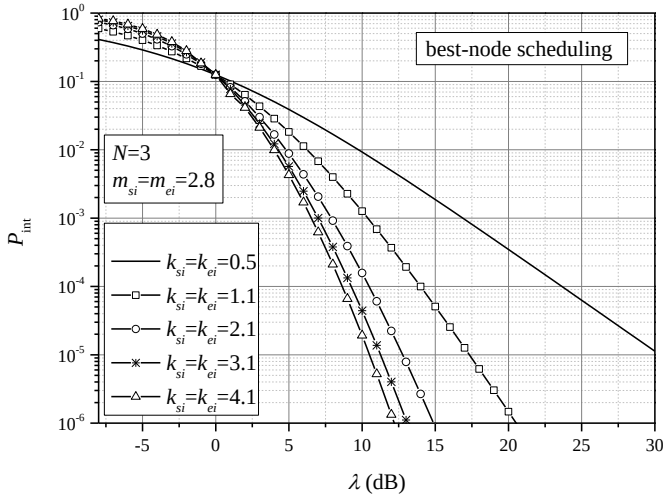


Fig. 3. Intercept probability for the best-node scheduling scheme over various shadowing channel conditions

The intercept probability of optimal sensors' scheduling scheme versus shadowing shaping parameter of the main and wiretap links is presented in Fig. 3. Lighter shadowing channel conditions allow securer sensor-to-sink communications. It is obvious that the worst shadowing channel condition scenario ($k_{si} = k_{ei} = 0.5$) implies the worst security WSN case. The possibility of interception events' occurrence decreases when shadowing shaping parameters increase. In addition, we can notice that for $\lambda < 0$dB the influence of channel state conditions on the probability of intercept is vice verse.

V. Conclusion

In the paper, we investigated the physical layer security of WSN over composite fading channels employing the optimal sensors' scheduling scheme. We derived the closed-form expression of intercept probability, under given circumstances. Obtained results showed that increasing number of WSN sensors benefits only when the best-node scheduling scheme is applied. Favorable channel conditions i.e. higher values of fading/shadowing shaping parameters do improve the secrecy in sensor-sink communications.

Proposing of novel optimal sensors' scheduling scheme in order to enhance WSN security will be considered in our further work.

Acknowledgement

This work was supported by Ministry of science and technology development of Republic of Serbia (grants III-44006, TR-32028 and TR-32051).

References

[1] V. C. Gungor, B. Lu and G. P. Hancke, "Opportunities and Challenges of Wireless Sensor Networks in Smart Grid," IEEE Trans. Ind. Electron., vol. 57, no. 10, pp. 3557-3564, 2010.

[2] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam and E. Cayirci, "Wireless sensor networks: a survey," Computer Networks, vol. 38, no. 4, pp. 393-422, 2002.

[3] R. Liu and W. Trappe, *Securing wireless communications at the physical layer*, New York, Springer, 2009.

[4] G. Pan and C. Tang and X. Zhang and T. Li and Y. Weng and Y. Chen, "Physical-Layer Security Over Non-Small-Scale Fading, " IEEE Trans. Veh. Technol., vol.65, no.3, pp. 1326 – 1339, 2016.

[5] H. Lei, C. Gao, Y. Guo, and G. Pan, "On Physical Layer Security Over Generalized Gamma Fading Channels," IEEE Commun. Lett., vol. 19, no. 7, pp. 1257-1261, 2015.

[6] H. Lei, H. Zhang, I. S. Ansari, C. Gao, Y. Guo, G. Pan, and K.A. Qaraqe, "Performance Analysis of Physical Layer Security Over Generalized-K Fading Channels Using a Mixture Gamma Distribution," IEEE Commun. Lett., vol. 20, no. 2, pp. 408-411., July 2016.

[7] H. Lei, I.S. Ansari, C.G. Yongcai, G.G. Pan, K.A. Qaraqe, "Secrecy Performance Analysis of SIMO Generalized-K Fading Channels," Frontiers of Information Technology & Electronic Engineering, vol. 17, no.10, 2016.

[8] Y. Zou and G. Wang, "Intercept behavior analysis of industrial wireless sensor networks in the presence of eavesdropping attack," IEEE Trans. Indust. Inform., vol. 12, no. 2, pp. 780-787, 2016.

[9] Y. Zou, X. Wang, and W. Shen, "Optimal relay selection for physical layer security in cooperative wireless networks," IEEE J. Sel. Areas Commun., vol. 31, no. 10, pp. 2099-2111, 2013.

[10] S. Goel and R. Negi, "Guaranteeing secrecy using artificial noise," IEEE Trans. Wire. Commun., vol. 7, no. 6, pp. 2180-2189, 2008.

[11] M. Bloch, J. Barros, M. Rodrigues, and S. McLaughlin, "Wireless information-theoretic security," IEEE Trans. Inf. Theory, vol. 54, no. 6, pp. 2515-2534, June 2008.

[12] M. Kostic, "Analytical approach to performance analysis for channel subject to shadowing and fading," IEE Proc., vol. 152, no. 6, pp. 821–827, 2005.

[13] I. S. Gradshteyn and I. M. Ryzhik, *Tables of integrals, series, and products*, fifth edition, New York, Academic Press, 1994.

[14] A. P. Prudnikov, Y. A. Brychkov, and O. I. Marichev, *Integral and Series: Volume 3, More Special Functions*, New York, CRC Press Inc, 1990

[15] V. S. Adamchik, and O. I. Marichev, "The algorithm for calculating integrals of hypergeometric type functions and its realization in reduce system," in *Proc. of the inter. symp. on Sym. and comp.*, Tokyo, Japan, pp. 212-224, 1990.

Low-Complex Real-Time Edge Detection with a Compact FPGA Architecture for Embedded IoT Applications

Nikola Rendevski, Darko Pajkovski, Zoran Kotevski, Ilija Hristoski, Ramona Markoska

Abstract – **In this work, we present a digital design of compact Field Programmable Gate Array (FPGA) - based architecture for real-time edge detection. At this point, we are focused on pure-hardware realizations of filtering and edge-detection algorithms as widely used techniques for various industrial and entertainment applications. The advantage of FPGAs as digital signal processing (DSP) platforms for real-time video and image processing lays mainly in their reconfigurable and re-usable structure, capable to efficiently exploit spatial and temporal parallelism. Based on the results of the digital FPGA synthesis of pipelined Sobel edge detection architecture, we discuss on the potential and challenges towards design of low-complex pure hardware-based video processing devices for embedded IoT applications.**

Keywords – **FPGA, Edge Detection, Sobel Algorithm, IoT**

I. INTRODUCTION

Image Processing (IMP) and Video Processing (VP) have been used for a long time in various industrial, medical, security and entertainment systems. In example, within the industrial environments, such techniques are still used to control the product quality and inspection procedures by replacing manual activities traditionally conducted by humans. In any case, the control, inspection and detection procedures are based on off-line IMP, or real-time VP processing, mainly by implementation of preprocessing algorithms such as thresholding and filtering or processing algorithms such as pattern matching and edge detection. Considering the Industry 4.0 and Tactile Internet paradigms, efficient IMP and VP capable to provide responses towards meeting the requirement for sub-millisecond latencies are one of the key aspects for entirely remote operated facilities and smart factories. Furthermore, nowadays we are all witnesses of the mass proliferation of image capturing devices (different forms of cameras) in our daily lives: in computers, mobile devices, cars etc. In such developmental trends, with the advances of the microelectronics, cameras are becoming drastically smaller, conduct faster image acquisition and produce high-definition output as expected by the modern application scenarios which require strict QoS levels. Moreover, such sensitive QoS inevitably requires more and more processing power which directly affect the complexity, cost and energy conservation capabilities of the systems. The future views of the fully networked society, based on technologies such Internet of Things (IoT), Cyber Physical Systems (CPS) or Wireless Sensor Networks (WSN), embedded as an inevitable part of the Fifth Generation Network (5G) framework, are expected to provide ubiquitous processing and multi-interface communication with limited DSP resources, mainly constrained by the size and energy consumption, towards justifying the vision for massive deployment on a battery-operated power sources. Nowadays, energy-efficient wireless sensor nodes falling within the concept of WSN, are mainly driven by low-power RISC processors and microcontrollers which are capable to provide readings from simple conventional sensors, low-sampling A/D conversions, simple processing and short-range wireless transmission. Expectations from such reduced functionality devices (RFDs) for any type of image or video processing are impossible, except some ultra-low-resolution imaging for realization of simple detection functionality. From the other side, fully functional devices (FFDs) pose processing capabilities enabled by certain type of embedded processor with higher processing power, capable to realize IMP and real-time VP, but still with limited performance. Such devices in the WSN model are equipped with additional power sources and energy harvesting mechanisms to achieve longer operation. Considering the above facts, the WSN FFD devices concept could be treated as model of a typical IoT embedded device. Various IoT development platforms exist such Raspberry Pi, Orange Pi and Arduinos, equipped with ARM 32- and 64-bit embedded processing cores. However, efficient higher-definition IMPs and VPs are still heavy load for those miniature computers as the dynamic imaging sensor fusion cause serious bottlenecks mainly because of speed-limited execution of software-based algorithms and memory access. This fact could lead to a conclusion that processor-centric systems based on general-purpose processors (GPP) designed only on the conventional Von-Neumann (so called program-stored computation) architectures are not standalone candidates for the processing unit in next-generation IoT embedded devices. The recent advances in FPGA manufacturing processes, as recently declared by Intel (Altera), is moving to feasible 10 nm based on FinFET technology, which makes the future FPGAs speeds comparable to ASIC processor chips. Such trends truly support the System-on-Chip (SoC) integrated digital hardware architectures consisted of so-called Hard Processor Systems (HPS), interconnection networks (buses) and reconfigurable FPGA, not only on the same board, but also on the same VLSI

Nikola Rendevski, Darko Pajkovski, Zoran Kotevski and Ramona Markoska are with the Faculty of Information and Communication Technologies, "St. Kliment Ohridski University" – Bitola, Republic of Macedonia. Email: {nikola.rendevski, darko.pajkovski, zoran.kotevski, ramona.markoska}@fikt.edu.mk.
Ilija Hristoski is with the Faculty of Economics, "St. Kliment Ohridski" University - Bitola, Republic of Macedonia. Email: ilija.hristoski@uklo.edu.mk.

chip. Considering the fact that 65 nm FPGAs nowadays are widely available with reasonable sizes, price and performance suitable for the vast IMP and VP IoT applications, positions them as valuable alternatives for processing and co-processing subsystems of a typical IoT device.

Considering the above-mentioned available technologies, a question naturally rises: Which IMP and VP functions and routines are more suitable for certain processing architecture (FPGA, HPS-FPGA, soft processor etc.)? In this paper, based on the practical experiences from the physical FPGA synthesis of digital logic architecture for performing Sobel Edge Detection of real-time video and images, we conduct analyses with goal to propose processing scheme suitable for design of IoT devices with IMP and VP capabilities. The rest of the paper is organized as follows: In Chapter II we present the introduction to low-complex edge detection (ED) algorithms. Chapter III is dedicated to the experimental testbed. The results from the FPGA synthesis and complexity analyses and the conclusions are presented in Chapter IV.

II. Edge Detection Algorithms

The understanding of edge detection (ED) techniques and algorithms and their mathematical background is crucial to analyze the complexity of appropriate digital processing architecture and FPGA synthesis, which are the main focus of this work towards realization of low-complex IoT devices. In the following we present brief explanation of what ED is, its purpose and types. Edge detection is one of the fundamental tools in image and video processing, machine and computer vision - particularly in all areas where feature detection and feature extraction are crucial. The main purpose of edge detection is capturing important events and variations in different properties of the physical objects which are observed, such changes in the material structure, surface orientation, and variations in scene illumination. Sometimes, few lines in an illustration are often sufficient to unambiguously describe an object or a scene. Numerous IoT usage models ranging from industrial applications to entertainment require such analyses, in both, still images and video. From the design and processing point of view, the edge detection of complex graphic and multimedia content is not a trivial task.

ED, in general, is a mathematical method for identifying points within an image at which there is a significant change in the color, brightness or discontinuity along a particular orientation. Such points are typically organized into a set of curved line segments which enable identification of edges. The stronger the local intensity change, the higher is the evidence for an existence of edge at that position.

Although certain literature has considered the detection of ideal step edges, the edges obtained from natural images are usually not that simple. Instead, they are normally affected by one or several of the following effects: focal blur, penumbral blur caused by shadows of various light sources and shadings of a smooth objects. In mathematics, the amount of change with respect to spatial distance is known as the first derivative of a function, and the simplest detectors are based on analyses of this function.

Let us assume a simple image with white region surrounded by a dark (black) background. The intensity profile along one image line is one-dimensional function *f(x)* and the first derivative of the function, $\boldsymbol{f'(x) = \frac{df}{dx}(x)}$, results in a positive peak where the intensity increases, and a negative where the function decays, as shown on the Fig.1.

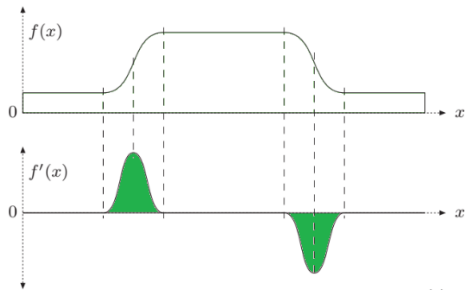


Fig. 1. The intensity profile along one image line, one-dimensional function *f(x)*, and the first derivative of the function

However, the line profile of real image is a discrete function *f(u)* and suitable estimation method is needed as the first derivative is undefined in the discrete domain (Fig. 2.).

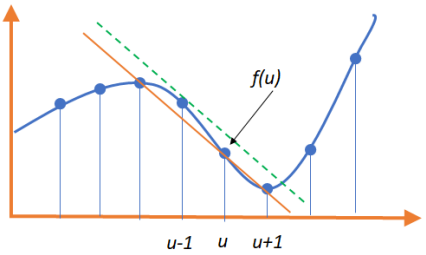


Fig. 2. Line profile of real image - discrete function *f(u)* and estimation method

A simple method for roughly approximating the slope of the tangent for a discrete function *f(u)* at position *u* is to fit a straight line through the neighboring function values *f(u−1)* and *f(u+1)*. The same applies for both horizontal and vertical direction to estimate the first derivative Eq.1. [1].

$$\frac{\boldsymbol{df}}{\boldsymbol{du}}(\boldsymbol{u}) \approx \frac{\boldsymbol{f(u+1) - f(u-1)}}{\boldsymbol{(u+1) - (u-1)}} = \frac{\boldsymbol{f(u+1) - f(u-1)}}{\boldsymbol{2}} \quad (1)$$

A derivative of a multidimensional function at one of its coordinate axes is called a partial derivative. For example: $\frac{\partial \boldsymbol{I}}{\partial \boldsymbol{u}}(\boldsymbol{u},\boldsymbol{v}), \frac{\partial \boldsymbol{I}}{\partial \boldsymbol{v}}(\boldsymbol{u},\boldsymbol{v})$ are partial derivatives of an image function *I(u,v)* along the *u* and *v* axes. The function $\boldsymbol{\nabla I(u,v)}$ is so called gradient or gradient vector of the image function *I* at a position *(u,v)*: $\boldsymbol{\nabla I(u,v)} = [\frac{\partial \boldsymbol{I}}{\partial \boldsymbol{u}}(\boldsymbol{u},\boldsymbol{v}) \ \frac{\partial \boldsymbol{I}}{\partial \boldsymbol{v}}(\boldsymbol{u},\boldsymbol{v})]^{\boldsymbol{T}}$.

The magnitude of the gradient (Eq.2.):

$$|\nabla \boldsymbol{I}(\boldsymbol{u},\boldsymbol{v})| = \sqrt{\left(\frac{\partial \boldsymbol{I}}{\partial \boldsymbol{u}}(\boldsymbol{u},\boldsymbol{v})\right)^2 + \left(\frac{\partial \boldsymbol{I}}{\partial \boldsymbol{v}}(\boldsymbol{u},\boldsymbol{v})\right)^2} \quad (2)$$

is not dependent from the image rotation and orientation of the underlying image structures. This property is important for isotropic localization of edges, and thus $|\nabla I|$ is the basis of many simple edge detection methods. The approximation of the first horizontal derivatives can be easily implemented by a linear filter (see with the coefficient matrix $\boldsymbol{L_x} = [-\frac{1}{2} \quad \mathbf{0} \quad \frac{1}{2}]$ and $\boldsymbol{L_y} = [-\frac{1}{2} \quad \mathbf{0} \quad \frac{1}{2}]^{\boldsymbol{T}}$, where the center pixel $I(u, v)$ is weighted with the zero coefficient or with other words - ignored.

The Sobel operator performs a 2D spatial gradient measurement of the image detecting regions of high spatial frequency that correspond to edges. It finds the approximate absolute gradient magnitude at each point of an input grayscale image. Corresponding to the application of the following filter masks to the image data, the Sobel operator is based on the following filters:

$$\boldsymbol{L_x} = \begin{bmatrix} -\mathbf{1} & \mathbf{0} & \mathbf{1} \\ -\mathbf{2} & \mathbf{0} & \mathbf{2} \\ -\mathbf{1} & \mathbf{0} & \mathbf{1} \end{bmatrix} \textbf{ and } \boldsymbol{L_y} = \begin{bmatrix} \mathbf{1} & \mathbf{2} & \mathbf{1} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\mathbf{1} & -\mathbf{2} & -\mathbf{1} \end{bmatrix} \quad (3)$$

The estimates for the local gradient components are obtained from the filter results by appropriate scaling. For the Sobel operator [1]:

$$|\nabla \boldsymbol{I}(\boldsymbol{u},\boldsymbol{v})| \approx \frac{\mathbf{1}}{\mathbf{8}}\begin{bmatrix} \boldsymbol{I} * \boldsymbol{L_x}(\boldsymbol{u},\boldsymbol{v}) \\ \boldsymbol{I} * \boldsymbol{L_x}(\boldsymbol{u},\boldsymbol{v}) \end{bmatrix} \quad (4)$$

It is worth to note that edge detection methods based on second derivatives exists, and the main idea behind this approach is that edges can be found at zero-crossings of the second derivatives within the image function. Since the second derivatives amplify the noise, methods for presmoothing must be applied using low-pass filters which will certainly affect the processing architecture complexity. Very used example of second derivative ED is Laplacian-of-Gaussian" (LoG) operator.

Considering the fact that in this work we are focused on analyses of low-complex ED solutions, digital FPGA synthesis will be conducted using first derivative method using Sobel operator.

III. Experimental FPGA-Camera Testbed

In this paper, we synthesize digital architecture for processing Sobel Algorithm on video source using Altera DE-1 SoC FPGA board. The capturing device is simple low-complex OmniVision OV7670 camera module. The OV7670 is a low voltage CMOS image sensor that provides full functionality of a single-chip VGA camera and image processor in a small footprint package. The OV7670 provides full-frame, sub-sampled or windowed 8-bit images in a wide range of formats, controlled through the Serial Camera Control Bus (SCCB) interface. The module has an image array of 656 x 488 pixels for a total of 320,128 pixels, of which 640 x 480 pixels are active (307,200 pixel). After the Analog Processing block, the raw signal is fed to a 10-bit analog-to-digital (A/D) converter shared by G and BR channels (see Ref. [2]). This A/D converter operates at speeds up to 12 MHz and is fully synchronous to the pixel rate (actual conversion rate is related to the frame rate). For the Altera DE-1 FPGA Development board, interested reader could refer to [5] for full board specification. The crucial specifications of the board, important for this demonstration, are presented in Table I.

TABLE 1
ALTERA DE-1 SOC SPECIFICATIONS

FPGA Chip	Cyclone V (5CSEMA5F31C6)
Memory Controllers	2 Hard Memory Controllers
Embedded memory	4,450 Kbits
FPGA Memory	64MB (32Mx16) SDRAM on FPGA
Display Port	24-bit VGA DAC
Video Input	TV Decoder (NTSC/PAL/SECAM)
A/D	sample rate: 500 KSPS, Channel number: 8, Resolution: 12 bits, Analog input range: 0 ~ 4.096 V
Power	12V DC input
Clock	Four 50MHz clock sources from the clock generator

There are two possible approaches towards FPGA implementations of the digital architecture for processing Sobel ED algorithm. The first approach is by implementing a processing clock that is 4 times the maximum pixel clock. In this implementation, the FPGA will use up to two clock cycles to process a single pixel. However, such approach is not useful for Sobel ED, but for some other convolution operations that use different kernel coefficient which requires multiplication instead of shifting. This implementation is suitable for low resolutions like VGA at up to 60 fps (24MHz pixel clock and 100MHz processing clock). Second option is to design fully pipelined architecture, enabling the pixel clock serving as the processing clock. Such implementation naturally uses significantly more resources, but it is very frequency-efficient (pixel clock can be higher than 130MHz) and is capable to process HD images.

The Sobel filter, is nothing else, but a 2D spatial high-pass FIR (Finite Impulse Response). FIR filters process the output based on the input history, opposed to an Infinite Impulse Response Filter (IIR) that computes the output based on input history and output history. Considering the Sobel 3x3 filter mask, to generate a single pixel in the output image, access to

9 pixels in the input image is required, then multiply them by 9 values and adding it to the partial result. As 6 out of all 9 components are non-zero (Eq.3.), only 6 multiplications and 5 addition operations are processed. The procedure of multiplication of the kernel values with values from the input to generate a single output is pure convolution operation.

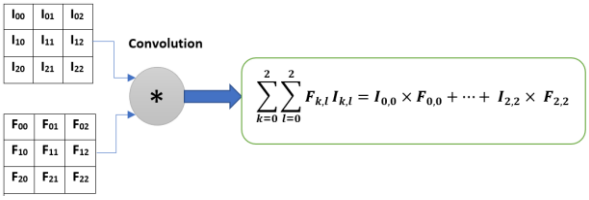


Fig. 3. Convolution of the kernel (F) and input image pixels (I)

To develop a Sobel filter, we practically synthesize two main modules, fully described in VHDL: a 3x3 filter mask module and the Sobel arithmetic module. The used camera does not provide output in 2D stream of pixels, but 1D stream with a system signal indicating when to increment the second dimension (*hsync*). The minimum amount of data to store is two lines of the image and 3 pixels, to realize a 3x3 sliding window in the image [3] (Figure 4).

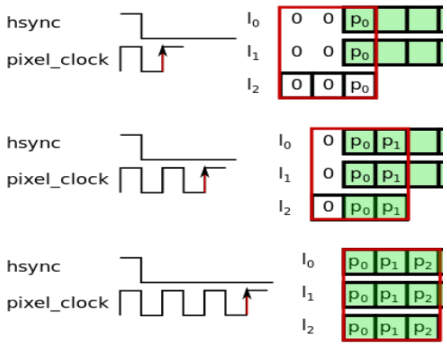


Fig. 4. Realization of the Sobel Sliding Window

IV. Synthesis Results and Conclusion

We realized synthesis of FPGA architecture for Sobel ED on real-time images captured from simple 0.3 Megapixel camera. The examples of the resulting EDs are presented on Fig. 5. Comparing them with the results we got from MATLAB and C++, performing both Sobel and Canny Filters, the quality of the experimental ED is acceptable and very comparable from the aspect of the quality. On a Windows PC equipped with Intel Core I7 Processor with 8 GB of RAM, the Sobel ED in MATLAB took 5-7 seconds on the equal resolution still images from the same scenario, fed from the local hard drive. The FPGA architecture provides real-time almost instantaneous Sobel ED response. We experienced insignificant edge misses in images from more complex scenarios. Table I summarize the FPGA synthesis results. It can be seen that the Sobel kernel and arithmetic architecture fully accommodate on the FPGA chip of Altera DE-1 SoC board occupying only 2% of the adaptive logic modules (ALMs). Each ALM on Altera FPGA board consists of combinational logic, two registers, and two adders [5].

Considering the System on Chip (SoC) nature of the board, where ARM hard processor system exists and interconnection bus architecture is available to connect with the FPGA synthesized ED logic, the HPS-FPGA processing model and FPGA accelerated computing is serious candidate for the future embedded IoT devices with image processing capabilities. The future work is seriously considering involvement of the HPS system on the FPGA board for realization of IMP and VP.

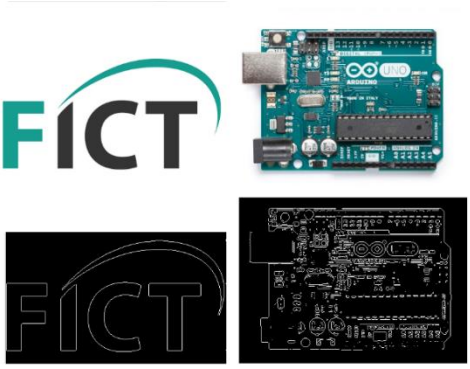


Fig. 5. Results from the real-time Sobel ED taken from the VGA output of the FPGA board connected to standard monitor

TABLE 1
FPGA Synthesis Summary

Family	Cyclone V
Device	5CSEMA5F31C6
Logic utilization (in ALMs)	754 / 32,070 (2%)
Total registers	870
Total pins	94 / 457 (21%)
Total block memory bits	3,152,384 / 4,065,280 (78 %)
Total RAM Blocks	386 / 397 (97 %)
Total PLLs	1 / 6 (17%)

References

[1] Burger, Wilhelm, Mark James Burge, Mark James Burge, and Mark James Burge. *Principles of digital image processing*. London: Springer, 2009

[2] OV760/761 Camera Module Specification. https://www.voti.nl/docs/OV7670.pdf (Accessed April 10,2018)

[3] Gradient Filter Implementation on FPGA, Online Article. https://www.element14.com/community/groups/fpga-group/blog/2015/05/27/gradient-filter-implementation-on-fpga-part2-first-modules. (Accessed April 12, 2018)

[4] A. Ben Amara, E. Pissaloux and M. Atri, "Sobel edge detection system design and integration on an FPGA based HD video streaming architecture," *2016 11th International Design & Test Symposium (IDT)*, Hammamet, 2016, pp. 160-164.

[5] DE1-SoC FPGA Board Specifications: https://www.altera.com/solutions/partners/partner-profile/terasic-inc-/board/de1-soc-board.html. (Accessed April 12, 2018)

Graph-Based Neural Network for Handwritten Digit Recognition

Ivaylo Penev[1] and Milena Karova[2]

Abstract – **The paper presents building and training of a neural network for handwritten digit recognition. As opposed to the known solutions this work uses computation graphs for building, training and estimating the neural network. This approach has two main advantages: reduces the time for network building and training and achieves relative independence of the constructed network model from the runtime environment. The forming of the computation graphs at each step of the neural network building, training and estimating is described. The results from experimental tests with standard data patterns for handwritten digits are discussed.**

Keywords – **Machine learning, Neural networks, Computation graphs, Classification, Recognition, Parallel calculation.**

I. Introduction

Handwritten digit recognition is a famous problem with multiple applications in theory and practice. It is a classification problem, where the input data (images of handwritten digits) are classified to one of a set of classes (digits from 0 to 9).

Neural networks (NN) are classical method for recognition and classification. Many solutions of NN for handwritten digit recognition have been proposed. Many of them achieve accuracy of more than 99.7% (i.e. the error is lower than 0.3%) (e.g. [3], [4], [5]), but the following main disadvantages could be pointed out in them:

- The NN model is complex and therefore hard for scaling;
- The NN model is dependent of the implementation tools (libraries, frameworks);
- The NN training requires much time.

The NN building and training typically consists of lots of independent calculation operations. The natural approach for NN building and training is using parallel calculations via the available computing devices in many systems - multicore central processing units (CPUs) and graphic processors (GPUs). In some literature sources solutions, using GPU are proposed (e.g. [1], [2]). Other sources present the application of the MapReduce method for NN training (e.g. [7]). Although the published results are very good, the implementation is still specific to platforms and libraries. For example the implementation of a NN for a system CPU and GPU needs serious modification for a system with multicore CPU but without GPU support.

[1]Ivaylo Penev is with the Faculty of Computer Sciences and Automation at Technical University of Varna, 1 Studentska str. Varna 9010, Bulgaria, E-mail: ivailo.penev@tu-sofia.bg.

[2]Milena Karova is with the Faculty of Computer Sciences and Automation at Technical University of Varna, 1 Studentska str. Varna 9010, Bulgaria, E-mail: mkarova@ieee.bg.

Recently have been developed tools (programming frameworks and runtime environments), which allow the building and training process of NN to be presented by computation graphs. In graph nodes are placed calculation operations, in the arcs are the data for the operations. The main goals are two - possibility for automatic distribution of the graph between different devices of the computing system and achieving of relative independence of the NN model from the runtime environment (architecture, programming languages and frameworks). Some of the popular frameworks, providing computation graphs, are Theano, Torch. In the last two years has grown the usage of TensorFlow, developed by Google [6].

The present paper describes NN for recognition of handwritten digits from 0 to 9. Building, training and estimating the network are implemented as computation graphs, suitable for parallel execution by different devices. Results from experimental tests of the trained network in a machine learning framework, providing computation graphs and parallel execution of the graphs, are discussed.

II. Neural Network Building and Training by Computation Graphs

A simplified model of a computation graph is shown on Fig. 1. The operations are placed in the nodes and their operands (sets of calculation values) are placed in the arcs. The framework and runtime environment divide the graph into subgraphs and submit each subgraph to a device of the system.

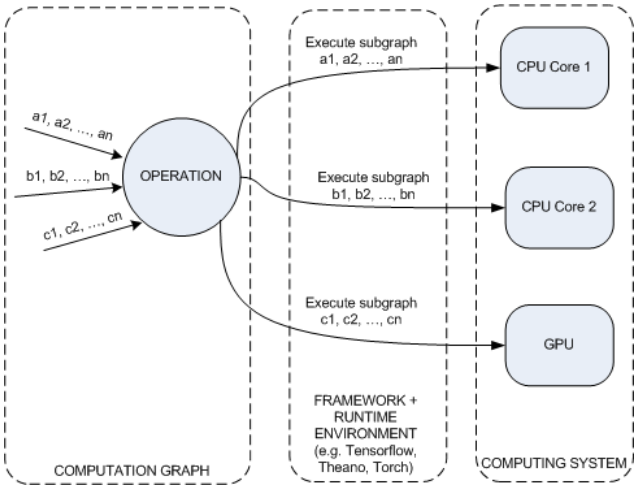


Fig. 1. Simple computation graph, divided into subgraphs

A. Presentation of the input data

For training and testing the NN data sets from the MNIST data base with patterns of handwritten digits are used [8]. A digit image is presented as a two-dimensional matrix. The algorithm processes the digit as a set of 0s and 1s (Fig. 2)

The two-dimensional array is converted to a vector, thus constructing the digit working set, suitable for algorithmic processing:

digit = {x1 x2 ... xn}, where xi = {0,1}, 1 - available pixel in the digit image, 0 - not available pixel in the digit image.

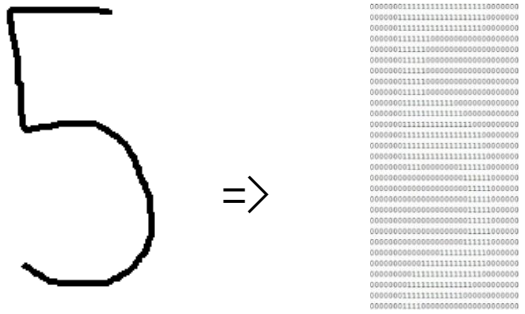


Fig. 2. Presentation of an example digit from MNIST

B. Structure of the neural network

For building the NN model logarithmic softmax regression is used. The process follows the next two steps:

1. Calculation of digit distance to a definite class

The distance is a numeric value, estimating the similarity of a digit with a definite class (from 0 to 9). This value is calculated by a weighted sum of the intensities of the input digit image pixels. If a pixel has high intensity, which does not belong to a definite class, then the pixel weight is negative. Otherwise the pixel weight is positive.

For a class i the distance value is calculated as follows:

$$H_{y'} = -\sum_{i} y_i' \log(y_i) \quad (1)$$

where

W_i - digit weights for a class i

b_i - biases of the digit for the class i

j – index for all pixels weights of digit x.

2. Calculation of probability for belonging of the digit to a definite class

The distance value (Eq. 1) is used for calculation of the probability, showing how likely is the jth digit from the training data set to belong to a class i from 0 to 9. The softmax function is used (Eq. 2):

$$H_{y'} = -\sum_{i} y_i' \log(y_i) \quad (2)$$

The model of the neural network is constructed (Fig. 3).

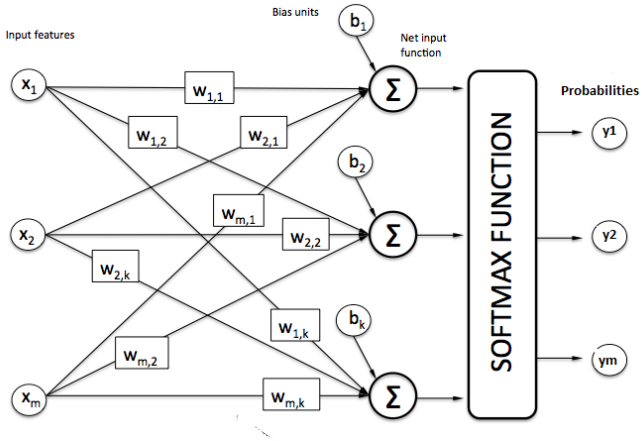


Fig. 3. NN structure for handwritten digit recognition

The multiplication of the sets W and x consists of a lot of independent calculations for each data item from the training data set. These calculations could be performed in parallel. The computation graph of the NN model includes the possibilities for parallel calculation of W and x multiplication (Fig. 4).

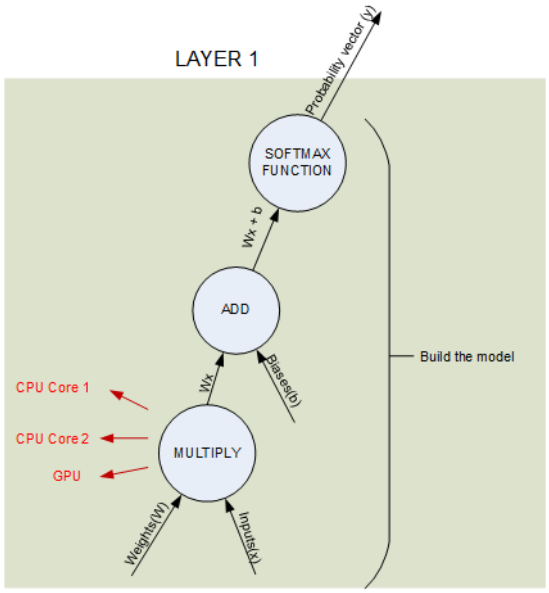


Fig. 4. Computation graph of the NN building

C. Training of the neural network

For estimating the classification error cross-entropy function is used. For each digit from the training data set the error $H_{y'}$ is calculated (Eq. 3)

$$H_{y'} = -\sum_{i} y_i' \log(y_i) \quad (3)$$

where

y_i - the calculated probability for digit to belong to class i,

y' - the true class label of the digit, known from the training data set.

The NN training is performed by backpropagation and stochastic gradient descent algorithms. The purpose is finding the minimum of the $H_{y'}$ error by iteratively changing the values of the weighted parameters W.

The cross-entropy function consists of a set of independent calculations, which could be performed in parallel. At this step of the NN training a new vertex to the computation graph is added (Fig. 5).

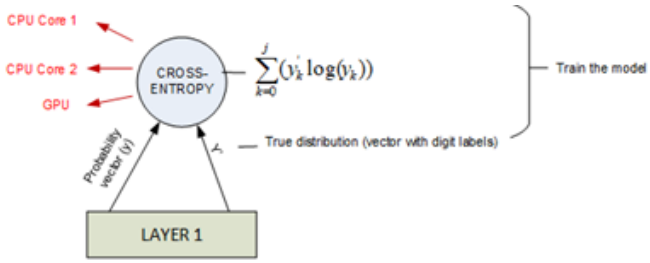


Fig. 5. Computation graph for the NN training

D. Estimation of the neural network accuracy

The accuracy is estimated using the training data sets. Each predicted value y is compared to the corresponding class label y' from the training data set. The result from the comparison is a set of values 1 (TRUE) if y is equal to y' and 0 (FALSE) otherwise. The average value of the equalities is finally determined as $\frac{\sum_j (y == y')}{j}$ for each jth item of the training data set.

For example the result [1 0 1 1] -> [TRUE FALSE TRUE TRUE] means accuracy of 0.75, i.e. 75%.

A new vertex for the accuracy estimation is added to the computation graph (Fig. 6).

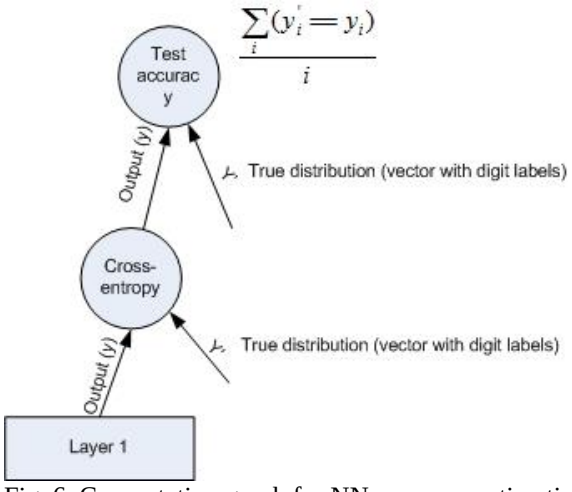


Fig. 6. Computation graph for NN accuracy estimation

III. Experimental Results

A. Framework for the tests

The process of building, training and estimating the NN, described in the previous sections, is implemented by the TensorFlow framework, which provides Python-based programming tools and runtime environment for machine learning problems. The calculation operations and the data values for them (called tensors) are presented as computation graphs. The runtime environment automatically distributes the graphs to the devices of the system (CPU cores, GPU).

The experiments are performed on two computing systems - computer 1 (with GPU support) and computer 2 (without GPU support).

The TechPowerUp GPU-Z application is used for measuring the video card work in real time. The system parameters are monitored by the Windows task manager.

B. Measuring the NN training time

For measuring the NN training two parameters are used:

number of iterations – how many iterations are performed in the gradient descent algorithm for minimizing the error function;

batch size - number of data sets from MNIST, processed at each iteration of the gradient descent.

The data size for each test is determined as $data\ size = number\ of\ iterations * batch\ size$. For example a test with 500 iterations and 50 MNIST input digits is performed by data size of 500 * 50 = 25000.

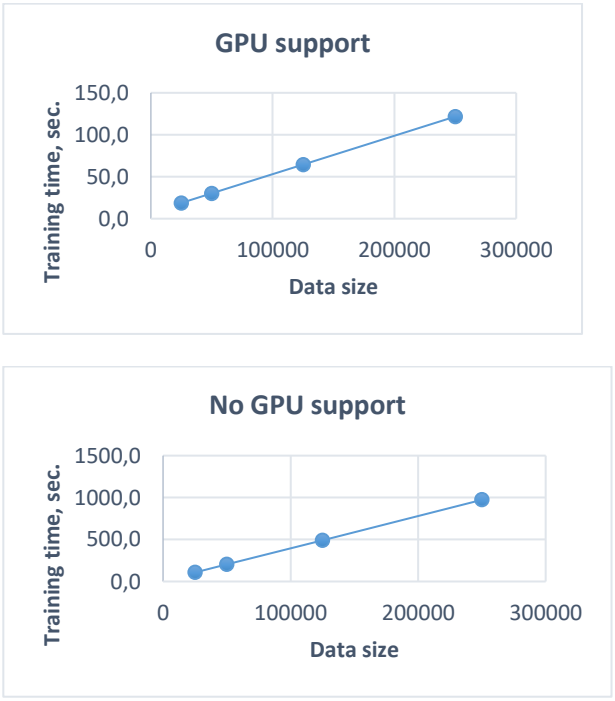


Fig. 7. Training time with 500 iterations and 50 batch size

The results show, that the NN training on a system with CPU and GPU reduces significantly the training time - about 8-10 times in comparison to a system with CPU only (Fig. 7). When increasing the values of the test parameter values the difference between the times decreases, but the distributed computation graph is still calculated 6-8 times faster than the same graph on a CPU only (Fig. 8).

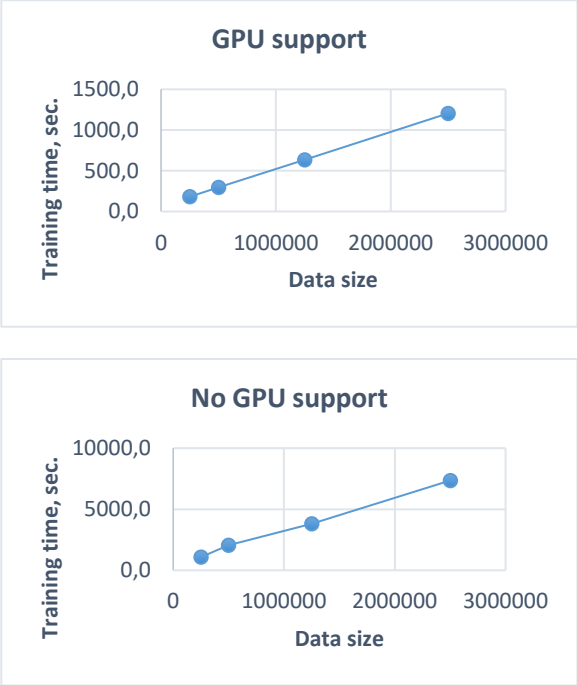


Fig. 8. Training time with 5000 iterations and 250 batch size

C. Measuring the NN training accuracy

The best training accuracy achieved is 99.5% (Fig. 9). This result is close to the best known results for handwritten digit recognition by NN. For example in [1] an accuracy of 99.65% is reported (i.e. error 0.35%), but the NN network is more complex with multiple hidden layers and is performed on a system with many video cards - precondition, which is not available in massive computer systems

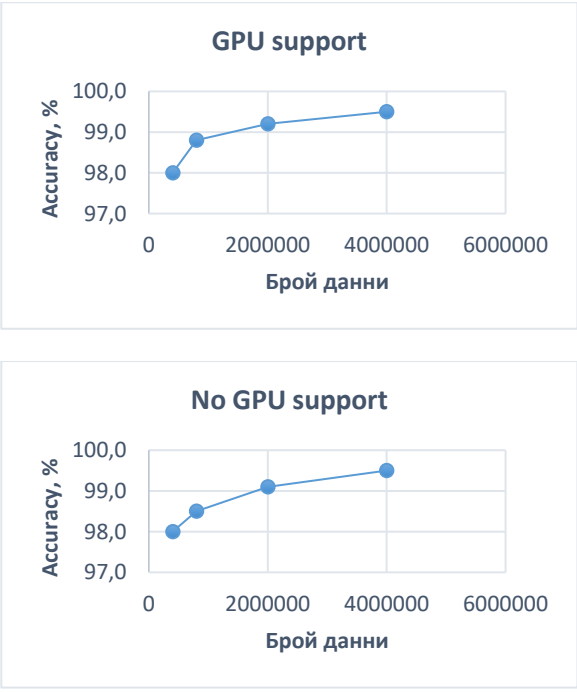


Fig. 9. Training accuracy with 8000 iterations and 500 batch size

IV. Conclusion

The following conclusions could be derived from the presented results:

- The building and training of neural network could be significantly accelerated by the usage of proper computation graphs, performed on systems with GPU;
- The presented approach could be used for other problems, because the process of NN building and training typically includes the same (or very similar) steps;
- NN, presented by computation graphs, could be adapted for implementation on other frameworks and environments, which provide distribution of the graphs to different system devices (e.g. frameworks as Theano or Torch).

The future work will be directed to application of computation graphs for various problems by other machine learning algorithms (e.g. kNN, SVM, etc.), where lots of independent calculation operations are typically performed.

References

[1] D. Ciresan, Meier, U., Gambardella, L., Schmidhuber, J. Deep Big Simple Neural Nets Excel on Handwritten Digit Recognition, Neural Computation, Vol. 22, MIT, USA, 2010.

[2] D. Cireşan, Meier, U., Schmidhuber, J. Multi-column Deep Neural Networks for Image Classification, Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference, DOI: 10.1109/CVPR.2012.6248110, IEEE, 2012.

[3] I. Sato, Nishimura, H., Yokoi, K. APAC: Augmented Pattern Classification with Neural Networks, https://arxiv.org/abs/1505.03229, 2015.

[4] J.-R. Chang, Chen Y.-S. Batch-normalized Maxout Network in Network, https://arxiv.org/abs/1511.02583, 2015.

[5] L. Chen-Yu, Gallagher, P., Zhuowen, T. Generalizing Pooling Functions in Convolutional Neural Networks: Mixed, Gated, and Tree, https://arxiv.org/abs/1509.08985, 2015.

[6] M. Abadi, Agarwal A., et. al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems, arXiv:1603.04467v2, 2016.

[7] N. Basit, Zhang Y., Wu H., Liu H., Bin J., He Y., Hendawi A. MapReduce-based deep learning with handwritten digit recognition case study, Big Data (Big Data), 2016 IEEE International Conference, DOI: 10.1109/BigData.2016.7840783, IEEE, 2016.

[8] http://yann.lecun.com/exdb/mnist/

Evaluation and improvement of stemmers for Serbian language

Petra Antić

Abstract – **In this paper the existing stemmers for Serbian language were evaluated based on the percentage of incorrect stems they produce. According to the obtained results, new rules were introduced to minimize the errors for three different word types. The evaluation after the improvement has showed that the new rules had the positive effects to the stemmer correctness for all three types.**

Keywords – **Stemmers, Serbian language, Error metric**

I. Introduction

Stemming is a low level task in natural language processing, whose goal is to map different word variations to the same form, called stem, by removing suffixes.

The corpora with pairs of words and their correct stems are not available most of the times, so authors mainly use manual ways for assesing the stemmer correctness. Milošević presented two methods [1]: in the first method, a news article was manually stemmed, and then the produced text was compared with the outcome of the stemmer applied to the same article. The other method used machine stemming as the first step, and then a person was reading the stemmed text. The stems were evaluated based on the possibility for the human reader to conclude the original meaning (no overstemming), and the ability of the stem to cover all morphological variations of its lemma (no understemming).

To the best of the author's knowledge, there are three publicly available stemming algorithms for Serbian and one for Croatian (which can also be applied to Serbian, given the similarities between these two languages). Two of them are by the authors Kešelj and Šipka [2] – the optimal and the greedy stemmer, and the third one is an improved version of the aforementioned greedy algorithm, given by Milošević [3]. They all employ a suffix subsumption approach, while the stemmer for Croatian, by the authors Ljubešić and Pandžić [4], relies on regular expressions.

In order to discover deficiencies of the existing stemmers with a goal to minimize them, but also to decrease human effort needed for this task and obtain more robust results, the author has defined an error metric expected to be applied to a Serbian language lexicon.

The next section covers the lexicon used for conducting the evaluation and the definition of the metric used. Section III presents the results obtained by the evaluation, while in Section IV the improvements and their effects are discussed. Finally, Section V summarizes the findings of this paper and gives directions for the future work.

Petra Antić is with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, e-mail: petra.antic92@gmail.com.

II. Evaluation method

A. Serbian Language Lexicon

Stemmer evaluation was conducted on the Serbian language lexicon, srLex [5]. This is a flective lexicon, where each item is presented as the following 5-tuple: inflectional form, lemma (canonical form), morphosyntactic description, absolute frequency, frequency on million occurences (ženu, žena, Ncfsa, 15838, 0.028556). An inflectional form represents morphological variation of a lemma, such as a case for nouns or a tense for verbs. Morphosyntactic descriptions are given by MULTEXT-East Morphosyntactic Specifications for Serbian, Version 5 [6]. This lexicon contains 108,829 different lemmas, and 5,326,726 inflectional forms of these lemmas.

Before applying stemmers on the words from the lexicon, preprocessing was performed with a goal to reduce the amount of data and to remove the irrelevant data. The first step was the removal of the words whose occurrence frequency is less than 0.00002, as well as interpunction signs. This has reduced the number of items in the lexicon more than 10 times, which has enabled evaluation to be performed in reasonable time. Finally, only unique words were extracted from the first column, because stemmers do not differentiate between the different cases or the same word forms obtained from different word types, and proper nouns were removed, which gave a final number of 247,518 words for the stemmers' application and evaluation.

B. Stemmer Error Metric

The goal of stemming is to reduce a word to its base, which should be the same for all morphological variations of that word. This further means that a stem obtained from a word in a morphological form is expected to be the same as a stem obtained from that word's lemma. Since the used lexicon contains pairs of lemmas and their variations, it is possible to easily compare stems of these two words. If there is a difference between these stems, it can be concluded that there is an error in stemming.

Based on this idea, the following metric was defined – the percentage of items in the lexicon where a mismatch between stem of lemma and stem of its morphological variation exists. It is important to note that, while this metric can be used to locate the errors in stemming, it can not define overall stemmer correctness – even if the obtained stems are the same, it doesn't necessarily means they are correct. However, in most cases, if all of the variations are giving the same stem,

and there is no understemming, the stem can be accepted as correct.

The evaluation and the improvements were applied to Java implementation of the stemmers from the SCStemmers package [7], presented by Batanović [8].

III. Stemmer evaluation

With the aim to locate word types which are the most error prone, this metric was firstly calculated using all four stemmers on the main word types, shown in Table I.

Word type with the most errors are found to be pronouns, for which each stemmer made an error for almost a half of the word pairs. For adjectives, verbs and numbers, at least one stemmer gave acceptable results, however, it was decided to make further analysis for adjectives and verbs, and check whether some of their subtypes show significant errors in stemming.

The next step was calcuating the metric for subtypes of verbs, adjectives and pronouns. The idea was to extract the subtypes for which all of the stemmers make error of 0.5 or more, and use those subtypes as candidates for further analysis and improvement. Given the large number of possible subtypes, only the extracted candidates are shown in Tables II-IV. The rows in all tables in this paper are labeled with morphosyntactic descriptions used in srLex lexicon [6].

Table I
Percentage of different stems for word and lemma for main word types

	KSG	KSO	M	LjP
P	**0.61**	**0.59**	**0.59**	**0.48**
Q	0.12	0.12	0.12	0.12
A	**0.46**	**0.26**	**0.57**	**0.33**
R	0.28	0.17	0.23	0.18
C	0.02	0.02	0	0.02
S	0.07	0.09	0.06	0.06
V	**0.2**	**0.15**	**0.4**	**0.39**
M	0.62	0.45	0.52	0.17
N	0.21	0.18	0.33	0.07

A. Prounouns

For pronouns, two the most error prone groups are demonstrative (Pd) and indefinite pronouns (Pi), as well as some forms of personal (Pp), interrogative (Pq), reflexive (Px) and possessive (Ps) pronouns, with metric results shown in Table II.

The first problem is with short demonstrative pronouns (*taj*, *ovaj*), where two types of errors are present: either the word ending is completely removed, leaving the stem of only one letter (*toga-t*), or no change was made at all.

The other problem is related to demonstrative pronouns of quality, such as *takav/onakav/ovakav*, which show presence of a voice change of absent A, and are stemmed to base *takv* for every case form except nominative, which gets stemmed to *takav*.

Table II
Percentage of different stems for word and lemma for pronouns

	KSG	KSO	M	LjP
Pd	0.73	0.72	0.79	0.66
Pi	0.62	0.61	0.51	0.5
Pp2	0.69	0.69	0.5	0.69
Pq-	0.63	0.63	0.63	0.5
Px-n	0.79	0.5	0.79	0.5
Ps2mp	0.9	0.5	0.9	0.6

An additional problem are volatile vowels [8] which can be present in some case forms at the end of the word: *takvog*(*a*). Their presence or absence doesn't change a word in a grammatical or a semantical way, and its usage depends only on the writing style.

Interogative pronouns show errors similar to demonstrative ones, with short forms being stemmed to only one letter (*koga – k*), but also errors caused by different bases for nominative and other cases (*šta – čega*). This difference in base is also a cause of errors for personal pronouns (*ti – tebe, on-njemu*).

B. Verbs

Table III shows errors for verb subtypes, and it can be seen that all of them are the forms of auxiliary verbs (Va), which, because of their irregular nature, have a mismatch for almost all cases.

Table III
Percentage of different stems for word and lemma for auxiliary verbs

	KSG	KSO	M	LjP
Vam	1	1	1	1
Var	0.93	0.93	0.57	0.7
Vae	0.5	0.5	0.5	1

Different authors had different approaches regarding auxiliary verbs. For example, the stemmer by authors Kešelj and Šipka does not have special rules for this word type, so there are stemming examples such as *biti-b*. Milošević uses a dictionary of mappings for processing these verbs, where every form of an auxiliary verb is mapped to its infitive form. However, this dictionary does not cover all the possible forms, so there are still mismatches such as *budite-bud* and *biću-biti*. Finally, the authors Ljubešić and Pandžić have the third approach – they consider auxiliary verbs to be stopwords and ignore them during stemming. Still, there are also incosistencies in this approach, because word *bude* gets stemmed to form *bud*.

C. Adjectives

For adjectives, subtypes with the highest rate of errors are comparative (Agc) and superlative (Ags) forms, with the

results shown in Table IV. High values for superlative are expected, given that none of the stemmers are removing prefixes and prefix *naj-* always remains in the stem.

TABLE IV
PERCENTAGE OF DIFFERENT STEMS FOR WORD AND LEMMA FOR ADJECTIVES

	KSG	KSO	M	LjP
Agc	0.75	0.83	0.93	0.62
Ags	0.99	0.99	0.99	1

The analysis of the cases where a mismatch for a comparative exists, shown that one of the problems is the inconsistency in stemming – for some words, suffix for the comparative is removed, while for the others it is not. Then, all of the stemmers are making a mistake when iotation voice change occurs (*mlad– mlađi*), where only *-i* of the suffix *-ji* is removed, while stem keeps the changed consonant (*đ*). Another problem is also caused by a voice change, in this case the absent A, when the letter A is missing in the stem of a case form (*ružn*), while it is present in the stem of lemma (*ružan*). For examples such as *spontaniji – spotanij* only a partial removal of suffix *–iji* can be noticed.

IV. STEMMER IMPROVEMENTS

All four stemmers have shown advantages and disadvantages, so it was not obvious based only on the results which stemmer should have been chosen for the improvements. However, by comparing their implementations, the one by Ljubešić and Pandžić has proven to be the most appropriate for the expansion. Its biggest advantage is that, by using regular expressions, suffix removal is dependent on the letter combination which precedes it. Also, it is structured in a way that each rule pattern covers more suffixes, so it is easier to deduct which type of word a rule covers. In addition, the number of the rules is the smallest – it has 70 rules, opposed to 285 in the stemmer by Milošević, or even 1000 and over 17000 rules in the greedy and the optimal version of stemmer by Kešelj and Šipka.

One of the proposed improvements was to add a new structure for post-transformations, which will solve problems with different bases of an inflection and a lemma, or problems with base transformations due to iotation changes. It was implemented as a two level hash map, where keys on the first level are the regular expressions for the word start, and keys on the second level are the regular expressions of the word end. The value obtained based on these two keys is post transformation which should be applied.

In addition to this new structure, new rules were also introduced for all word types classified as the most error prone.

A. New Rules

Since almost all pronoun subtypes showed problems with stemming, and a number of existing pronouns is limited and many are different only in prefixes (*kakav – ikakav – nekakav*), it was decided to define special rules for all pronouns. These rules would be placed on the beginning of the list, to be applied before more general rules.

Personal pronouns were stemmed in a way to keep the gender (*njemu – on, njoj – ona*), while other, adjective-like pronoun types, were stemmed to the male gender. Using this approach, twenty new rules were added.

For defining rules for verbs, the author has used rules for building simple verb forms. Using the rules instead of the dictionary was chosen for its structure and smaller posibility to exclude some of the forms. By Klajn's recommendation [9], and in complience with lemmas in the lexicon, forms of the verb *jesam* were reduced to the base *biti*. Using this aproach, six new rules were added.

Rules for adjectives included four irregular verbs stated in the grammar books (*dobar, zao, mali, veliki*), as well as three adjectives known to be the only ones with suffix *–ši* (*lep, lak, mek*). In addition, a rule was created to try to solve the problem of iotation in the adjective's base in the comparative form. This rule included the usage of post tranformation to transfrom the changed letter of the base (*đ* in *mlađ*) to its original letter (*d*). Another rule was added to ensure the consistency in removing suffix *-iji* and its inflections.

B. Results of Improvement

TABLE V
PERCENTAGE OF DIFFERENT STEMS FOR WORD AND LEMMA FOR PRONOUNS AFTER THE IMPROVEMENTS

	LjP	A
P	0.48	0.06
Pp	0.61	0.15
Pq	0.31	0.19
Ps	0.18	0
Pd	0.66	0.02
Px	0.41	0
Pi	0.5	0.1

After introducing the new rules, the error metric was recalculated for the relevant word subtypes, and the results are shown in Tables V-VII.

There is a significant decrease in errors for all subtypes of pronouns, which can be seen in Table V. The errors that are still present are mostly dependent on:

- The lexicon structure – there are words which are not common in Serbian, such as *tko* instead of *ko*; personal pronouns are given in the lexicon with lemma in the male gender, and an assumption was made that for these pronouns the gender should be preserved.
- The overlapping forms for a different gender and number (e.g. *one* can be genitive, 2nd person, singular, female gender, or acusative, 3rd person, plural, male gender).

Improvements are noticeable also for auxiliary verbs (Table VI), but the problems similar to those with the pronouns still

remain: the lexicon contains some iecavic forms which are not covered by the rules, and there is also overlapping with other word types (*je* can be shortened version of verb *jesam*, but also a shortened version of the personal pronoun *ona* in acusative singular form).

TABLE VI
PERCENTAGE OF DIFFERENT STEMS FOR WORD AND LEMMA FOR AUXILIARY VERBS AFTER THE IMPROVEMENTS

	LjP	A
Vam	1	0
Var	0.7	0.13
Vae	1	0.5

For comparative forms of adjectives there is only a slight decrease in errors, which can be seen in Table VII. The reasons for those results are various:

- Some adjectives do not form a comparative of its full positive form, but from shortened from (*dub-ok – dub-lji*). In this cases, because of *-ok* and similar suffixes not consistently covered by the rules, there is still a mismatch even if comparative is correctly stemmed.
- The absent A is still an unsolved problem, given that it is very hard to differentiate between situations where it is present and where it is not (e.g. *smotan – smotaniji* and *verovatan – verovatniji*: by simply looking at the suffixes and letters which are preceding the suffix, a difference can not be noticed, so this type of a problem should be approached in a more sofisticated way).
- Some of the comparative forms are already covered by a rule for female nouns ending with the suffix *-ija* (e.g. *galija*), so that suffix is not removed for them. Again, a more sofisticated approach would be needed to differentiate between these two cases.

TABLE VII
PERCENTAGE OF DIFFERENT STEMS FOR WORD AND LEMMA FOR ADJECTIVES AFTER THE IMPROVEMENTS

	LjP	A
Agc	0.62	0.56
Ags	1	1

The final recalculation of the error metric was conducted for all word types, in order to evaluate the impact of the new rules on the whole word set. The results are given in Table VIII, where we have excluded, for the clarity, word types without a changed result of the error metric.

The sigificant improvement can be noticed for pronouns in general. For adjectives and verbs, the improvements are minor, since only the specific word types were targeted – however, it can be concluded that the additional errors were not created by the new rules. Interesting side effect can be seen for adverbs, where a small decrease in error metric was caused due to the equivalent forms for the adverbs and the adjectives in the neutral gender. Only the nouns have suffered the increase of the error metric, most probably due to the introduced iotation rules for adjectives.

TABLE VIII
PERCENTAGE OF DIFFERENT STEMS FOR WORD AND LEMMA FOR MAIN WORD TYPES AFTER THE IMPROVEMENTS

	LjP	A
Total	0.238557	0.236956
P	0.478708	0.057269
A	0.334851	0.329862
R	0.175959	0.174632
V	0.395296	0.391382
N	0.066764	0.074437

V. CONCLUSION

This paper has presented the common errors of existing stemmers for Serbian language and the attempt to minimize those errors. For particulat word types there were noticeable improvements, and for others there are remaining problems that need attention. However, the efficiency of the improved stemmer was not evaluated in a real application. One of the following steps would be to apply this stemmer in a task such as information extraction or sentiment analysis, and verify if the introduced changes are showing better results for that task.

REFERENCES

[1] N. Milošević, "Sentiment Analysis of Sentences in Serbian language", Master's Degree Thesis. School of Electrical Engineering, University of Belgrade, Belgrade, Serbia, 2012. ("Mašinska analiza sentimenta rečenica na srpskom jeziku")

[2] V. Kešelj and D. Šipka, "A Suffix Subsumption-Based Approach to Building Stemmers and Lemmatizers for Highly Inflectional Languages with Sparse Resources," INFOtheca, vol. 9, no. 1–2, pp. 23a–33a, 2008.

[3] N. Milošević, "Stemmer for Serbian language." arXiv 1209.4471, 2012.

[4] N. Ljubešić, D. Boras, and O. Kubelka, "Retrieving Information in Croatian: Building a Simple and Efficient Rule-Based Stemmer," in INFuture2007: Digital Information and Heritage, Zagreb, Croatia: Department for Information Sciences, Faculty of Humanities and Social Sciences, 2007, pp. 313–320.

[5] N. Ljubešić, F. Klubička, Ž. Agić, and I.-P. Jazbec, "New Inflectional Lexicons and Training Corpora for Improved Morphosyntactic Annotation of Croatian and Serbian," 10th International Conference on Language Resources and Evaluation (LREC 2016), Conference Proceedings, pp. 4264–4270, Portorož, Slovenia, 2016.

[6] MULTEXT-East Morphosyntactic Specifications, Version 5. http://nl.ijs.si/ME/V5/msd/html/msd-hr.html#msd.R-hr

[7] SCStemmers – GitHub repository. https://vukbatanovic.github.io/SCStemmers/

[8] V. Batanović, B. Nikolić, and M. Milosavljević, "Reliable Baselines for Sentiment Analysis in Resource-Limited Languages: The Serbian Movie Review Dataset," 10th International Conference on Language Resources and Evaluation (LREC 2016), Conference Proceedings, pp. 2688–2696, Portorož, Slovenia, 2016.

[9] I. Klajn. *Serbian language Grammar*. Beograd, Zavod za učenike i nastavna sredstva, 2005. (*Gramatika srpskog jezika)*

Extreme learning machines for prediction of the stock market trend

Miloš Stojanović[1], Ivana Marković[2], Jelena Z. Stanković[3] and Stevica Cvetković[4]

Abstract – **In this study we investigate the application of Extreme Learning Machines (ELM) to the problem of stock market trend prediction. We started with feature formation, based on the analysis of technical indicators. Classification is done using recently introduced specific single layer neural networks called Extreme Learning Machines (ELM). Our tests performed on the BELEX15 stock market index have shown competitive results in this area, while model training and testing is executed almost instantly.**

Keywords – **Stock market trend prediction, technical indicators, Extreme learning machines.**

I. INTRODUCTION

Prediction of stock market movements is a challenging task, taking into consideration the fact that the financial market is a complex, evolving and dynamic system whose behavior is pronouncedly non-linear [1].

Although some researchers have suggested that there is evidence that stock prices are not purely random, the general consensus still is that their behavior is approximately close to the random walk process. Therefore, degrees of accuracy of an approximate 60% hit rate in predictions are often considered satisfactory results in this area [2].

Predicting the direction of movement of the price of financial instruments is a current area in academic research where technical analyses in combination with machine learning have proven to be highly effective [3].

The choice of technical indicators substantially affects the overall performance of the classification system. Assuming that investors trade continuously and that past prices provide sufficient information, the most commonly used technical indicators for securities' price trend prediction are trending, volume and oscillating indicators. Trending indicators identify and monitor the securities trends, while the volume indicators are based on the change in the volume of trading in securities and complete the information which is offered by the trending indicators in forming trading strategies. Oscillating indicators are leading indicators which generate early warning signals of changes in the securities trend and determine the strength of the current trend, as well as the moment when a change in the trend occurs.

Since the prediction of the movement of stock market indices plays an important role in the development of effective market trading strategies, it is important to point out that every increase in precision is considered an exceptional contribution since it leads to an increase in the return and the decrease in the risk involved in trading. However, due to illiquidity, any selected indicators, especially on emerging markets such as Belgrade stock exchange, may provide trading signals that cannot be used to form profitable trading strategies. Market liquidity is an important factor for portfolio managers and large institutional investors and it refers to the ability to execute a trade promptly, at low cost or no cost, risk or inconvenience [4]. Therefore, in this study liquidity risk is in particular considered in the calculation of technical indicators.

The second crucial part of the system is for machine learning technique to be applied for stock market trend prediction. In [5] it was indicated that the Least Squares Support Vector Machines (LS-SVMs), and SVMs - Support Vector Machines outperform other machine learning methods, since in theory they do not require any previous *a priori* assumptions regarding data properties. In this study we investigate the application of Extreme Learning Machines (ELM) [6, 7] for stock market trend prediction, as an alternative to the commonly used SVM. ELM is a single hidden layer feed-forward neural network (SLFN), which overcomes an important drawback of traditional artificial neural networks (ANNs) - their slow learning speed. It increases training speed by randomly assigning weights and biases in the hidden layer, instead of iteratively adjusting its parameters by gradient based methods. As well as minimizing training error, ELM finds smallest norm of output weights and hence has a better generalization performance than gradient based training algorithms, such as backpropagation.

In the rest of the paper we first describe the technical indicators selection that will be used as features for the prediction model. The problem of predicting the direction of the stock index movements is then modeled as a problem of a binary classification, after that we give an overview of ELM for classification. Finally, the experimental evaluation and conclusion are presented.

[1]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

[2]Ivana Marković is with the Faculty of Economics at the University of Nis, Trg Kralja Aleksandra Ujedinitelja 11, Niš, Serbia, E-mail: ivana.markovic@eknfak.ni.ac.rs.

[3]Jelena Z. Stanković is with the Faculty of Economics at the University of Nis, Trg Kralja Aleksandra Ujedinitelja 11, 18000 Niš, Serbia, E-mail:jelenas@eknfak.ni.ac.rs

[4]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

II. Technical Indicator Analysis and Feature Construction

Feature construction is an essential step for defining an accurate prediction model. The arbitrary application of a large number of explanatory features to ELM or any other machine learning based algorithm could lead to low prediction accuracy. A proper feature selection procedure on the other hand would lead to higher method accuracy. This process is of great importance, but there is no general rule that can be followed.

This study represents a continuation of our prior work [8, 9] with new results and a more in-depth analysis.

Firstly, in order to adjust the technical analysis to the specifics of trading in the observed emerging market, technical indicators are calculated using liquidity risk adjusted returns:

$LAr_t = r_t - c_t$, r_t – *logarithmic return*, c_t – *illiquidity cost.*

The usual approach for liquidity risk modeling in emerging and frontier markets uses proxies. Since detailed transaction data on bid-ask spreads are not available, we employ Amihud's illiquidity measure [10] that can be calculated using daily data on price and trading volumes:

$c_t = \frac{|r_t|}{V_t}$, V_t – *trading volume in ten millions of monetary units.*

Considering the fact that returns are negatively correlated with illiquidity [10], illiquidity cost decreases returns (Fig. 1). Involving liquidity risk that is significantly high in emerging markets [11] should provide higher accuracy of technical indicators reflected in the decrease in the number of trades according to the realistic conditions on the market.

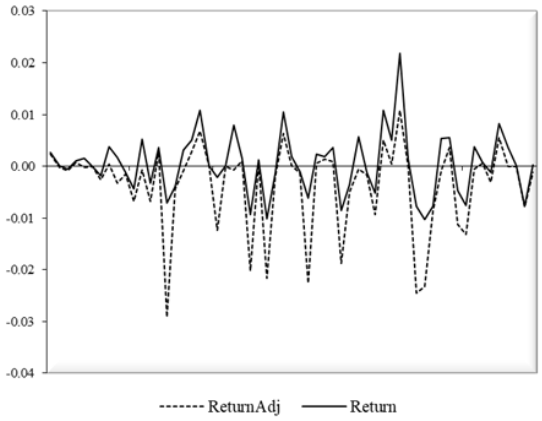


Fig. 1. The relation between the logarithmic return (Return) and liquidity risk adjusted return (ReturnAdj) on the BELEX15 index

In our prior studies different technical indicators were analyzed based on their common use in the literature [1-3]. Briefly, it can be summarized in the following way. Considering that the response variable predicts the stock market trend (either an increase or decrease), the explanatory features need to measure changes as well. In effect, observing feature changes over time is more significant for prediction than the absolute value of each feature. In this study a more extended analysis is conducted regarding modeling stock market tendencies with comprehensive indicators which are able to more accurately model dynamic changes on the stock market.

In the group of technical indicators of oscillations, which are used to discover short-term overbought or oversold conditions, in this study we introduce the TRIX indicator. In order to remove the noise and reveal the market trend for a specific time interval we have facilitated a smoothing process using the TRIX indicator.

The Triple Exponential Moving Average Oscillator (TRIX) is a technical indicator that shows the slope of the security's return trend line. A rising trend line indicates an uptrend and in that case the TRIX takes positive value. On the other hand, if the TRIX indicator has negative value, it can be concluded that there is downtrend in the concrete security's return, while crossing the signal line indicates a trend-change. Table I shows a selected list of the technical indicators.

TABLE I
THE LIST OF TECHNICAL INDICATORS

Indicators	Formula
Closing price	CP_t, t= 1,2, ... N
Lowest price	LP_N – Lowest price in the past N days
Highest price	HP_N – Highest price in the past N days
Logarithmic return	$r_t = \log CP_t - \log CP_{t-1}$
Liquidity risk adjusted return	$LAr_t = r_t - c_t$
Trend indicators	
Exponential Moving Average	$EMA_N = LAr_t * k + EMA_{t-1} * (1-k)$; $K = 2/(N+1)$
Oscillating indicators	
Relative Strength Index	RSI = $100-(100/(1+\frac{\frac{1}{T}\sum_{t=o}^{T} LAr_t^+}{\frac{1}{T}\sum_{t=o}^{T} LAr_t^-}))$
Parabolic "Stop and Reverse" Indicator	$SAR_t = SAR_{t-1} + AF(ELAr_{t-1} - SAR_{t-1})$ AF – acceleration factor (from 0.02 to 0.2), $ELAr_{t-1}$ – extreme return in the previous period
Moving Average Convergence Divergence	$MACD_t = EMA_{12} - EMA_{26}$ Signal Line = Simple 9-day moving average of MACD
Stochastic Oscillator	%K = 100*((LAr_{close}–LAr_{14})/ ($LAr_{14,high}$ – $LAr_{14,low}$)) %D = EMA_3(%K) Slow %D = EMA_3(%D)
Triple Exponential Moving Average Oscillator	$TRIX = (EMA3_t - EMA3_{t-1}) / EMA3_{t-1}$
Commodity Channel Index	$CCI = (TLAr_t - MA_{20}) / (0.015 * MD)$ $TLAr_t$=(LAr_{high}+LAr_{low}+LAr_{close})/3 $MD = \sum_{t=1}^{N} (TLAr_t - MA_{20}) / N$
Williams' R Indicator	%R = - (LAr_{high} – LAr_{close})/ (LAr_{high} – LAr_{low})*100

For all technical indicators, liquidity risk is considered in the calculation of their values.

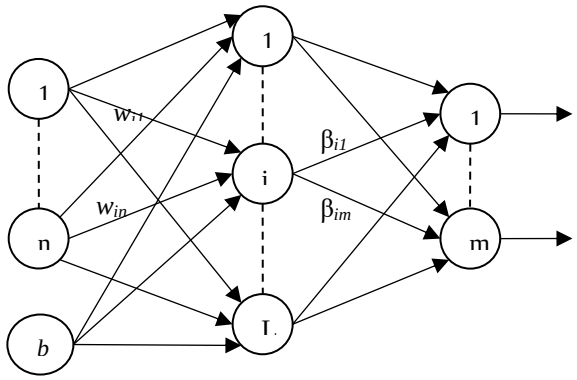
Finally, the trend is modeled as a categorical variable used to indicate the movement direction of the BELEX15 index over time t. If the liquidity risk adjusted return over time t is larger than zero, the indicator is 1. Otherwise, the indicator is −1.

III. Extreme Learning Machines (ELM)

Let us define N training examples as $(\mathbf{x}_j, \mathbf{y}_j)$ where $\mathbf{x}_j = [x_{j1}, x_{j2}, ..., x_{jn}]^T \in \mathbf{R}^n$ denotes the *j-th* training instance of dimension n and $\mathbf{y}_j = [y_{j1}, y_{j2}, ..., y_{jm}]^T \in \mathbf{R}^m$ represents the *j-th* training label of dimension m, where m is the number of classes. The set of features, that is previously explained – technical indicators, will further be denoted as $\mathbf{x}_j$, while $\mathbf{y}_j$ will denote m dimensional vector of binary class labels with value "*1*" denoting membership to the class. SLFN with an activation function $g(x)$ and L hidden neurons could be defined as:

$$\sum_{i=1}^{L} \beta_i g(\mathbf{w}_i \cdot \mathbf{x}_j + b_i) = \mathbf{f}_j, \; j = 1, \ldots, N \tag{1}$$

where $\mathbf{w}_i = [w_{i1}, w_{i2}, ..., w_{in}]^T$ denotes the vector of weights which connects the i^{th} hidden neuron and all input neurons, $\beta_i = [\beta_{i1}, \beta_{i2}, ..., \beta_{im}]^T$ is the weight vector which connects the i^{th} hidden neuron and all output neurons, and b_i is the bias of the i^{th} hidden neuron. In 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 Fig 2.



n input neurons | *L* hidden neurons | *m* output neurons

Fig. 2. Structure of the ELM network

SLFN in (1) should satisfy $\sum_{i=1}^{L} \|\mathbf{f}_i - \mathbf{y}_i\| = 0$, i.e., there exist β_i, $\mathbf{w}_i$ and b_i such that:

$$\sum_{i=1}^{L} \beta_i g(\mathbf{w}_i \cdot \mathbf{x}_j + b_i) = \mathbf{y}_j, \; j = 1, \ldots, N \tag{2}$$

If we denote as $\mathbf{H}$ a hidden layer output matrix of the ELM; the i^{th} column of $\mathbf{H}$ represents the i^{th} hidden neuron's output vector regard to inputs $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N$.

$$\mathbf{H} = \begin{bmatrix} g(\mathbf{w}_1 \cdot \mathbf{x}_1 + b_1) & \cdots & g(\mathbf{w}_L \cdot \mathbf{x}_1 + b_L) \\ \vdots & \cdots & \vdots \\ g(\mathbf{w}_1 \cdot \mathbf{x}_N + b_1) & \cdots & g(\mathbf{w}_L \cdot \mathbf{x}_N + b_L) \end{bmatrix}_{N \times L} \tag{3}$$

and

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times m} \quad \text{and} \quad \mathbf{Y} = \begin{bmatrix} \mathbf{y}_1^T \\ \vdots \\ \mathbf{y}_N^T \end{bmatrix}_{N \times m} \tag{4}$$

Then the equivalent matrix form of (2) can be represented as:

$$\mathbf{H}\beta = \mathbf{Y} \tag{5}$$

The output weights are then computed by finding the unique smallest norm least-squares solution of the linear system (5) as:

$$\beta = \mathbf{H}^{\dagger}\mathbf{Y} \tag{6}$$

where $\mathbf{H}^{\dagger}$ represents the Moore-Penrose generalized inverse of $\mathbf{H}$.

For a given training set $T=\{(\mathbf{x}_j, \mathbf{y}_j)\} | \; \mathbf{x}_j \in \mathbf{R}^n, \mathbf{y}_j \in \mathbf{R}^m, j = 1, ..., N\}$ with N instances of n-dimensional descriptors, the sigmoid activation function $g(x)$, and a hidden number of neurons L, the ELM algorithm for classification problems can be summarized as follows:

Training:

(a) Assign random input weights $\mathbf{w}_i$, and biases b_i, $i = 1,..., L$.
(b) Compute the hidden layer output matrix $\mathbf{H}$ using (3).
(c) Compute the output weights β using (6).

Testing:

(a) Compute the hidden layer output matrix $\mathbf{H}_{test}$ for instances from the test set, using (3)
(b) Compute the output matrix $\mathbf{Y}_{test}$ according to (5) using the β obtained in step 3 of the training.
(c) For every row in $\mathbf{Y}_{test}$ (i.e. every test instance), compute a class label as the index of the maximal value in that row.

IV. Experimental evaluation

To test the proposed method for stock trend prediction, we used data taken from the website of the Belgrade Stock Exchange (www.belex.rs). The available data were divided into two groups. The first group consisted of 2485 records required for the training model, from February 24, 2006 to December 31, 2015. For the second group of data, from January 4, 2016 to December 29, 2016, a total of 252 days of trading were selected.

As a general measure for the evaluation of the prediction, the Hit Ratio (HR) is used, which was calculated on the basis of the number of properly classified results within the test group:

$$HR = \frac{1}{m}\sum_{i=1}^{m} PO_i \;\text{ for }\; PO_i = \begin{cases} 1 & PV_i = AV_i \\ 0 & PV_i \neq AV_i \end{cases} \tag{7}$$

where PO is the prediction output of the *i* trading day, AV_i is the actual value for the *i* training day and PI_i is the predicted value for the *i* trading day and *m* is the number of data in the test group.

For the tests, we implemented ELM in MATLAB and used it to measure the classification hit rate and speed. Hit rates are measured on both the training and test set, with training and test times, measured in seconds on an *Intel Core i5* computer. The results are presented in table 2, where the data in the first column represent the number of neurons in a hidden layer. The sigmoid function is used as the activation.

TABLE II
A COMPARISON OF THE MODELS

Number of neurons	HR Training Set	HR Test Set	Training time (s)	Test time (s)
10	65 %	62 %	0.001	0.001
25	66 %	62 %	0.09	0.001
50	66 %	61 %	0.07	0.001
100	66 %	61 %	0.1	0.001
500	72 %	59 %	0.6	0.06
1000	77 %	58 %	1.5	0.07
10000	97 %	58 %	217	0.3

From table 2 we can note that HR on the training and test set remains stable for the number of neurons in a hidden layer in the range from 10 to 100. When increasing the number of neurons further, HR on test set increases significantly, while HR on test set decreases slightly, which implies possible overfitting. We can observe that increasing the number of neurons above a certain value does not improve classification results. Nevertheless, the obtained values of the hit rates for all ELMs, regardless of the number of neurons in the hidden layer are within the expected range of precision and comparable to the results obtained in other studies [1], [2], [12].

Training time for all 2458 instances in the training set is only 0.1 seconds with 100 neurons in a hidden layer, while classification for all 252 instances is done instantly (< 1ms). With an increase in the number of neurons of up to 1000, these times slightly increase. Only with 10000 neurons is training time increased significantly, while test time increases slightly (< 1s). These results demonstrate high performances in terms of training and test speed on this dataset.

In order to compare the results of the ELM with other common classification techniques, we measured the accuracy of the Linear SVM and kernelized RBF SVM [13], on the same dataset. Both Linear SMV and RBF SVM reached approximately a 62 % hit rate. Thus, it can be noted that ELM reaches results comparable with Linear SVM, as well as to the kernelized SVM while operating significantly faster during the training and testing. Regarding the results obtained using the random walk model (RW) as a benchmark that used current value to predict the future value, assuming that the latter in the following period (y_{t+1}) will be equal to the current value (y_t), ELM significantly outperformed the obtained results of 51.19%.

V. CONCLUSION

In this paper we presented the results of our research in the field of stock market trend prediction. A standard set of technical indicators is used as features in combination with a fast and powerful ELM classifier. A hit rate of around 60% is the expected range in this area. It can be concluded that a combination of technical indicators with an ELM classifier is reasonable choice for stock trend prediction applications. The ELM classifier could be used as an alternative to the commonly used SVM. In the future, we plan to investigate the performance of other types of features combined with an ELM classifier, particularly integration of trading strategies.

REFERENCES

[1] W. Huang; Y. Nakamori and SY. Wang, "Forecasting stock market movement direction with support vector machine," Computers & Operations Research vol. 32, no. 10, pp. 2513–2522, 2005.

[2] S. Lahmiri, "A Comparison of PNN and SVM for Stock Market Trend Prediction using Economic and Technical Information", International Journal of Computer Applications, vol. 29, no.3, 2011.

[3] V. Chsherbakov, „ Efficiency of Use of Technical Analysis: Evidences from Russian Stock Market", Ekonomika a management , vol. 4, 2010.

[4] K. Dowd, "Measuring Market Risk", Chichester: John Wiley & Sons Ltd, 2002.

[5] 5 O. Phichhang and H. Wang, "Prediction of Stock Market Index Movement by Ten Data Mining Techniques" Modern Applied Science, vol. 3, no. 12, pp. 28-42, 2009.

[6] 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.

[7] 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.

[8] I. Marković, J. Stanković, M. Stojanović. M. Božić, "Prediction of the stock market trend using LS-SVMs based on technical indicators", presented at the 49th Int. Conf. ICEST 2014, Nis, Serbia, June 25-27, 2014.

[9] M. Stojanovć, S. Cvetković, G. Stančić, "Poređenje metoda mašinskog učenja za predikciju trenda promene finansijskih vremenskih serija", Simpozijum INFOTEH Jahorina 2017, pp 398-401, 2017.

[10] Y. Amihud, "Illiquidity and stock returns: cross-section and time-series effects", Journal of Financial Markets, vol. 5, no. 1, pp. 31–56, 2002.

[11] J.Z. Stankovic and E. Petrovic, "Liquidity Risk Implications for Market Risk Assessment in Emerging Markets", *Journal of Contemporary Economic and Business Issues*, vol. 5, no. 1, pp. 5-23, 2018.

[12] L. Yuling, H. Guo and J. Hu, "An SVM-based Approach for Stock Market Trend Prediction", Neural Networks (IJCNN), IEEE Press, pp. 1-7, 2013.

[13] 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.

Performance of VoIP Services with Integrated Analog Peripherals

Dragan Martović[1], Zoran Veličković[2], Zoran Milivojević[3], Marko Veličković[4]

Abstract – **In this paper, the performance of VoIP services with integrated analog telephones is determined. The integration of analog phones was carried out with the analog telephone adapter Grandstream HT702. By measuring the value of key QoS parameters of conversational speech realized through the IP PBX AXON switch, the performance of VoIP services is determined. Packet delay, as the most important QoS parameter of conversational speech, was within the given limits. The package loss was also within the given limits, while the mean value of the jitter oscillated around the allowed value. Jitter was successfully compensated by the receiving buffers and had no influence on the conversation speech. Based on the experimental results, it was found that the key QoS parameters defined for the conversation speech were satisfied, which confirms the successful integration of analog telephones into the VoIP service.**

Keywords – ***VoIP, IP PBX, AXON, ATA, QoS*.**

I. INTRODUCTION

The great expansion and distribution of computer networks made it a very suitable medium for the transmission of various content. In addition to the exchange of textual documents in electronic form, computer networks have become the basic medium for the distribution of multimedia content [1]. Regardless of the fact that they are not designed for this purpose, with increasing network flow and the emergence of modern network communication protocols this service has become possible [2]. In order to offer satisfactory resources to different network applications, it is necessary to classify them according to their needs. Standardization of network applications has been carried out by international standardization bodies such as IETF (*Internet Engineering Task Force*) and ITU (*International Telecommunication Union*). From the point of view of users, network applications can be divided into those tolerant to packet errors (multimedia applications), and those tolerant to packet delay (Web search or e-commerce applications) [3]. Multimedia network applications are further classified into three basic classes. Network applications that share stored audio or video content in real time are classified in the first class. The second class includes applications that support conversation and video telephony. Network applications that allow streaming of audio and video content streaming are classified in the third class. This application class is similar to classic radio and television diffusion. The basic problem with all applications for distributing multimedia content through computer networks is the realization of the appropriate QoS (*Quality of Service*) [4]. The origin of this problem is in the concept of computer networks based on TCP / IP protocols that do not guarantee QoS, but only "best effort" QoS services are provided. This means that there is no guarantee that the package will be delivered at all, and if it is delivered, its timeliness and order are not guaranteed. For each of the classes of multimedia network applications, specific QoS parameters are defined. The key standardized QoS parameters of multimedia network applications are packet delay, packet delay variation, and packet loss. The realization of conversational speech through computer networks, or the Internet, is called Internet telephony or VoIP (*Voice over Internet Protocol*) [5]. From a user perspective, VoIP is similar to a traditional analog telephone service that is realized by circuit switching. Analog peripheral devices characteristic of a traditional phone service are still present in homes and small business environments. In order to take advantage of the existing analog telephone service infrastructure, technology for the integration of analog telephone devices in VoIP has been developed. Multimedia applications belong to the class of network applications that are tolerant of errors, but they are extremely intolerant to the package delay. Thus, for a conversation speech, the delay of 150 ms is allowed, but the delay of 150 ms to 400 ms can be tolerated. The acceptable variation of the delay is 1 ms, and the allowed PLR (*Packet Lost Rate*) is 3% [4].

A. State of Art

Getting the certain QoS in IP network is crucial and is usually achieved by setting up access and service policies [6]. Realization of QoS in IP VPN networks must provide the recommended lower limits for satisfying QoS requests at the network node level [7]. Enabling multilayer QoS in VoIP calls requires speed adjustment and resource management to integrate SIP and QoS mechanisms to wireless networks [8]. Different queueing policies have been considered to achieve the end-to-end QoS requirements for different types of network trafic [9]. In this paper, an analog phone is integrated into the digital IP PBX (*Internet Protocol Private Branch Exchange*) [10]. The integration of the analog phone into the IP PBX provides, among other things, additional features to an analog peripheral device such as voicemail, call waiting, call hold, personalization of ring tones, and so on. Integration of analog phones into IP PBX requires a set of time-consuming signal and protocol conversions so that QoS satisfaction can be critical.

[1]Dragan Martović is with the College of Applied Technical Sciences of Niš, A. Medvedeva 20, 18000 Niš, Serbia, E-mail: martovic.dragan@gmail.com

[2]Zoran Veličković is with the College of Applied Technical Sciences of Niš, A. Medvedeva 20, 18000 Niš, Serbia, E-mail: zoran.velickovic@vtsnis.edu.rs.

[3]Zoran Milivojević is with the College of Applied Technical Sciences of Niš, A. Medvedeva 20, 18000 Niš, Serbia, E-mail: zoran.milivojevic@vtsnis.edu.rs.

[4]Marko Veličković is with the College of Applied Technical Sciences of Niš, A. Medvedeva 20, 18000 Niš, Serbia, E-mail: marko.velickovic.rsni@yahoo.com

This paper discusses the performance of VoIP services in which analog peripherals ATA (*Analog Telephone Adapter*) are integrated. In the experimental part of the work, the Grandstream HT702 ATA device was used [11], which allows the analog telephone to be connected to the digital IP network. The Grandstream HT702 conducts the conversion of one communication protocol (analog) to another (digital) and vice versa. In the terminology of the ITU-T ATA, the Grandstream HT702 is a gateway that provides two-way communication in real time. The measurement of the parameters of the VoIP application was carried out in the AXON IP PBX realized in the local computer network. The AXON IP PBX central software phones are directly integrated through the network interface, while the integration of analog phones uses the Grandstream HT702 ATA device. The key QoS parameters for the conversational speech connections made between all terminal devices are measured. For this work, it is important to analyze the QoS parameters lured by communication between software IP phones and analog telephone devices. Using the open-source *Wireshark* software package, measurements were made and the results obtained were presented.

The second section describes the need to integrate analog telephone devices into the IP PBX central. A basic overview of the basic characteristics of the AXON central is given, its web interface is described and the basic characteristics of the used ATA Grandstream HT702 are shown. The third section describes the test environment and shows the obtained QoS results. The experimental results are compared with the target values of QoS and certain performance of VoIP services within the AXON IP PBX. In the fourth section, appropriate conclusions were drawn based on the experiments carried out and certain recommendations were given.

II. INTEGRATION OF ANALOG TERMINAL DEVICES IN IP PBX

The integration of analog terminal devices - analog phones or FAX devices - provides several business benefits to the IP PBX network platform. The basic benefit of integrating multiple communication networks into one is to reduce costs. Combining multiple business communications networks into one convergent network infrastructure greatly reduces capital costs, cabling costs, and operational costs associated with managing separate network infrastructures. New business opportunities by combining network data with voice calls are obtained. In this way, an integrated information exchange system is formed which saves time and enables more efficient decision making. The telecommunications industry has many years of experience in IP voice integration in traditional analog networks. It is estimated that organizations save 30% to 40% on costs associated with long distance communications [12]. In addition, significant savings are realized on administrative costs and costs related to maintenance of equipment. Integrated network services use a single network architecture to simplify system maintenance and upgrading. From the standpoint of modern users, the characteristics of analog terminal devices are very limited. In order to improve the functionality of analog terminal devices - standard analog telephone devices, and at the same time utilize the existing network infrastructure, they can

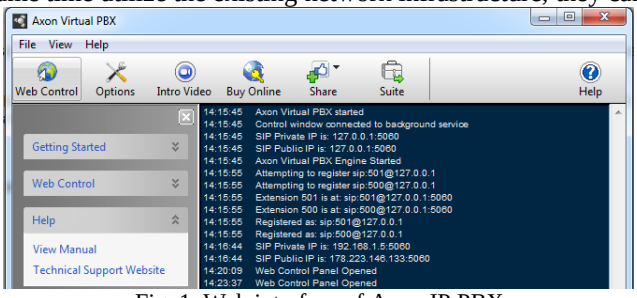


Fig. 1. Web interface of Axon IP PBX

be integrated into digital PBX telephone exchanges. In the telecommunications industry, the term IP PBX is a private telephone exchange using a TCP/IP protocol for managing telephone calls through a computer network.

The integration of the analog phone into the IP PBX provides additional features to an analog terminal device that is implicit in digital exchanges. Some standard IP phone functions cannot be easily implemented, while others can not be implemented at all because of the hardware limitations of the analog phones. IP PBX enables multimedia communication as well as the communication of sending simultaneous messages between two or more participants over the Internet or other compatible networks. They are mostly used by small and medium-sized enterprises, but they can also be found in large multinational corporations. IP PBX can exist as a physical hardware or just as a virtual software realization. In addition, it is possible to connect to the public telephone network PSTN.

In this paper the IP PBX of the company "*NCH Software*" was used [13]. Axon is a virtual IP PBX software designed to manage phone calls in a business environment or at a call center. AXON IP PBX is installed on a PC and supports up to 64 phone lines, as well as an unlimited number of extensions. AXON IP PBX works as a fully equipped telephone exchange that connects telephone lines and extensions using state-of-the-art VoIP technology. AXON offers all the usual features of a traditional PBX as well as routing all calls within the company. For the setting and controlling of the AXON switch, the Web interface shown in Fig. 1 is realized.

B. Analog Telephone Adapter Grandstream HT702

In order to integrate analog terminal devices into the digital IP network, the Grandstream HT702 analog telephone adapter (*Analog Telephone Adapter*) (Fig. 2) is used in this work. Analog telephone adapters are devices for connecting traditional analog phones, fax machines and similar user devices to a digital telephone system, or to a VoIP telephone network. An analog telephone adapter's digital interface usually

Fig. 2. Interfaces of ATA HT702 for connection from one side to the LAN, and on the other side, to analog terminal devices.

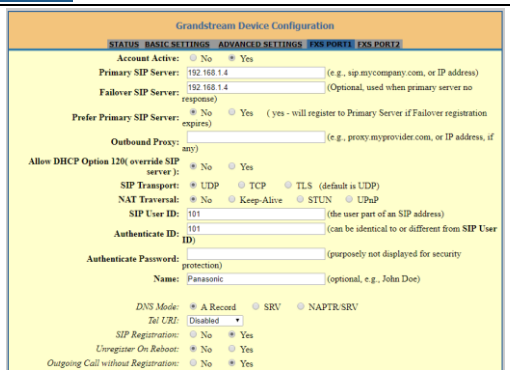


Fig. 3. Example of configuring the ATA HT702 FXS port 1

contains an Ethernet port for connecting to an IP LAN (Fig. 2). ATA communicates with the server (in our case with the AXON IP PBX) using one of the protocols H.323, SIP, MGCP (*Media Gateway Control Protocol*), SCCP (*Signaling Connection Control Part*) or IAX (*Inter Asterisk eXchange*). When an ATA is used to integrate an analog telephone, it encodes / decodes audio signals using one of the standard voice codecs such as G.711, G.729, GSM (*Global System for Mobile Communications*) and iLBC (*Internet Low Bitrate Codec*). The HT702 supports SIP signaling protocols for the VoIP network used in the experimental work. The ATA Grandstream HT702 is fully compatible with the industry-standard SIP standard and works with all compatible devices. Some of the features of the ATA HT702 are comprehensive support in the form of a callout for outgoing calls, G.168 echo cancellation, VAD (*Voice Activation Detection*), CNG (*Comfort Noise Generation*), hiding packet loss PLC (*Packet Loss Concealment*), etc. The HT702 has two RJ-11 FXS (*External eXchange Station*) ports for connecting to analog terminal devices such as analog phones or FAX machines (Fig. 2). ATA communicates directly with AXON IP PBX, and does not require additional hardware or any additional software. The Web interface for setting and managing ATA and a SIP environment is shown in Fig. 3.

III. EXPERIMENTAL ENVIRONMENTS

The experimental part consists in establishing telephone connections between software and integrated analog phones in the LAN through the AXON IP PBX [14]. A LAN is formed around a router that connects the wired and wireless part of the network as shown in Fig. 4. Measured QoS parameters for all performed experiments are shown in Tab I. QoS parameters are named according to [4]. The open source software "Wireshark" was used to measure the QoS parameters of realised VoIP traffic on the network. The DHCP (Dynamic Host Configuration Protocol) server is activated on the router so that all network components are assigned the appropriate IP addresses. Most available ATAs on the market have similar characteristics. The HT 702 model of Grandstream was available to us. Grandstream HT702 is assigned an IP address through which it can communicate with other network components. An analog telephone device is connected to one FXS port of the ATA HT702 device while the analog cordless telephone is connected to another. On FXS ports can be connected both analog dial phones and analog phones with

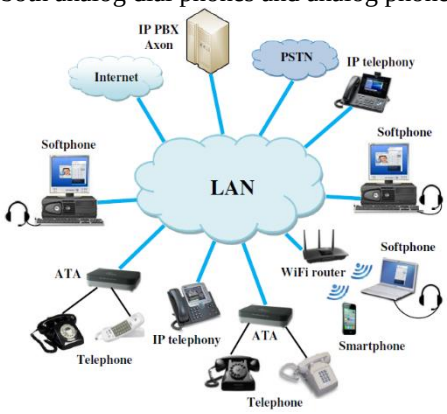


Fig. 4. Axon IP PBX with integrated analog telephones by ATA Grandstream HT702

DTMF dialing. An analog FAX machine can be connected to one of the FXS ports. All computers in the network are running Windows, and an AXON IP PBX telephone exchange is installed on one PC. All analog and software telephones from a local computer network are registered at the AXON telephone exchange before making telephone connections as required by the SIP protocol. NCH Software's Express Talk software phones have been installed on all network scanners so each of them can be used as a terminal device. In fact, each one of them can start and accept the call. Also, from each analog phone, it can be established, that is, accept a telephone connection with software phones. Fig. 4 shows a local computer network where the experimental part of the work is performed. Fig. 5 shows the interaction between the analog and the software telephone during the establishment, duration, and termination of the connection to the already known - received IP address (the phase of receiving the caller's IP address is not shown in Fig. 5). The provision of caller IP addresses is the task of the SIP protocol, which is used in the experimental part of this paper. In this case, the AXON IP PBX has the role of a SIP proxy that provides the current IP addresses of the users. Before any connection is established, it is necessary to perform SIP registration on AXON IP PBX of all software or integrated analog phones. Before any connection is established, it is necessary to perform SIP registration on AXON IP PBX of all software or integrated analog phones.

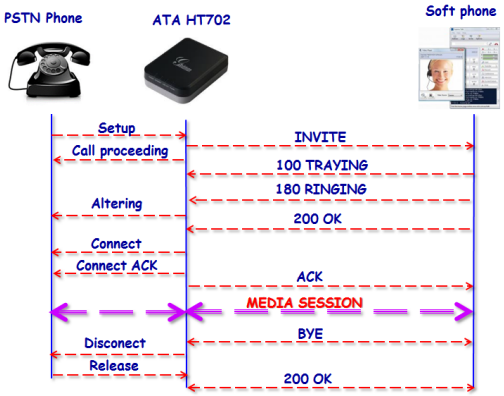


Fig. 5. Establishing a call to the known IP address between the analog phones integrated with the ATA H702 in the VoIP system.

TABLE I
EXPERIMENTAL RESULTS OF KEY QOS PARAMETERS

No	MEDIUM	APPLICATION	DEGREE OF SYMMETRY	KEY PERFORMANCE PARAMETERS AND TARGET VALUES						
				ONE-WAY DELAY (ms)		DELAY VARIATION (ms)			INFORMATION LOSS (%)	
				MEASURED	TARGET	MAX	MEAN	TARGET	MEASURED	TARGET
1	Audio	Convers. voice	Two-way	30.649	< 150	1.481	0.232	< 1	0	< 3
				129.319		7.135	1.852		0	
2	Audio	Convers. voice	Two-way	129.440		8.291	1.521		0	
				40.057		6.019	0.802		0	
3	Audio	Convers. voice	Two-way	55.775		5.537	1.057		0.2	
				40.796		5.012	1.158		0	
4	Audio	Convers. voice	Two-way	47.453		6.947	2.470		0	
				64.366		5.284	0.977		0.6	
5	Audio	Convers. voice	Two-way	26.501		1.286	0.176		0	
				139.625		11.742	1.357		0	
6	Audio	Convers. voice	Two-way	131.087		7.254	1.232		0	
				26.462		0.997	0.126		0	

After a successful SIP registration, connections can be established. Tab. I shows the results of key QoS parameters for six realized telephone connections. All connections have been made on one side of analog phones, and on the other side of software phones from LAN. In all experiments, it was a two-way conversation that formed two audio streams. Connections 1, 3, and 5 were initiated by analog phones integrated into VoIP with ATA Grandstream HT702, and accepted by software phones installed on computers from LAN. Software phones initiated the other three connections, and the integrated analog phones were accepted. Measured key QoS parameters that relate to one-way delay and loss of packets in all cases meet the prescribed ranges in [3] and [4]. This standard prescribed that the delay in one direction must not exceed 150 ms, which is also satisfied in all connections. The allowed loss of the package is 3%, so by insight into Tab. 1 can determine that this QoS parameter is satisfied. The variation of the delay (maximum and average) is allowed to be less than 1 ms. In some connections; considerably larger peaks are measured, while the mean value of this QoS parameter oscillates around the allowed. It should be noted that the variation of the delay did not significantly affect the quality of the speech, so it did not influence the conversation. The variation of the delay is simply compensated by the use of the receiving buffer.

IV. CONCLUSION

In this work, the integration of analog telephone devices into the VoIP system was made using the ATA Grandstream HT702. Checking of the realized QoS parameters of VoIP services of integrated analog phones was performed in the AXON IP PBX. Several bi-directional communication links between software phones and integrated analog phones were realized and key QoS parameters were measured. Based on experimental results, it was found that the key QoS parameters defined for the conversational speech were met. Packet delay, the most important QoS parameter for the conversion speech application, was within the given limits. Also, the loss of the package was within the given limits, while the variation of the case-by-case variation oscillated around the allowed value. The receiving buffers compensated the value of the jitter so that it did not have an effect on the conversation speech. Experimental results show that successful integration of analog terminal devices into a VoIP system based on the AXON IP PBX has been performed.

REFERENCES

[1] H. Sinnreich, A. Johnston, "*Internet Communications Using SIP: Delivering VoIP and Multimedia Services with Session Initiation Protocol*", Wiley Publishing, Inc., 2006.

[2] M. Jevtović, Z. Veličković, "Komunikacioni protokoli prepletenih slojeva", Akademska Misao, Beograd, 2013.

[3] ETSI TR 102 805-1, User Group; End-to-end QoS management at the Network Interfaces; Part 1: User's E2E QoS - Analysis of the NGN interfaces (user case), 2009.

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

[5] K. Wallace, "Implementing Cisco Unified Communications Voice over IP and QoS (CVOICE) Foundation Learning Guide", Cisco Press, 2011.

[6] A. Corte, S. Sicari, "Assessed quality of service and voice and data integration: A case study", Computer Communications, Vol. 29, Iss. 11, pp. 1992-2003, 2006.

[7] M. Elezi, B. Raufi, "Conception of Virtual Private Networks Using IPsec Suite of Protocols, Comparative Analysis of Distributed Database Queries Using Different IPsec Modes of Encryption", Soc. and Beh. Sci., vol.195, pp. 1938-1948. 2015.

[8] J. Chen, L. Lee, Y. C. Tseng, "Integrating SIP and IEEE 802.11e to support handoff and multi-grade QoS for VoIP-over-WLAN applications", Comp. Net. vol. 55(8), pp. 1719-1734, 2011.

[9] N. E. Rikli, S. Almogari, "Efficient priority schemes for the provision of end-to-end quality of service for multimedia traffic over MPLS VPN networks", Computer and Information Sciences pp. 89–98, 2013.

[10] G. Kavitha, K. Kumari, S. Kumar, "Internet Protocol Private Branch Exchange", Int. Jour. Rec. Trends in Eng. vol. 1(2), 2009.

[11] Grandstream Networks, Inc., "User Manual: Analog Telephone Adaptors HT701/HT702/HT704 FXS Port", 2017.

[12] Cisco White Paper, "Network Technology Integration Drives Business Success", available at: https://www.cisco.com/

[13] http://www.nch.com.au/pbx/index.html.

[14] J. Jocić, Z. Veličković, "Measurement QoS Parameters of VoIP Codecs as a Function of the Network Traffic Level", ICIST 2015 Proc. Vol. 2, pp. 422-426, 2015.

Inquiry algorithm and Message transmission algorithm in VANET

Tsvetan Marinov[1], Maria Nenova[2], Georgi Iliev[3]

Abstract – **Wireless communications is developing very fast in the last few years. Vehicular Ad-Hoc Networks (VANET) is also wireless network to exchange information among cars, buses and trucks. VANET uses moving cars as nodes in a network to create a mobile network. VANET turns every participating car into a wireless router, allowing cars of each other to connect and create a network with a wide range. It can provide wide variety of services such as Intelligent Transportation System (ITS) e.g. safety applications.**

The research algorithms and protocols in VANET are important and necessary for ITS. This paper describes Inquiry algorithm and Message transmission algorithm. The main idea is to propose algorithm which spread information about congestions, accidents, bad roads, dangerous weather and so on. All of this will make our life safer and easier.

Keywords – **VANET, Inquiry algorithm, Message transmission algorithm**

I. INTRODUCTION

The problem with road traffic has existed for a long time. There are many incidents with vehicle in all over the world. Every year numbers of dead cases increase resulting in 1.3 million people died. By developing the Intelligent Transport System (ITS) this number can decrease. VANET is a future network which promises to decide this problem. Providing the necessary connectivity and access to network services in this case requires the construction of a mobile ad-hoc network. Mobile ad-hoc network VANET consists of devices, which are in condition to self-configuration. Mobile stations can move randomly. They have dynamic topology. This network can be independent or connected to others networks like Internet.

One of the purposes of VANET is transferring data with high speed and minimizing the delay in a communication range. In a VANET, communication occurs from vehicle to vehicle (V2V), vehicle to infrastructure (V2I) and vehicle to broadband (V2B). When nodes are moving and if we have information for them, we can predict the future position on the road. This information could enable us to compute the traffic and in this way to decrease congestions, accidents and harmful air emissions.

[1]Tsvetan Marinov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1756, Bulgaria, E-mail: ts_marinov@abv.bg

[2]Maria Nenova is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1756, Bulgaria, E-mail: mariamnenova@gmail.com

[3]Georgi Iliev is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1756, Bulgaria, E-mail: gli@tu-sofia.bg

Given the distinct features of VANET, different problems are tackled by the researchers. The main part of research works has focused on algorithms.

Additionally, characteristics like the lack of fixed infrastructure and the fact that the mobile stations are not only routers, but are also routing packets to other stations on their way to the final recipient (the so-called multihop routing), are unique to this type of network and they hide new research topics such as network configuration, neighbor's detection, topology support, addressing and routing [1].

The paper aim is to present algorithm for message transmission. This is a very important topic for the future network – VANET.

II. VANET ARCHITECTURE

This part describes the system architecture of vehicular ad hoc networks. The architecture integrates features of traditional ad hoc networking technologies and VANET technologies.

As shown in Figure 1, the in-vehicle domain is composed of an on-board unit (OBU) and one or multiple application units (AUs). The connections between them are usually wired and wireless. However, the ad hoc domain is composed of vehicles equipped with OBUs and roadside units (RSUs). An OBU can be seen as a mobile node of an ad hoc network and RSU is a static node likewise. An RSU can be connected to the Internet via the gateway; RSUs can communicate with each other directly or via multihop as well. There are two types of infrastructure domain access, RSUs and hot spots (HSs). OBUs may communicate with Internet via RSUs or HSs. In the absence of RSUs and HSs, OBUs can also communicate with each other by using cellular radio networks (GSM, GPRS, UMTS, WiMAX, and 4G) [2], [3], [4].

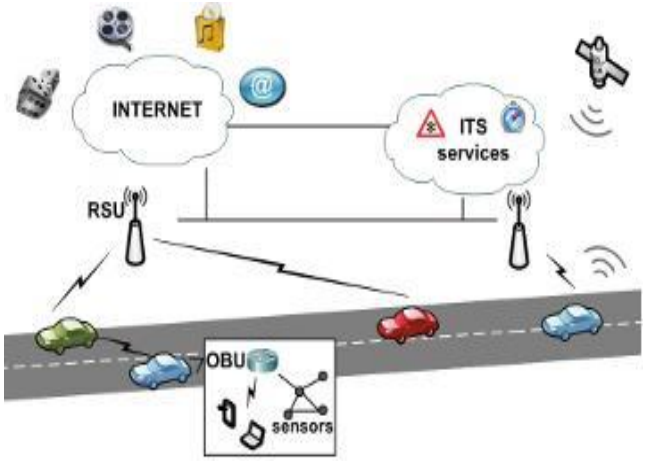


Figure 1. Network Architecture and communication types in VANET

Communication types in VANETs can be categorized into four types. The category is closely related to VANETs components as described above. Figure 1 describes the key functions of each communication type [5]. In-vehicle communication, which is more and more necessary and important in VANETs research, refers to the in vehicle domain. In-vehicle communication system can detect a vehicle's performance and especially driver's fatigue and drowsiness, which is critical for driver and public safety. Vehicle-to-vehicle (V2V) communication can provide a data exchange platform for the drivers to share information and warning messages, so as to expand driver assistance. Vehicle-to-road infrastructure (V2I) communication is another useful research field in VANETs. V2I communication enables real-time traffic/weather updates for drivers and provides environmental sensing and monitoring. Vehicle-to-broadband cloud (V2B) communication means that vehicles may communicate via wireless broadband mechanisms such as 3G/4G. The broadband cloud may include more traffic information and monitoring data as well as infotainment, this type of communication will be useful for active driver assistance and vehicle tracking [6].

III. ALGORITHMS DESIGN

It is assumed that the congestion has occurred and the cars that have come into it are aware of its presence. In the figure 2 those are the dark-moving automobiles moving to the left. The cars on the opposite side of the road are moving and one of them, whose turn has arrived, sends a request message. In the message it includes its location and moving direction. The message is received by all other vehicles in its transmission range. Vehicles that are moving in the same direction as the requesting vehicle will not transmit request packages for a certain period. Vehicles from the opposite lane that have registered events begin to respond one at a time. In the case in question, the vehicles are stuck in traffic and therefore the incoming car will receive traffic jams messages. Then it will make a decision about the authenticity of the registered event based on the number of responses received and the number of lanes in the opposite direction of the road. Depending on that, it does or does not generate a message indicating that there is a traffic jam. Inaccurate information that could be generated by the not completely precise method used by the vehicles is filtered at this step of process.

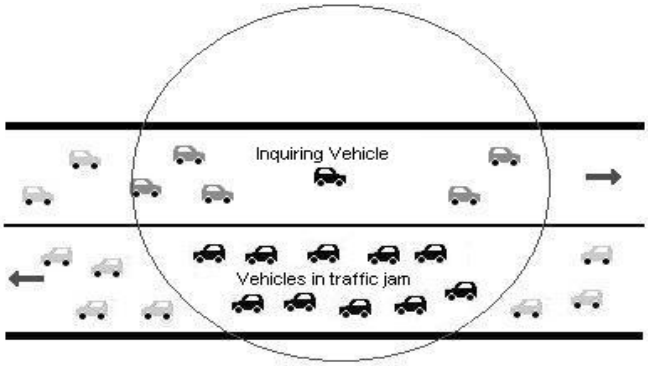


Figure 2. Process of inquiry

In case that the congestion is confirmed a message with the following structure is generated:

- •Current location of the vehicle transmitting the message
- • Time that the event was registered
- • Location of the traffic jam (road lane)
- • Direction of the event
- • Status field (Active, Passive)
- • Event code
- • Time To Live (TTL)
- • Other data

Most of the field names are self-explanatory. Current location is used by other vehicles to determine whether they are before the vehicle that has transmitted the message and therefore start transmitting the message as well. The next three fields contain data about the event that can be used to update old event data the other vehicles have. Status and TTL fields will be discussed in more details later on. The Event code field is used to differentiate traffic jam from other possible events. Since only the case of congestion has been addressed, to allow for future changes there is an option to add more fields.

There are two methods for exchanges of messages. They are directly dependent on the type of message. In the first (Figure 3 (a)) the station that wants to transmit, first advertises the data by sending only part of the message. If the station receives a request, then sends the entire message. The first approach is more appropriate if we have large messages with relatively small meta data. Since this is not the case in the present case, the second approach is chosen.

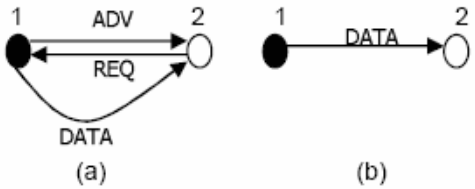


Figure 3. Methods of exchanging messages

For the case it is assumed that the congestion has been registered. The interrogating car generates a congestion message, which among other fields, marks itself as active in the status field. The generated message is then sent to other nearby vehicles that stop the queries and copy the message. These cars mark themselves as passive to this message and continue the normal process of asking the other cars.

The time each vehicle makes a query is set to s2 and is a system parameter. Determining an appropriate value for s2 is a compromise between frequent bandwidth and early detection of congestion. The distribution process is carried out by periodically broadcasting the message using a broadcast address.

Once a new message is generated and active and passive vehicles are determinate, it remains to be seen how the transmitted and received of the message to other cars will be. The status field, as already mentioned above, shows the status of the vehicle for the specific message. A certain vehicle may be active for one message and passive for another. The message is distributed only by active cars. The distribution

period is s3, where s3 is a system parameter and its choice is a compromise between this message being received more than once by some cars (with less s3) and there are cars that do not receive it at all (with a larger s3).

The reason for having passive cars is to increase the reliable transmission of the message in the event of a drop in the active station and to extend the area of the vehicles receiving the message. However, since passive cars do not distribute the message, we have fewer exchanged messages. The following rows look at when a car is marked as active when it passes into a passive state and when it erases the message.

The options for a car to go into active status for a message are three. The first one was already mentioned, the car that generates the message marks itself as active. The second is when a vehicle that has previously been in a passive state stops "hearing" the message from the active for a period longer than the distribution period s3. Then it is supposed to be dropped and then the station changes the value of the status field to active and begins to spread the message. Thus, after intersection a part of the vehicles start to go on different roads, they will expand the area of distribution of the message. The third possible scenario for a vehicle to become active for a given message is when it receives a message from a vehicle that is further behind in the direction of travel. Then it copied the message and marked itself as active with respect to it.

The possibilities for a car to be marked as passive for a message are two. The first is because the vehicle was within the reach of the interrogator, copied the generated message and marked as passive. The second case is when an active car receives the same message from a car ahead of it in the direction of travel. In this case, it passes into a passive state by previously notifying other passive vehicles that delete the message. Here, however, an unstable situation can occur, leading to a frequent change in the status of the second car, therefore a possible optimization would be the second car to remain active while the distance to the previous one is not less than 70% of the maximum range.

Finally, it is considered when the messages are deleted. A message is destroyed first when its TTL field becomes zero. This controls the distribution area of the message. The second reason to delete a message is when a passive station receives an order for it from the active one. The third reason is when a message with more up-to-date information is received. In this case, it is superfluous to continue to distribute the old message [7], [8], [9].

IV. PSEUDO-CODES AND BLOCK DIAGRAMS OF THE ALGORITHMS

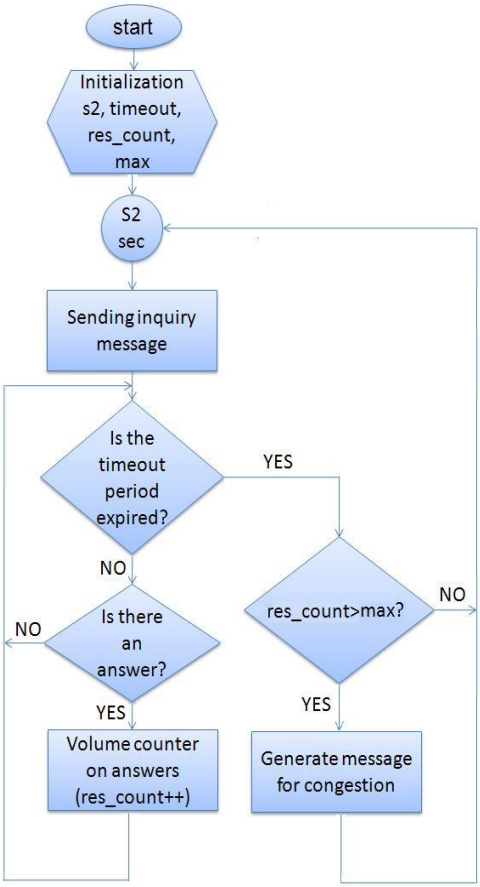
This part presents pseudo-codes and blog diagrams of Inquiry algorithm and Message transmission algorithm (MTA). Thise algorithms can work together with other algorithms. The main idea is to transmission information about congestions, accidents, emergency cases, bad weather, road conditions and more.

A. Pseudo-code and bloc diagrams of Inquiry algorithm

Pseudo-code diagrams of Inquiry algorithm:

1. **start**
2. **initialization:** s2 ,timeout, res_count, max
3. send (enquiry message) to all neighbors;
4. **while** receiving responses **do**
5. responses_count++;
6. **If** responses_count > (max * Number of lanes) **then**
7. generate_message (jam);
8. Sleep (s_2 seconds);

where:
s2 - the time each vehicle makes a query
timeout – this is the period
res_count – this is the response counter
max – parameter



Picture 4. Block diagram of Inquiry algorithm

B. Pseudo-code and block diagram of Message transmission algorithm

Pseudo-code of message transmission algorithm:

1. **start**
2. **initialization:** time to live, s3, offset
3. **If** status of message m is Active **then**

4. **If** new message m1 is received and m1 is new version of m **then**
5. Erase (m);
6. m.status = passive;
7. **Else**
8. send (m) to all neighbors;
9. Sleep (s3 seconds);
10. **Else if** new message m2 is received and m2.status = active **then**
11. Sleep (s3 + random offset);
12. **Else**
13. m.status = active ;
14. Sleep (s3 seconds);
15. **End if**

where:
Time to live (TTL) – message lifetime
s3 - the distribution period

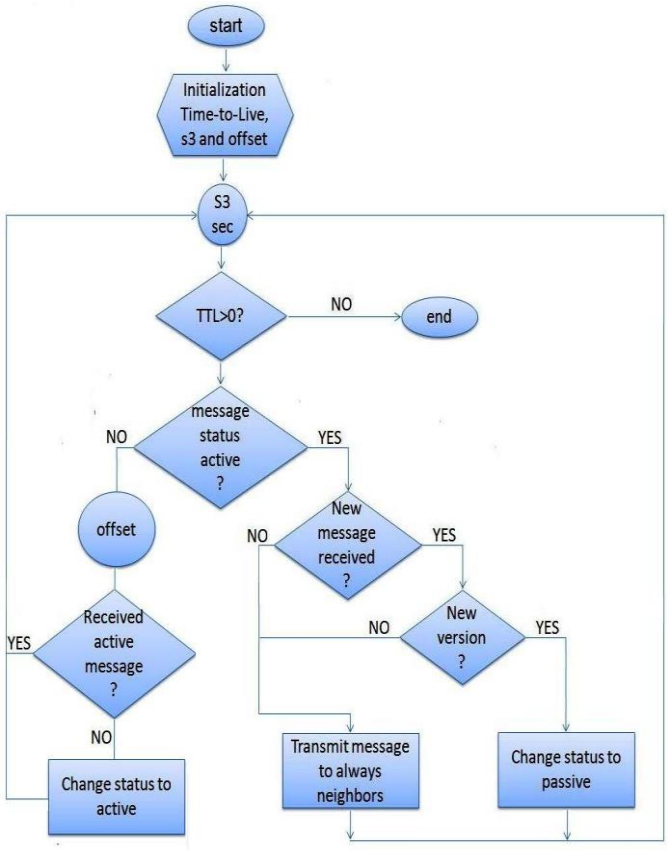


Figure 5. Block diagram of Message transmission algorithm

The developed algorithms take into account the latest developments in the area of ad-hoc networks. They are reliable and flexible.

V. Conclusion

One of the most important areas of investigation nowadays are Intelligent Transport Systems – ITS. Ad-hoc network VANET is a new and perspective class, which is gaining popularity and promises to decide traffic problems. In conclusion, there are not most optimal and design works on new algorithms have to continue.

The paper presents VANET network and algorithms. We have proposed Inquiry algorithm and Message transmission algorithm. Part three describe algorithms design. Inquiry process and methods of exchanging messages have been considered. Part four contains pseudo codes and block diagrams of algorithms. Message transmission algorithm in VANET is a novel traffic broadcasting algorithm designed for improving communication performance of a particular VANET. In MTA, the approaches of the message delivery very relying on the features of nodes, including the type, position direction of each communication vehicle at the moment. The collision in this algorithm is low and reliability is moderate. As a result – it improves road safety like the vehicles transmit and receive messages for congestions, accidents, bad weather, poor road condition and more.

Future work should concentrate on working the algorithm in real urban environments and doing researchs and simulations.

References

[1] M. Arif Khan, Tanveer A. Zia, Lihong ZhengJ. Vehicular Ad-Hoc Networks (VANETs) - An Overview and Challenges Sabih ur Rehman*, Journal of Wireless Networking and Communications 2013

[2] T. Kosch, C. Schroth, M. Strassberger, and M. Bechler, Automotive Internetworking, Wiey, New York, NY, USA, 2012.

[3] Wenshuang Liang, Zhuorong Li, Hongyang Zhang, Shenling Wang, and Rongfang Bie, Vehicular Ad Hoc Networks: Architectures, Research Issues, Methodologies, Challenges, and Trends, *College of Information Science and Technology, Beijing Normal University, Beijing 100875, China*

[4] Fall, Kevin, and Kannan Varadhan. "The network simulator (ns-2)." [Online]. Available: http://www.isi.edu/nsnam/ns/ [Accessed: May- 2013].

[5] F.J. Ros, P.M. Ruiz, and I. Stojmenovic, "Acknowledgment-Based Broadcast Protocol for Reliable and Efficient Data Dissemination in Vehicular Ad Hoc Networks," Mobile Computing, IEEE Transactions *on* , vol.11, no.1, pp.33-46, Jan. 2012.

[6] D. Emery and R. Hilliard, "Every architecture description needs a framework: expressing architecture frameworks using ISO/IEC 42010," in Proceedings of the Joint Working IEEE/IFIP Conference on Software Architecture & European Conference on Software Architecture (WICSA/ECSA '09), pp. 31–40, Cambridge, UK, September 2009.

[7] Y. Luo, W. Zhang, and Y. Hu, "A new cluster based routing protocol for VANET," in Networks Security Wireless Communications and Trusted Computing (NSWCTC), 2010, pp. 176–180, Second International Conference on, vol. 1, 2010

[8] V. Naumov, R. Baumann, and T. Gross, "An evaluation of inter-vehicle ad hoc networks based on realistic vehicular traces," Proceedings of the seventh ACM international symposium on Mobile ad hoc networking -Mobi-Hoc '06, p. 108, 2006.

[9] Z. Rawashdeh and S. Mahmud, "A novel algorithm to form stable clusters in vehicular ad hoc networks on highways," *EURASIP Journal on Wireless Communications and Networking*, no. 1, pp. 1–13, 2012,vol. 2012

SIGNAL PROCESSING AND ENGINEERING EDUCATION – POSTER SESSION

Possibilities for Open Source Software in Mechatronic Systems Design

Gordana Janevska[1] and Mitko Kostov[1]

Abstract – **This paper considers open source software for designing mechatronic systems. Thereby, emphasis is placed on Scilab/Xcos. Through simple example the paper presents the way how to formalise dynamic system with known physical and mathematical model in this environment.**

Keywords – **Mechatronics, model-based design, open source software, Scilab/ Xcos.**

I. Introduction

The design of mechatronic systems could be done with the so-called V model (VDI guideline 2206). A major role in the VDI 2206 guideline has modelling and model analysis in the domains of mechanical engineering, electrical engineering and information technology. Usually, commercial software tools are used to design mechatronic systems. MATLAB/Simulink and its toolboxes are well-known as leading software for model-based design of mechatronic systems, but many others can be added. On the other hand, it is of interest to know Open Source alternatives for it.

A dozen years ago, free software was approved as useful in the area of information technology. The expression "Free and Open Source Software" (FOSS) was introduced to highlight the difference of "free" in terms of free of charge and free access to the source code of the software. Since then FOSS has been developed to a successful business model for computer programs. In information technology there is no doubt in Linux as a very good operating system for network server. The largest supercomputer runs under Linux, and Linux is already a market leader as an embedded operating system for embedded systems. But FOSS is, of course, more than Linux, it also includes other very successful server applications.

From the view point of using software in the field of mechatronics, a small enterprise that develops mechatronic systems according to instructions of the VDI 2206 guideline, should invest in appropriate software and must pay annually high price for software maintenance. Otherwise, it will lose the opportunity for further development of the software, and therefore its customers. From this aspect, the facts about using FOSS as an addition to many useful existing software tools can be considered.

Bruce Perens began using the term FOSS and provides a good analysis of the economic benefits. It includes free access to source code, encouraging other companies, even competitors, to join in software development and share costs so as to accelerate this development.

[1]Gordana Janevska and Mitko Kostov are with the Faculty of Technical Sciences at St. Kliment Ohridski University, Bitola, Makedonska Falanga 33, Bitola 7000, Macedonia, E-mail: gordana.janevska@uklo.edu.mk

II. Open Source Alternatives to MATLAB/Simulink

MATLAB/Simulink is one of the leading software packages for model-based design of mechatronic systems and market leader in the development and simulation of control systems. This also can be concluded from the MathWorks pricelist. Small engineering offices have financial problems with buying and maintaining the products of commercial software packages. It is therefore good to know some alternatives.

In addition to MATLAB, there are other partially specialized tools such as Modelica/Dymola, ADAMS, ANSYS, ASCET MD/RP, AMESim, CarMaker, DSpace, SimulationX and many others.

Wikipedia lists some alternatives to MATLAB/Simulink at http://en.wikipedia.org/wiki/MATLAB#Alternatives. It is advisable to work with GNU Octave.org and Scilab.org.

GNU Octave, or short *Octave*, has been developing as a MATLAB clone for decades. It runs on operating systems like MS Windows, Linux, Mac OS X and others. It uses MATLAB files, i.e. MATLAB's Control System Toolbox functions, but a disadvantage is that in Octave there is no graphic editor for building block diagrams such as Simulink in MATLAB. This software can be used, for example, as an introduction to MATLAB, instead of MATLAB, and each student can use it without any problems with licenses. Considering industrial use, there are no interfaces for other programs and extensions for real-time work. Documentation is available on internet, and there already exists some literature.

Another way of modelling systems is proposed by the University of Berkeley with the development of *Ptolemy II*. This is a promising tool and already contains tools to work in real time. Perhaps it will allow not only modelling, but also realization of experiments in real time, and it would be particularly useful with compilers for rapid software prototyping directly from the tested models.

There are some other free software tools, such as FreeMat which is a MATLAB clone and processes MATLAB scripts, then JMathLib and so on.

Another widely used free open-source software, which is emphasized in this paper is Scilab/Xcos. Scilab/Xcos is a software tool with graphical editor for block diagrams that is comparable to MATLAB/Simulink.

III. SCILAB/XCOS

Developed since 1990 by researchers from the French Government's INRIA ("Institut Nationale de Recherche en Informatique et en Automatique", i.e. National Institute for Research in Informatics and Automation), SciLab is now maintained and developed by the Scilab Consortium after the Consortium was founded in May 2003. It is distributed freely over the Internet and has been open-source since 1994. SciLab is currently widely used in educational and industrial environments around the world.

A key feature of the SciLab syntax is its ability to work with matrices: basic matrix calculations like as transpose are immediately executed, as well as basic operations such as addition or multiplication of matrices. SciLab can also work with more complex objects than numerical matrices. For example, those working in the field of control systems can work with rational or polynomial transfer matrices i.e. SciLab provides a natural symbolic display of complicated mathematical objects such as transfer function, linear systems or graphs.

Starting with version 5.4, which was released in October 2012, Scilab comes along with a convenient user interface, with a files browser, a variables browser, command history, and a command window. The SciNotes editor opens as a separate window.

Similar to Simulink, it has a graphical interface for block diagrams called Xcos, and components library. Xcos is a graphic based software for modelling dynamic systems. Xcos can create block diagrams that will be used to model and simulate the dynamics of hybrid dynamic systems and models can be compiled in executable code. The new extensions allow the execution of control in real time as well as component-based modelling of electrical and hydraulic circuits.

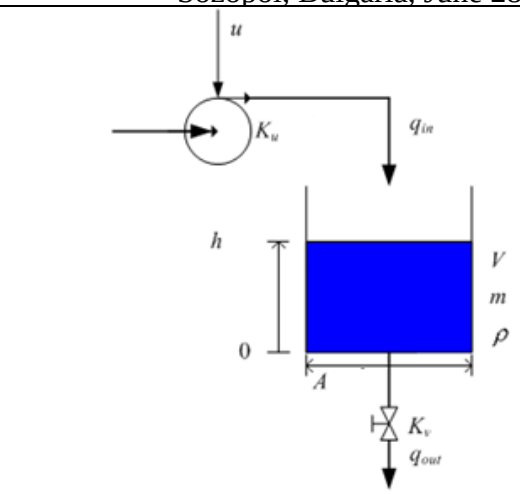
IV. CASE STUDY

Real system can be presented with a theoretical model through four steps. The *real system* is simplified, resulting in a *physical model* and mathematical equations can be written, resulting in a *mathematical model*. The mathematical model can be solved with a *computational model*, and the results are *visualized* by simulation. Xcos is used for the two last steps of theoretical modelling.

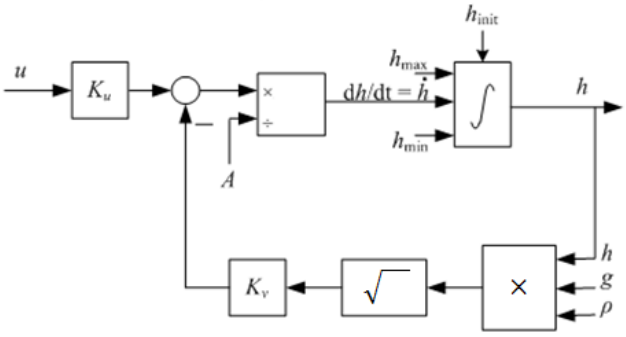
A. Mathematical Model

The simulated system is a liquid tank, with a pump inflow and outflow valve (Fig. 1). The simulator will calculate and display the level *h* over time. The simulation is in real time, thus giving a feeling of a "real" system. In fact, since the reservoir is a bit slow, the simulation is accelerated in order the simulation time to run faster than real time, to avoid unnecessary waste of time. The user can adjust the input by adjusting the pump control signal, *u*.

Each simulator is based on a mathematical model of the system that needs to be simulated. Accordingly, a mathematical model of the tank should be developed.



(a)



(b)

Fig. 1. a) Physical model of liquid tank b) Mathematical block diagram of the model.

By introducing relevant assumptions, a real system can be simplified and afterwards a corresponding physical model can be created. Following assumptions are introduced (parameters used in the expressions below are defined in Fig. 1a):

- The liquid is incompressible, i.e. the density of the liquid is ρ = const;
- The tank has straight vertical walls, i.e. A = const;
- The mass and level of the liquid in the reservoir are related to the equation:

$$m(t) = \rho A h\ (t)$$

- The inlet volume flow through the pump is proportional to the pump control signal:

$$q_{\text{in}}(t) = K_u u(t)$$

- The outlet volume flow through the valve is proportional to the square root of the pressure drop of the valve. This drop in pressure is assumed to be equal to the hydrostatic pressure at the bottom of the tank:

$$q_{\text{out}}(t) = K_{\text{v}}\sqrt{\rho g h(t)}$$

Based on the law of conservation of mass, the continuity equation for the liquid in the reservoir can be written:

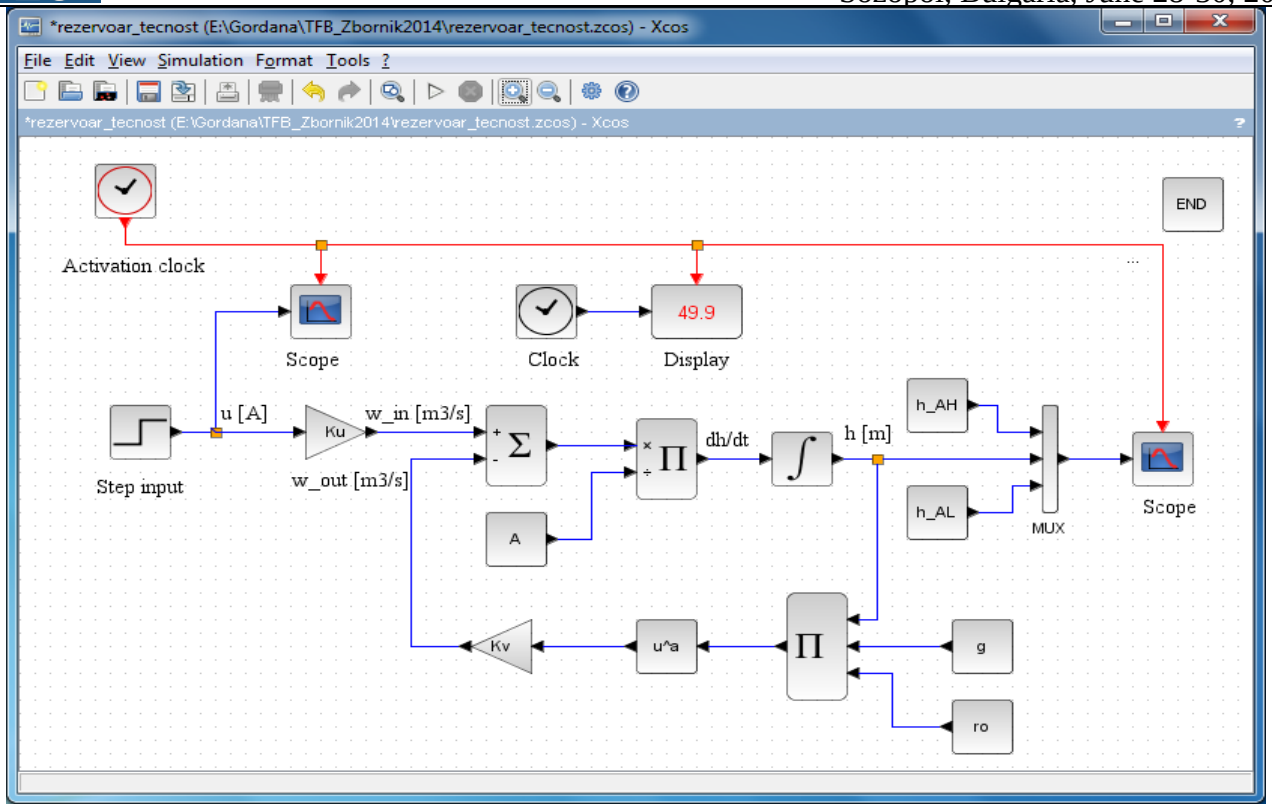


Fig 2. Simulation block diagram of the liquid tank in Xcos.

$$\frac{dm(t)}{dt} = \rho q_{\text{in}}(t) - \rho q_{\text{out}}(t) \qquad (1)$$

or, taking into account the above relations, it is obtained:

$$\frac{d[\rho Ah(t)]}{dt} = \rho K_u u(t) - \rho K_v \sqrt{\rho g h(t)} \qquad (2)$$

Based on the above equation, a mathematical block diagram of the model can be drawn. This block diagram can be implemented in the simulation block diagram. As an appropriate starting point for drawing the mathematical block diagram, the differential equation (2) will be written as a *state space model*, that is, the first-order time derivative to be alone on the left side, so the differential equation obtains the following form:

$$\frac{d[h(t)]}{dt} = \frac{1}{\text{A}}\left[K_u u(t) - K_v \sqrt{\rho g h(t)}\right] \qquad (3)$$

The upper equation is a differential equation for $h(t)$, i.e. *mathematical model* of the considered reservoir. According to it, $h(t)$ can be calculated by using the simulator when $dh(t)/dt$ is integrated with respect to time from 0 to t, with the initial value $h(0)$, which is marked here as h_{init}. Drawing a block diagram of the model (3) can began with adding an integrator to an empty block diagram. The input of this integrator is $dh(t)/dt$, and the output is $h(t)$. Afterwards, blocks with mathematical functions are added in order to construct the expression for $dh(t)/dt$ represented by the right side of the equation (3). The final block diagram for the model (3) is shown in Fig. 1b).

B. Xcos Block Diagram and System Simulation

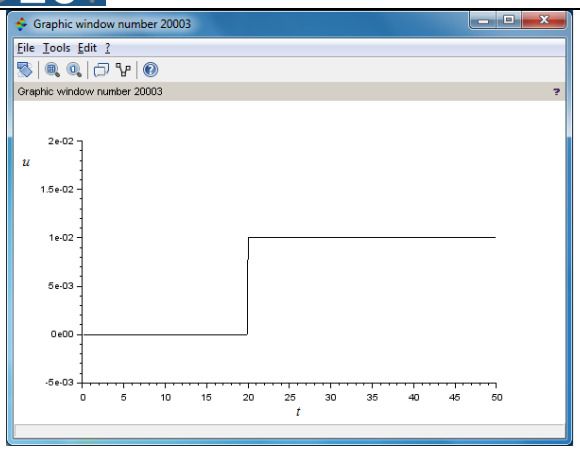
Xcos provides many elementary blocks organized in different palettes (**Palettes**) that can be accessed through the **Palette browser - Xcos**. Blocks from palettes can be copied to the main Xcos window by clicking first on the desired block, and then at the location where the block should be copied to the Xcos window. The blocks are then interconnected with connections, which is accomplished in a natural way by using the computer mouse. The blocks can be configured by double clicking on the block itself and entering the corresponding numerical or (preferably) context variables (parameters) with a specific name.

The simulation block diagram of the liquid tank built in Xcos is shown in Fig. 2.

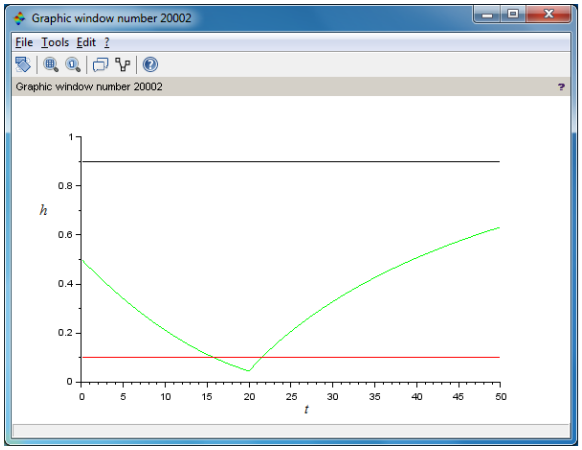
In the simulation of the level of the considered reservoir with liquid, the initial value of the level is 0.5 m. The control signal of the pump u is 0 till 20 s simulation time, and at 20 s it changes with step 0-0.01A. The simulation is from the starting time 0 s to the final time 50 s. The numerical values of the parameters are:

ρ = 1000 kg/m^3	A = 1 m^2
g = 9.81 m/s^2	h_{init} = 0.5 m
K_v = 0.0005	h_{max} = 1 m
K_u = 5 m^3/A	h_{min} = 0 m

It is assumed that there are level "alarm" limits, which should be drawn along with the level in the simulator. The boundaries



(a)



(b)

Fig. 3. a) Simulation of the control signal of the pump u; b) Level h and level alarm values h_{AL} = 0.1 m and h_{AH} = 0.9 m.

are: h_{AH} = 0.9 m (alarm high) and h_{AL} = 0.1 m (alarm low). The results of the simulation are shown in Figs. 3a and 3b.

V. Conclusion

In order to enable better organized development of FOSS (Free and Open Source Software) with companies as clients of these tools, OSADL (Open Source Automation Development Laboratory) was established in 2006. Here companies in mechanical engineering, electrical engineering and electronics, computer technology, and universities have been joined in order to define new and open standards in automation.

MATLAB/Simulink alternatives are indicated in this paper. The open source software packages Scilab/Xcos and GNU Octave are two alternatives that are ready for use in education and industry. The tools are supported by the company (Scilab / Xcos), or by the Internet community (Octave). Documentation, instructions and other help are available online.

The paper also presents some aspects of Xcos formalism. Dynamic systems modelling is presented in a precise and simple manner using the Xcos environment as a model for calculation and simulation, which are the last steps in theoretical modelling.

However, it should be noted that the main advantages of MATLAB/Simulink are, for example, the vast amount of pallet tools (toolboxes) and professional support by the MathWorks worldwide.

References

[1] M. Lohöfener, “Design of Mechatronic Systems and Benefit of Open Source Software Tools”, 9th International Workshop on Research and Education in Mechatronics, Bergamo, IT, Sep.18th– 19th 2008.

[2] S. L. Campbell, J. Chancelier and R. Nikoukhah, *Modeling and Simulation in Scilab/Scicos with ScicosLab 4.4*, 2nd ed., Springer, 2010.

[3] Scilab Enterprises, *Xcos for very beginners*, Scilab Enterprises, 2013.

[4] *Scilab Open source software for numerical computation*, Documentation, http://www.scilab.org/resources/documentation

The Prediction of Acoustic Parameters of the Occupied Concert Hall

Violeta Stojanović[1], Zoran Milivojević[2] and Momir Praščević[3]

***Abstract* –This paper includes the prediction of acoustic parameters: reverberation time, speech intelligibility and music clarity parameters for the fully occupied concert hall Promenadikeskus Pori, Finland. In the first part of this paper, we present the assessed values of the analyzed acoustic parameters according to the acoustic analysis of the impulse response of the concert hall and the estimation algorithms. The results are presented in tables and graphs. In the second part of the paper we determine the acoustic purpose of the concert hall, based on the comparison of the achieved results.**

Keywords – **Room Impulse Response, Acoustic Parameters, Prediction.**

I. Introduction

In concert halls music and acoustics are inseparable. There are held concerts and events that require, by their features, very different acoustic properties of the space.

A concert hall can be seen as a big resonance box of an instrument, where sound waves reflected off the walls and thereby initiate the phenomenon of reverberation. This phenomenon can be experienced subjectively as pleasant or irritating [1]. An objective classification is achieved with acoustical parameters extracted from the impulse response of the room. Thus, accurate impulse response measurements, already regulated by ISO 3382 standard [2] are desired for the assessment of acoustical properties.

This work contains the analysis of the predicted values of acoustic parameters which characterize the speech intelligibility and clarity of music for the occupied concert hall "Promenadikeskus" in the Pori (Finland).The analysis is based on the data of impulse response of the concert hall, available on http://legacy.spa.aalto.fi/projects/poririrs/. The impulse responses were recorded at three locations of the sound source at multiple measurement points. Using software packages EASERA and Matlab, as well as and theoretical models, the following results were achieved: a) for an empty concert hall for each location of the sound source and all measuring points we determined: reverberation time by octaves $RT_{30unocc}$, mean value of the reverberation time by octaves $\overline{RT_{30unocc}}$, speech transmission indexSTI_{unocc}, mean value of the speech transmission index $\overline{STI_{unocc}}$, clarity of music $C_{80unocc}$ and mean value of the clarity of music $\overline{C_{80unocc}}$ and b) provided for the fully occupied concert hall for each sound source and measurement point: the mean value of the reverberation time by octaves $\overline{RT_{30occ}}$, speech transmission indexSTI_{occ},the mean value of the speech transmission index $\overline{STI_{occ}}$, clarity of music C_{80occ} and mean value of the clarity of music $\overline{C_{80occ}}$. The obtained values of these parameters were analyzed in relation to standard values and certain conclusions were presented.

The work is organized as follows: Section II presents the estimation algorithms of the analyzed acoustic parameters, Section III describes the experiment and the results are presented in tables and graphs, Section IV presents the result analysis, whereas Section V is the conclusion.

II. The Evaluation of the Parameters for the Full Room

A. Reverberation Time

The evaluation of the reverberation time for the full room RT_{occ} can be carried out using the following formula [3]:

$$RT_{occ} = RT_{unocc} - DT \tag{1}$$

where RT_{unocc} is estimated reverberation time for the empty room, and DT is Shulc's diffusiion time which is calculated using the equations from the Table I [3].

TABLE I
SHULC'S ESTIMATED VALUES FOR RT

f (Hz)	DT (s)
125	0.510RT-0.708
250	0.605RT-0.867
500	0.668RT-0.929
1000	0.696RT-0.935
2000	0.694RT-0.889
4000	0.652RT-0.752

B. Speech Intelligibility

The evaluation of the speech intelligibility parameters for the full room STI_{occ} can be carried out using [4]:

$$STI_{occ} = STI_{unocc} + \Delta STI\ . \tag{2}$$

[1]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

[2] Zoran Milivojević is with the College of Applied Technical Sciences of Niš, 20. Aleksandra Medvedeva, St, 18000 Niš, Serbia.

[3] Momir Praščević is with the Faculty of Occupational Safety in Niš University of Niš, 10A Čarnojevića, St, 18000 Niš, Serbia.

STI_{unocc} is the index of the speech transmission for the empty room and ΔSTI je the value which is calculated using the following formula:

$$\Delta STI = 0.45\, ln \frac{RT_{unocc}(2kHz)}{RT_{occ}(2kHz)} + 0.012\,, \qquad (3)$$

where RT_{unocc} is the reverberation time of the empty room and RT_{occ} is the reverberation time of the full room. These values are taken at f=2 kHz [4].

C. Music Clarity

The evaluation of the music clarity in the room, C_{80}, can be carried out using Baron's theory of sound decreasing in which the direct sound follows an exponential decrease in sound energy [5, 6]. Propagation time of the direct sound from the source to the receiver at the distance r is t_D=r/c, where c is the speed of sound.

For the room of the volume V integrated acoustic energy l for the time $t \geq t_D$, where $t \to \infty$, at certain location in the room is:

$$l(t) = 31200\, \frac{RT}{V} e^{-\frac{13.82t}{RT}}\,, \qquad (4)$$

where RT is the reverberation time.

The evaluation of music clarity parameters in the room, C_{80}, includes three components of the received sound at the receiver position: the direct sound d (Eq. 5), the early reflected sound e with the delay less than 80 ms (Eq. 6) and the late reflected sound l with the delay of more than 80 ms (Eq. 7):

$$d = \frac{100}{r^2}\,, \qquad (5)$$

$$e = 31200\, \frac{RT}{V} e^{-\frac{0.04r}{RT}} \left(1 - e^{-\frac{1.11}{RT}}\right) \qquad (6)$$

and

$$l = 31200\, \frac{RT}{V} e^{-\frac{0.04r}{RT}} e^{-\frac{1.11}{RT}}\,. \qquad (7)$$

where r is the distance between the receiver and the source. The estimated parameter of the sound clarity in the room C_{80} is:

$$C_{80} = 10\, log \frac{\int_{t_0-0.08}^{t_0} p^2(t)dt}{\int_{t_0+0.08}^{\infty} p^2(t)dt} = 10\, log \frac{d+e}{l} \qquad (8)$$

III. EXPERIMENTS

The evaluation of the acoustic parameters has been carried out for thePromenadikeskusconcert hall in Pori, Finland, which is shown in the Fig. 1.

According to the impulse responses [7] we determined the value of the objective acoustic parameters RT_{30}, STI and C_{80} for the unoccupied concert hall. Evaluation of these parameters was performed using theoretical models and Matlab program package.

For the analysis of acoustic parameters, we considered three locations of the sound source S and 4 (of 7) measurement locations, MP, the receiver, with the arrangement shown in Fig. 2.

Fig. 1. Promenadikeskus concert hall in Pori.

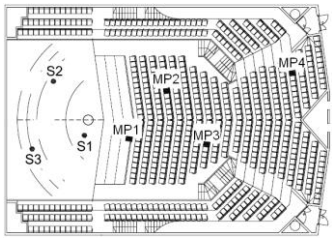


Fig. 2. Location of sound sources (S1, S2 i S3) and measuring points (MP1 ÷ MP4).

For the purpose of the room acoustic analysis, we calculated mean values of acoustic parameters considered for each position of the sound source at the measuring point and for each position of the sound source at all measuring points in the analyzed hall. The results are presented in tables and graphs.

A. The Basis

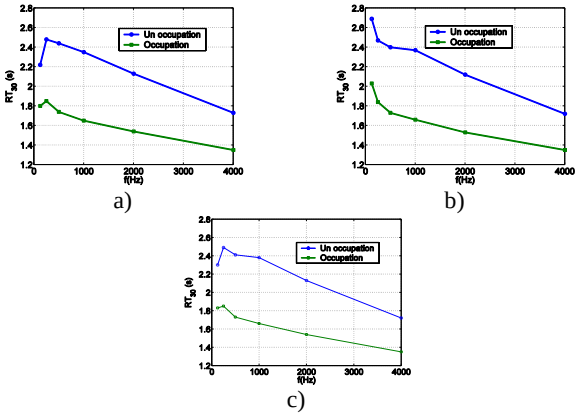
The impulse responses base contains audio recordings of the Pori Promenadikeskus concert hall impulse responses that have been recorded within the TAKU/V Ä RE technical project [7]. The dimensions of the hall are: $33 \times 23 \times 15$ m. Total volume is approximately 9500 m^3 and capacity of about 700 people.

The responses were acquired using the IRMA measurement system. For all measurements, 48 kHz sampling rate and 16-bit A/D and D/A conversion were used. Two sound sources were utilized in the measurements. The measurements were performed using an omnidirectional loudspeaker DPA 4006 and a subwoofer. The impulse responses were measured using logarithmic sinusoidal sweep excitation. For measurements, the sweeps were synthesized in the time domain.

B. Results

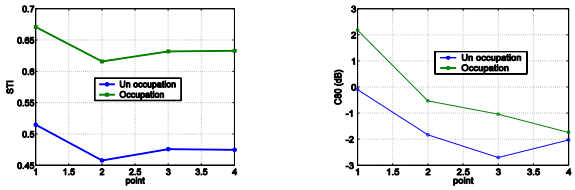
Table II shows the values: a) the reverberation time $RT_{30unocc}$,and mean value of the reverberation times, $\overline{RT_{30unocc}}$, for the unoccupied hall, b) of the Schultz diffusion, DT and c)

mean value of the reverberation times $\overline{RT_{30occ}}$ for the occupied hall. Tables III – V show the values: a) the reverberation time (1kHz, 2kHz), the speech transmission index and the clarity for the music for the unoccupied and occupied hall and b) the mean values of these parametars for MP points for sound source locations S1, S2 and S3. The change in mean values of reverberation time at measuring points for the unoccupied and occupiedconcert hall for the sound soucers positions S1, S2 and S3 is shown in Fig. 3. The change in the mean values of transmission index and clarity for music at measuring points for the unoccupied and occupied concert hall for the sound soucers positions S1, S2 and S3 is shown in Fig. 4. – Fig. 6., respectively.



a) b) c)

Fig.3. The mean value of the reverberation time at the measured points for both unoccupied and occupied concert hall for sound source locations: a) S1, b) S2 and c) S3.



a) b)

Fig. 4. The mean value of the transmission index (a) and clarity for music (b) at the measuring points for unoccupied and occupied concert hall for positions of sound soucers S1.

IV. The Results Analysis

Based on results shown in the Tables II – V and in the Figs. 3. - 6. we can draw the following conclusions:

1) The predicted mean values $\overline{RT_{30occ}}$, obtained by using the values to predict Schultz values (Table. I), are for all sound sources and measuring points less than the mean values $\overline{RT_{30unocc}}$. The decrease in the values of the reverberation time is especially important at frequencies close to or equal to the frequency f = 1 kHz because they are most important to the quality of speech intelligibility. At the positions of the sound source S1, S2 and S3, the decrease of the value of this parameter is $\overline{\Delta RT_{30}}(1\text{ kHz}) = \{0.7\,s, 0.715\,s, 0.72\,s\}$.

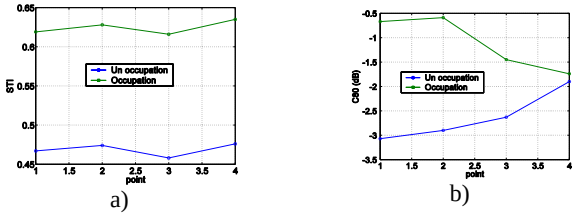
2) At medium frequencies, the predicted mean values $\overline{RT_{30occ}}$ for all positions of the sound source are the same and equal RT_{30srf} = 1.695 s. The calculated optimum reverberation time (depending on the volume) is RT_{opt} = 2.04 s. These

TABLE II
THE VALUES THE REVERBERATION TIME FOR UNOCCUPIED AND OCCUPIED CONCERT HALL AND SCHULTZ DIFFUSION TIME, *DT*

Sound source	*f* (Hz)	$RT_{30unocc}$ (s)				$\overline{RT_{30unocc}}$ (s)	*DT* (s)	$\overline{RT_{30occ}}$ (s)
		MP1	MP2	MP3	MP4			
S1	125	2.71	2.59	2.83	0.76	2.22	0.42	1.8
	250	2.44	2.42	2.56	2.51	2.48	0.63	1.85
	500	2.45	2.4	2.43	2.47	2.44	0.7	1.74
	1000	2.33	2.34	2.36	2.36	2.35	0.7	1.65
	2000	2.12	2.13	2.12	2.13	2.13	0.59	1.54
	4000	1.72	1.72	1.72	1.76	1.73	0.38	1.35
S2	125	2.62	2.64	2.7	2.79	2.69	0.66	2.03
	250	2.41	2.36	2.54	2.57	2.47	0.63	1.84
	500	2.45	2.33	2.42	2.4	2.4	0.67	1.73
	1000	2.36	2.37	2.4	2.35	2.37	0.71	1.66
	2000	2.09	2.11	2.13	2.15	2.12	0.59	1.53
	4000	1.7	1.68	1.75	1.75	1.72	0.37	1.35
S3	125	2.71	2.66	1.04	2.8	2.3	0.47	1.83
	250	2.46	2.52	2.48	2.5	2.49	0.64	1.85
	500	2.4	2.39	2.44	2.41	2.41	0.68	1.73
	1000	2.38	2.36	2.37	2.41	2.38	0.72	1.66
	2000	2.13	2.1	2.14	2.14	2.13	0.59	1.54
	4000	1.73	1.7	1.7	1.76	1.72	0.37	1.35

TABLE III
THE VALUES: THE REVERBERATION TIME (1KHZ, 2KHZ), THE SPEECH TRANSMISSION INDEX AND THE CLARITY FOR THE MUSIC FOR THE UNOCCUPIED AND OCCUPIED HALL FOR MP POINTS FOR SOUND SOURCE LOCATION S1

MP	$RT_{30unocc}$ (1 kHz) (s)	RT_{30occ} (1 kHz) (s)	RT_{30unoc} (2 kHz) (s)	RT_{30occ} (2 kHz) (s)	STI_{unocc}	STI_{occ}	$C_{80unocc}$ (dB)	C_{80occ} (dB)
1	2.33	1.64	2.12	1.54	0.515	0.671	-0.1	2.18
2	2.34	1.65	2.13	1.54	0.458	0.616	-1.83	-0.53
3	2.36	1.65	2.12	1.54	0.476	0.632	-2.7	-1.04
4	2.36	1.65	2.13	1.54	0.475	0.633	-2.03	-1.74
	$\overline{RT_{30unocc}}$ (1 kHz) (s)	$\overline{RT_{30occ}}$ (1 kHz) (s)	$\overline{RT_{30unocc}}$ (2 kHz) (s)	$\overline{RT_{30occ}}$ (2 kHz) (s)	$\overline{STI_{unocc}}$	$\overline{STI_{occ}}$	$\overline{C_{80unocc}}$ (dB)	$\overline{C_{80occ}}$ (dB)
	2.35	1.65	2.125	1.54	0.48	0.638	-1.67	-0.28



a) b)

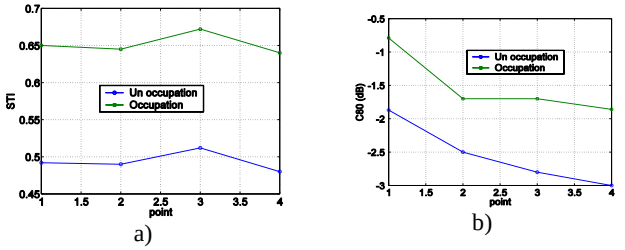
Fig. 5. The mean value of the transmission index (a) and clarity for music (b) at the measuring points for unoccupied and occupied concert hall for positions of sound soucers S2.

TABLE IV
THE VALUES: THE REVERBERATION TIME (1KHZ, 2KHZ), THE SPEECH TRANSMISSION INDEX AND THE CLARITY FOR THE MUSIC FOR THE UNOCCUPIED AND OCCUPIED HALL FOR MP POINTS FOR SOUND SOURCE LOCATION S2

MP	$RT_{30unocc}$ (1 kHz) (s)	RT_{30occ} (1 kHz) (s)	RT_{30unoc} (2 kHz) (s)	RT_{30occ} (2 kHz) (s)	STI_{unocc}	STI_{occ}	$C_{80unocc}$ (dB)	C_{80occ} (dB)
1	2.36	1.65	2.09	1.53	0.467	0.619	-3.07	-0.67
2	2.37	1.66	2.11	1.54	0.474	0.628	-2.9	-0.59
3	2.4	1.66	2.13	1.54	0.458	0.616	-2.63	-1.45
4	2.35	1.65	2.15	1.55	0.476	0.635	-1.9	-1.74
	$\overline{RT_{30unocc}}$ (1 kHz) (s)	$\overline{RT_{30occ}}$ (1 kHz) (s)	$\overline{RT_{30unocc}}$ (2 kHz) (s)	$\overline{RT_{30occ}}$ (2 kHz) (s)	$\overline{STI_{unocc}}$	$\overline{STI_{occ}}$	$\overline{C_{80unocc}}$ (dB)	$\overline{C_{80occ}}$ (dB)
	2.37	1.655	2.12	1.54	0.469	0.625	-2.63	-1.11

TABLE V
THE VALUES: THE REVERBERATION TIME (1KHZ, 2KHZ), THE SPEECH TRANSMISSION INDEX AND THE CLARITY FOR THE MUSIC FOR THE UNOCCUPIED AND OCCUPIED HALL FOR MP POINTS FOR SOUND SOURCE LOCATION S3

MP	$RT_{30unocc}$ (1 kHz) (s)	RT_{30occ} (1 kHz) (s)	RT_{30unoc} (2 kHz) (s)	RT_{30occ} (2 kHz) (s)	STI_{unocc}	STI_{occ}	$C_{80unocc}$ (dB)	C_{80occ} (dB)
1	2.38	1.66	2.13	1.54	0.492	0.65	-1.87	-0.79
2	2.36	1.65	2.1	1.53	0.49	0.645	-2.5	-1.7
3	2.37	1.66	2.14	1.54	0.512	0.672	-2.8	-1.7
4	2.41	1.67	2.14	1.54	0.48	0.64	-3	-1.86
	$\overline{RT_{30unocc}}$ (1 kHz) (s)	$\overline{RT_{30occ}}$ (1 kHz) (s)	$\overline{RT_{30unocc}}$ (2 kHz) (s)	$\overline{RT_{30occ}}$ (2 kHz) (s)	$\overline{STI_{unocc}}$	$\overline{STI_{occ}}$	$\overline{C_{80unocc}}$ (dB)	$\overline{C_{80occ}}$ (dB)
	2.38	1.66	2.13	1.54	0.49	0.65	-2.54	-1.51



a) b)

Fig. 6.The mean value of the transmission index (a) and clarity for music (b) at the measuring points for unoccupied and occupied concert hall for positions of sound soucers S3.

values are in the range determined by the value of the optimum reverberation time for the concert hall standard: RT = 1.5 ÷ 2.2 s.

3) According to Barron's recommended reverberation time values for the occupied concert hall [5], the analyzed concert hall with the value for RT_{30srf} = 1.695 s is suitable for performing the following: early classical music (RT = 1.6 ÷ 1.8 s), opera (RT = 1.3 ÷1.8 s) and chamber music (RT = 1.4 ÷1.7 s).

4) At the positions of the sound source S1, S2 and S3 the increase in the mean value of index transmission is uniform: $\overline{\Delta STI} = \{0.158, 0.156, 0.16\}$.

5) The mean values of this parameter for the unoccupied hall at all three positions of the sound source, $\overline{STI_{unocc}} = \{0.48, 0.469, 0.49\}$ are in the range determined by the values that represent the acceptable speech intelligibility: 0.45 ÷ 0.6.But the value of this parameter for an occupied hall, $\overline{STI_{occ}} = \{0.63, 0.625, 0.65\}$, are in the range determined by the values that represent good speech intelligibility: 0.6 ÷0.75.

6) The estimated values of the medium index of music clarity in all positions of the sound source, for both the unoccupied is ($\overline{C_{80unocc}} = \{-1.67\,dB, -2.63\,dB, -3dB\}$), and occupied hall is ($\overline{C_{80occ}} = \{-0.28\,dB, -1.11\,dB, -1.86dB\}$).

7) The increase in the mean value of the index of music clarity for the occupied hall at all sound source positions S1, S2, and S3 is $\overline{\Delta C_{80}} = \{1.39\,dB, 1.52\,dB, 1.14dB\}$.

8) The predicted average values of music clarity meet the criteria of optimum values for orchestral music: (0 ÷-3)dB

V. CONCLUSION

The work includes the analyzes of the predicted values of acoustic parameters: reverberation time, the index of speech intelligibility and music clarity , for an occupied concert hall. The presence of people in the concert hall led to a decrease in the mean value of the reverberation time for $\overline{\Delta RT_{30}}$ (1kHz) ≈ 0.7 s, increase in the mean value of the speech transmissionindex for $\overline{\Delta STI} \approx$ 0.16 and the maximum increase in the mean value of music clarity of $\overline{\Delta C_{80}}$ = 1.52 dB. The predicted values of acoustic parameters confirmed the fact that this concert hall has been acoustically designed for early classical, opera and chamber music.

REFERENCES

[1] L. Beranek, "Concert Halls and Opera Houses: Music, Acoustics and Architecture", Springer-Verlag New York, 2end edition, pp. 19-35, 2004.

[2] ISO 3382-1:2009, "Acoustics-measurement of room acoustic parameters-part 1: performance spaces".

[3] J. S. Bradley, "A comparison of three classical concert halls", JASA, vol. 89, no.3, pp. 1176-1192, 1991.

[4] V. Desarnaulds, A. P. O. Carvalho , G. Monay, "The effect of occupancy in the speech intelligibility in churches", INTER NOISE 2001, Conference Proceedings, pp.27-30, The Hague, The Netherlands, 2001.

[5] M. Baron, Lee L. J, "Energy relations in concert auditoriums", JASA, vol. 84, no.2, pp. 618-628, 1988.

[6] T. Zamarreno, S. Giron, M. Galindo, "Assessing the intelligibility of speech and singing in Mudejar-Gothic churches", Applied Acoustics, vol. 69, pp. 242-254, 2008.

[7] J. Merimaa, T. Peltonen, T. Lokki, "Concert Hall Impulse Responses-Pori, Finland: Reference, Helsinky University of Technology, 2005.

Performance of the VWM algorithm in presence of impulse noise

Bojan Prlinčević[1], Zoran Milivojević[2] and Stefan Panić[3]

Abstract –**The first part of the paper describes VVM (Visible Watermarking) algorithm for inserting - removing visible watermark in an image. Impulse noise is superimposed to visible watermarked image. In the second part of the paper, the results of the experiment are shown in which the image with the visible watermark is superimposed with the impulse noise. The experiment was performed for blocks dimension k = {2, 3, 4} and coefficient of insertion υ = 40, and noise is varied in the range p = {1, 5, 10, 20, 30, 40, 50, 60, 70}. Improving the quality of recovered image was performed using the MDB algorithm for removing impulse noise. A comparative analysis of the original and recovered image was performed. The comparison was performed on the basis of MSE and PSNR. The obtained results are detailed analysed and presented in a tabular and graphical manner.**

Keywords – **Visible watermark, Impulsive noise, Filtering.**

I. INTRODUCTION

In order to protect the copyrights of multimedia content (usually images), techniques based on the insertion of the digital watermark are used. The most widely used copyright protection technique is the technique for insertion the invisible digital watermark. Invisible watermark can be extracted later for ownership proofing. On the other hand, in order to identify ownership, the technique of inserting the visible watermark is also used. This technique has recently become a widely used technique to prevent the viewers for making unauthorized use. When visible watermark is inserted, it is most often impossible to remove the watermark from the image [1] - [3]. This is in order to protect it more efficiently [4], [5]. However, in certain situations, there may be a need to remove the visible watermark from the image [6]. Many researches were done by researching the techniques for inserting reversible watermark in an image but there was many difficulties. After recovering cover image was with visible damages. In the paper [7], scheme for visible-watermark removal and reversible image recovery are proposed. In this scheme, the image with embedded visible watermark is generated after embedding visible watermark into the original image. The recovered image is acquired after removing the visible watermark from the embedded image. Proposed method of inserting visible watermark is one of the most effective methods thus, original image can be exactly recovered.

During the image processing some level of degradation is inevitable. The authors of this paper came up with the idea to perform the VWM (Visible Watermarking) algorithm in presence of impulse noise. The authors set the question: the effect of applying the MDB (Modificated Decision - Based) algorithm for denoising [8] image with visible watermark to improve the quality of recovered image (image with removed watermark)? The answer on this question was sought through the realization of an experiment within which the VWM algorithm for inserting - removing visible watermark were performed, and after superimposing impulse noise MDB algorithm for removing impulse noise is applied. Standard test images was used: Lena, Boat and Girl and Watermark image. The experiment was performed by inserting the inverse binary watermark. The watermark is inserted into the blocks, dimensions k = {2, 3, 4} with the coefficient of insertion υ = 40. Watermarked images was superimposed with impulse noise, varied in the range p = {1, 5, 10, 20, 30, 40, 50, 60, 70}. The quality of improving recovered image was analyzed after superimposing impulse noise and after filtering the image using the MDB algorithm [8]. As a quality measure of improved quality of the recovered image was used Mean Square Error (MSE) and Peak Signal-to-Noise Ratio (PSNR). The results are presented in tabular form and graphically.

The paper is organized as follows: Section II describes the VWM algorithm for inserting - removing visible watermark, and MDB filtering algorithm. Section III gives results and analysis of the results. Conclusion is given in section IV.

II. ALGORITHMS

In the paper, two algorithms are used: a) VWM algorithm for watermarking [7], and b) MDB denoising algorithm [8].

A. VWM algorithm

The VWM embedding algorithm is realized through following steps:

Step 1: The original image I_0 size $M_0 \times N_0$ is divided into nonoverlapping blocks $k \times k$, where $2 \le k \le k_m$, $k_m \in \{\lfloor M_0 / M_w \rfloor, \lfloor N_0 / N_w \rfloor\}$, and $M_w \times N_w$ is size of embedded binary watermark image. The parameter k can be 2, 3 or 4.

Step 2: One bit of the watermark can be embedded in each block B of host image I_0. A manner how blocks can be partioned, depending of parameter k value is presented in Table 1.

Step 3: Calculate the mean value of the assigned visible watermark region in original image I_0:

$$avg_w = \frac{\sum_{m=0}^{k \times M_w - 1} \sum_{n=0}^{k \times N_w - 1} I_0(m + \mu_1, n + \mu_2)}{(k \times M_w) \times (k \times N_w)}, \quad (1)$$

[1]Bojan Prlinčević is with the Higher technical proffesional school in Zvecan, Serbia, E-mail: b.prlincevic@vts-zvecan.edu.rs.

[2]Zoran Milivojević is with the College of Applied Technical Sciences, Nis Serbia, E-mail: zoran.milivojevic@vtsnis.edu.rs.

[3]Stefan Panić is with the Faculty of Natural Science and Mathematics, K. Mitrovica, Serbia, E-mail: stefanpnc@yahoo.com.

where (μ_1, μ_2) present coordinate of watermark image W embedded in I_0, ($k \times M_w$, $k \times M_w$) denotes the region of embedded watermark.

Step 4: Threshold T_w is calculated:

$$T_w = \begin{cases} awg_w + 30, \text{ if } awg_w \le 128, \\ awg_w - 30, \text{ if } awg_w > 128. \end{cases} \quad (2)$$

Step 5: Mean value of each block B is calculated:

$$avg_B = \frac{1}{k^2}\sum_{j=1}^{k}\sum_{i=1}^{k} B(i,j). \quad (3)$$

Step 6: Embedding procedure of each bit of the Watermark into corresponding block is done:

$$B'(i,j) = \begin{cases} B(i,j) + \omega \times \nu, \text{ if } awg_w < T_w, \\ B(i,j) - \omega \times \nu, \text{ if } awg_w \ge T_w, \end{cases} \quad (4)$$

where ω denotes bit of the watermark and υ is the coefficient of inserting the bit (highest value of υ mean more visible watermark).

Step 7: Different values of parameters d_1 and d_2 are calculated for each block:

$$\begin{aligned} d_1 &= \sum_{B'(i,j)\in S_1} B'(i,j) - \sum_{B'(i,j)\in S_2} B'(i,j) - x, \\ d_1 &= \sum_{B'(i,j)\in S_1} B'(i,j) - \sum_{B'(i,j)\in S_2} B'(i,j) - y, \end{aligned} \quad (5)$$

where S_1, S_2, x and y are parameters from Table 1.

Step 8: Modification of the pixel x into the pixel x' and y into the y' was done:

$$x' = 2 \times d_1 + \omega + \left\lfloor \frac{\sum_{B'(i,j)\in\{S_1 \cup S_2\}} B'(i,j)}{2 \times \left\lfloor \frac{k^2-1}{2} \right\rfloor - 1} \right\rfloor,$$

$$y' = \begin{cases} 2 \times d_2 + 0 + \left\lfloor \frac{\sum_{B'(i,j)\in\{S_1 \cup S_2\}} B'(i,j)}{2 \times \left\lfloor \frac{k^2-1}{2} \right\rfloor - 1} \right\rfloor, \text{ if } avg_B < T_w, \\ 2 \times d_2 + 1 + \left\lfloor \frac{\sum_{B'(i,j)\in\{S_1 \cup S_2\}} B'(i,j)}{2 \times \left\lfloor \frac{k^2-1}{2} \right\rfloor - 1} \right\rfloor, \text{ if } avg_B \ge T_w. \end{cases} \quad (6)$$

After implementing the steps above visible watermarked image I_w is obtained.

The VWM algorithm procedure for recovering image is realized through following steps:

Step 1: The watermarked image I_w is divided into nonoverlaped $k \times k$ blocks in the same manner with the embedding procedure, and parameters S'_1, S'_2, x' and y' can also be obtained as in embedding procedure.

Step 2: Using the obtained values for parameters S'_1, S'_2, x' and y', differences d'_1 and d'_2 can be calculated for each $k \times k$ watermarked block B':

$$\begin{aligned} d'_1 &= x' - \left\lfloor \frac{\sum_{B'(i,j)\in\{S_1 \cup S_2\}} B'(i,j)}{2 \times \left\lfloor \frac{k^2-1}{2} \right\rfloor - 1} \right\rfloor, \\ d'_2 &= y' - \left\lfloor \frac{\sum_{B'(i,j)\in\{S_1 \cup S_2\}} B'(i,j)}{2 \times \left\lfloor \frac{k^2-1}{2} \right\rfloor - 1} \right\rfloor. \end{aligned} \quad (7)$$

Step 3: Extracted binary watermark ω' from each block can be obtained using:

$$\omega' = \mathrm{mod}(d'_1, 2). \quad (8)$$

Step 4: Differences d_1 and d_2 can be obtained using:

$$\begin{aligned} d_1 &= \left\lfloor \frac{d'_1}{2} \right\rfloor, \\ d_2 &= \left\lfloor \frac{d'_2}{2} \right\rfloor. \end{aligned} \quad (9)$$

Step 5: Pixels x'' and y'' can be recalculated according with following equations:

$$\begin{aligned} x'' &= \sum_{B'(i,j)\in S_1} B'(i,j) - \sum_{B'(i,j)\in S_2} B'(i,j) - d_1, \\ y'' &= \sum_{B'(i,j)\in S_1} B'(i,j) - \sum_{B'(i,j)\in S_2} B'(i,j) - d_2, \text{ if } \omega'=1. \end{aligned} \quad (10)$$

Step 6: Recovered blocks B'' can be obtained according to the parameter υ, pixels x'' and y'' according to following equation:

$$B''(i,j) = \begin{cases} B'(i,j) - \omega' \times \nu, \text{ if } \omega_1 = 0, \\ B'(i,j) + \omega \times \nu, \text{ if } \omega_1 = 1, \end{cases} \quad (11)$$

where $\omega_1 = \mathrm{mod}(d'_2, 2)$.

Step 7: Recovered image I_r is obtained.

TABLE I
COEFFICIENT MODE OBTAINING FOR EACH BLOCK

k	$k \times k$ block B	S_1	S_2	x	y
2	$b_{1,1}$ $b_{1,2}$ / $b_{2,1}$ $b_{2,2}$	$b_{1,1}$	0	$b_{2,2}$	$b_{1,2}$
3	$b_{1,1}$ $b_{1,2}$ $b_{1,3}$ / $b_{2,1}$ $b_{2,2}$ $b_{2,3}$ / $b_{3,1}$ $b_{3,2}$ $b_{3,3}$	$b_{1,2}$, $b_{2,1}$, $b_{3,2}$, $b_{2,3}$	$b_{1,3}$, $b_{2,2}$, $b_{3,1}$	$b_{1,1}$	$b_{3,3}$
4	$b_{1,1}$ $b_{1,2}$ $b_{1,3}$ $b_{1,4}$ / $b_{2,1}$ $b_{2,2}$ $b_{2,3}$ $b_{2,4}$ / $b_{3,1}$ $b_{3,2}$ $b_{3,3}$ $b_{3,4}$ / $b_{4,1}$ $b_{4,2}$ $b_{4,3}$ $b_{4,4}$	$b_{1,2}$, $b_{2,1}$, $b_{1,4}$, $b_{2,3}$, $b_{4,1}$, $b_{3,4}$, $b_{4,3}$	$b_{1,1}$, $b_{1,3}$, $b_{2,2}$, $b_{2,4}$, $b_{4,2}$, $b_{4,4}$	$b_{3,3}$	$b_{3,2}$

B. MDB denoising algorithm

The MDB denoising algorithm is performed as it is explained in paper the [8].

Denoising procedure was don in following steps:

Step 1: Watermarked image is superimposed with the noise $p = \{1, 5, 10, 20, 30, 40, 50, 60, 70\}$. I_{wn} image with added noise is obtained.

Step 2: Above the image I_{wn} MDB denoising algorithm is applied and filtered I_{wf} image is obtained.

III. EXPERIMENTAL RESULTS AND ANALYSIS

A. Experiment

Experiment is performed according following steps:

Step 1: Applying the VWM algorithm on the original image I_0 watermarked image I_w is obtained.

Step 2: Image I_w is superimposed with different value of impulse noise $p = \{1, 5, 10, 20, 30, 40, 50, 60, 70\}$. I_{wn} image with added noise is obtained.

Step 3: Recovered image I_{nr} is obtained from the watermarked image with the added impulse noise I_{wn}, applying the VWM algorithm.

Step 4: Above the watermarked image with noise I_n, the MDB denoising algorithm was applied. Filtered watermarked image I_{wf} is obtained.

Step 5: From I_{wf} image recovered image I_{fr} is obtained, applying the VWM algorithm.

Image is watermarked with the inverse binary watermark image.

In order to compare the results, recovering the image is performed from the image with superimposed noise and from the filtered image.

Energy of inserting the bit is set on the value $\upsilon = 40$.

As a measure of the quality of the reconstructed image was used Mean Square Error (MSE) and Peak Signal-to-Noise Ratio (PSNR):

$$MSE = \frac{\sum_{ij}\left(x_{ij} - y_{ij}\right)^2}{M \times N}, \qquad (12)$$

$$PSNR = 10\log 10\frac{(2^n - 1)^2}{MSE}, \qquad (13)$$

where: $i = 1....M$, $j = 1...N$, $x_{i,j}$ - i,j- pixel of the original image, $y_{i,j}$ - i,j - pixel of the recovered image, $M \times N$ – size of image.

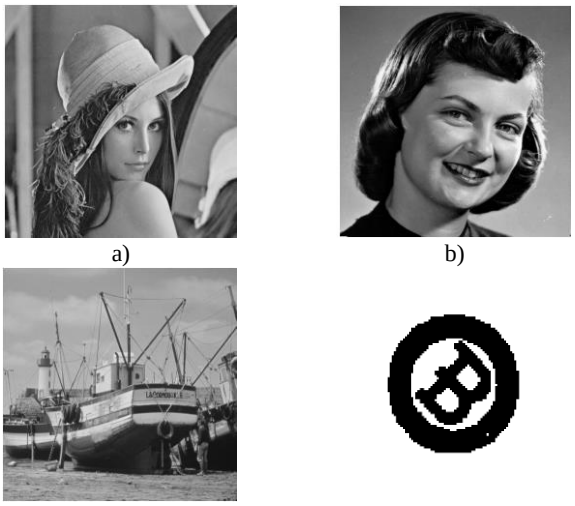
B. Base of experiment and results

For the purposes of experiment, the base of standard images: a) Lena, b) Girl and c) Boat was used and d) binary watermark Fig. 1..

Fig. 2. shows look of the image Lena after inserting the watermark with $k = 2$: a) after superimposed noise $p = 10$ %, b) after filtering, c) after recovering the image with noise, and d) after recovering the filtered image.

Fig. 3. shows look of the image Lena after inserting the watermark with $k = 4$: a) after superimposed noise $p = 10$ %, b) after filtering, c) after recovering image with noise, and d) after recovering filtered image.

Fig. 4. and Fig. 5. shows graphs of MSE and PSNR for recovered images, for different value of parameter $k = \{2, 3,4\}$, after superimposing the noise and after applying the MDB algorithm for filtering.



a) b) c) d)

Fig. 1. Base of images: a) Lena, b) Girl, c) Baboon and d) Watermark

a) b) c) d)

Fig. 2. Watermarked image Lena with $k = 2$ and $p = 10$: a) noised, b) filtered, c) recovered from noised image, d) recovered from filtered image.

Fig. 3. Watermarked image Lena with $k = 4$ and $p = 10$: a) noised, b) filtered, c) recovered from noised image, d) recovered from filtered image.

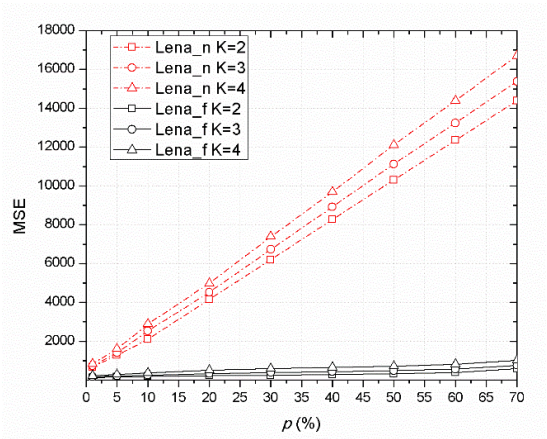


Fig. 4. Quality measure MSE for different value of $k = \{2, 3, 4\}$: Lena_n - Image Lena with noise, and Lena_f - filtered image Lena

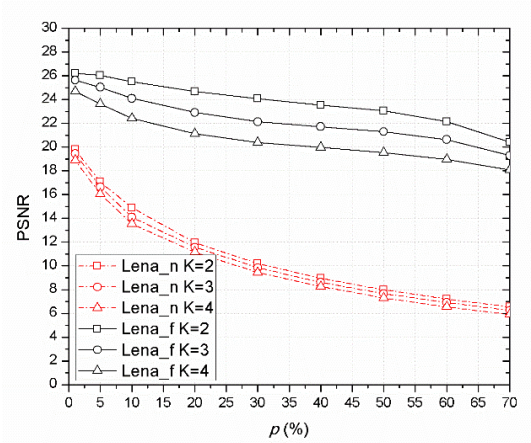


Fig. 5. Quality measure PSNR for different value of $k = \{2, 3, 4\}$: Lena_n - Image Lena with noise, and Lena_f - filtered image Lena

C. Analysis

Based on the results shown in Fig. 2. and Fig. 3. it can be concluded that the MDB algorithm is effective in removing noise and the image is visually better. Recovering image after filtering with MDB algorithm is good in regard to recovered image without filtering. Area of image with watermark after recovering is still with damage. Comparing the quality of images with the watermark inserted in blocks size $k = 2$ and $k = 4$ can be concluded that the recovered images with $k = 2$ have better visible quality comparing to visible quality of images with $k = 4$.

Based on graph shown in Fig. 4. can be concluded that MDB algorithm for filtering impulsive noise is very effective for deonising visible watermarked image. For noise level $p = 10$ %, and block size of inserting watermark bit $k = 2$, MSE for recovered image with noise is exciding value 1000, on the other side after applying MDB denoising algorithm MSE is near the value 100. This shows, that using the MDB algorithm MSE has been decreased 10 times. Based on Fig. 5. for $p = 10$ % and $k = 2$, PSNR for image with noise is decreasing from value 30.66 to 14.99, on other size after applying MDB algorithm PSNR is decrease on value 25.50.

IV. Conclusion

In this paper we has analyzed VWM algorithm for inserting visible watermark, and MDB algorithm for filtering image. For the purpose of experimental analysis of the efficiency of VWM algorithm for recovering image, the image of the inverse binary watermark was used. The watermark is inserted with the coefficient of insertion $\upsilon = 40$, in blocks size $k = \{2, 3, 4\}$. Noise is varied in the range $p = \{1, 5, 10, 20, 30, 40, 50, 60, 70\}$. After comparing the visual appearance of the recovered image with superimposed noise and recovered image from filtered image, it is concluded that the recovered image from the filtered image, is a better quality. Also, the quality measures of MSE and PSNR shows that the recovered image, after application of the MDB algorithm is much better quality.

On the basis of the obtained results, it is concluded that the MDB algorithm successfully eliminates noise from the visible watermarked image.

References

[1] M. S. Kankanhalli, Rajmohan, and K. R. Ramakrishnan, "Adaptive Visible Watermarking of Images," in *Proc. IEEE Int. Conf. Multimedia Comput. Syst.*, vol. 1. Florence, SC, Jul. 1999, pp. 568–573.

[2] B. B. Huang and S. X. Tang, "A contrast-sensitive visible watermarking scheme," *IEEE Multimedia*, vol. 13, no. 2, pp. 60–67, Apr.-Jun. 2006.

[3] R. Lukac and K. N. Plataniotis, "Secure single-sensor digital camera," *Electron. Lett.*, vol. 42, no. 11, pp. 627–629, May 2006.

[4] C. H. Huang and J. L. Wu, "Attacking visible watermarking schemes," *IEEE Trans. Multimedia*, vol. 6, no. 1, pp. 16–30, Feb. 2004.

[5] S. C. Pei and Y. C. Zeng, "A novel image recovery algorithm for visible watermarked images," *IEEE Trans. Inf. Forens. Security*, vol. 1, no. 4, pp. 543–550, Dec. 2006.

[6] Y. Yang, X. Sun, H. Yang, and C.-T. Li, "Removable visible image watermarking algorithm in the discrete cosine transform domain," *J. Electron. Imaging*, vol. 17, no. 3, pp. 033008-1–033008-11 Jul.–Sep. 2008.

[7] C.C. Chen, Y.H. Tsai, H.C. Yeh, "Difference-expansion based reversible and visible image watermarking scheme", *Multimedia Tools Appl.* 76 (6) (2017) 8497–8516.

[8] B. Prlincevic, Z. Milivojevic and D. Brodic, "Efficiency of MDB algorithm for filtering watermarked images", in *Proceeding IT'14, Zabljak*, pp.244-247.

Teaching Methods for Digital Phase Locked Loop

Dimiter Badarov[1] and Georgy Mihov[2]

Abstract –**This article discusses the methodology, the tasks, the material base and the problems in the teaching process of all digital phase locked loop. Different types of phase and phase-frequency detectors are examined. Their parameters are compared and the advantages and drawbacks for different applications are discussed. Different types of digital integrators and digital filters are developed and tested along with the different phase-frequency detectors. Different circuits for Digitally Controlled Oscillator are developed and tested. Several different phase locked loops are built and tested. Their parameters are compared and discussed.**

Keywords –**All Digital Phase Locked Loop, Teaching methodology, Phase-frequency detector, Digitally Controlled Oscillator, FPGA.**

I. Introduction

The application of Phase Locked Loop (PLL) in the modern electronics is very wide. Along with the Direct Digital Synthesis (DDS) it is one of the main methods for generating clock frequencies. One of the advantages of the PLL is that it uses one reference clock frequency. The stability of the generated frequencies is equal to that of the reference frequency. The wide application of the phase locked loops requires more detailed examination. While the standard analog PLL are well examined and their development is almost finished, their digital equivalent can offer some advantages which require more attention. All structural blocks of the analog PLL are replaced with their digital functional equivalents. The transition to the digital electronics gives the advantages of easy testing using programmable logic devices and lowers the cost of the system. Although the development of the digital equivalents of the analog blocks offers some freedom for different solutions with better properties. Due to that a different solution for every block is developed, and a series of tasks are proposed for the education:

- The dependence of the digital code at the output of the digital integrator from the phase difference between the two input signals is examined for the different phase detectors.
- Different types of digital integrators and filters are proposed and the PLL stability is examined with them. The advantages of the different solutions are discussed.
- The different methods for Digitally Controlled Oscillators (DCO) construction are examined. The advantages for the different applications are discussed.

All of the proposed schematics are implemented and tested on Field Programmable Gate Array(FPGA)device Spartan 3A by Xilinx. This gives flexibility for quick testing of the different solutions.

[1]Dimiter Badarov is with the Faculty of Electronic Engineering and Technologies at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: dbadarov@tu-sofia.bg

[2]Georgy Mihov is with the Faculty of Electronic Engineering and Technologies at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria.

II. Structural Diagram of the All Digital Phase Locked Loop

The structural diagram of the test setup is represented on Fig. 1 [1]:

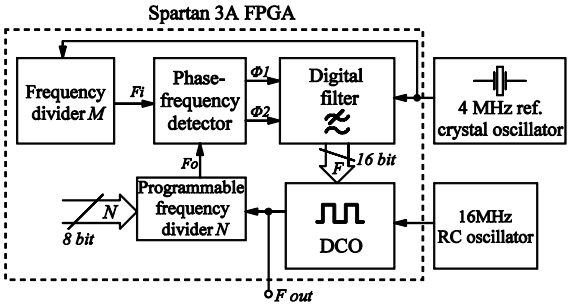


Fig. 1. Structural diagram of the test setup

A development board with FPGA programmable device Spartan 3A by Xilinx is used. It has USB interface for configuration of the FPGA and for power supply. It also has 4 capacitive touch sensitive buttons, 16 *MHz* RC oscillator and expansion connectors. An expansion board with 4 *MHz* quartz oscillator is developed for the different circuits analysis [1]. A digital USB oscilloscope Velleman PCSU1000 is used for examination of most of the proposed circuits and tasks. The basic block diagram of the developed PLL is drawn inside the FPGA chip.

III. Phase Frequency Detectors

There are different types of phase detectors. Their main differences are the width of the linear region in the phase response, and the frequency difference sensitivity. The detectors which are sensitive to the difference of the frequencies of the input signals are called phase-frequency detectors [2]. The other difference is the triggering event. The simplest phase detectors are insensitive to the difference in the frequencies of the input signals and are level triggered. These are called type 1 detectors. The input signals should be square wave and with the same 50% duty cycle [2]. Their linear range of the phase response is relatively narrow and repeats periodically. These detectors are used in PLL systems with narrow frequency range due to the risk of synchronization on multiple of the reference frequency. These types of detectors are most commonly built with an exclusive OR gate (XOR). The logical diagram with the timing diagrams and the phase response are represented on Fig. 2 [2]. The output of the

detector Φ is connected to an analog filter for demonstration of the phase response.

The schematic diagram of the detector with the analog filter is represented in *a*. The timing diagrams of the input signals and the Φ signal is represented in *b*. The phase response is represented in the form of voltage after the analog filter in *c*.

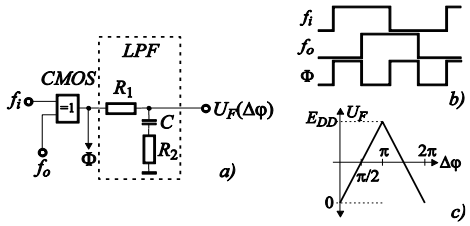


Fig. 2. XOR phase detector with LPF

As it can be seen from the diagrams the output frequency is twice the frequencies of the input signals. The phase response is with triangular shape and linear region from 0 to π. In order to work in the middle of the linear range the phase difference between the input signals should be π/2. In case of input signal loss, the detector works in the middle of the linear region [2].

The next type of phase detectors are edge triggered. These are also called type 2 detectors. In this type of detectors, the duty cycle of the input signals does not need to be 50% [3]. Its logical diagram along with the timing diagrams and the phase response is represented on Fig. 3:

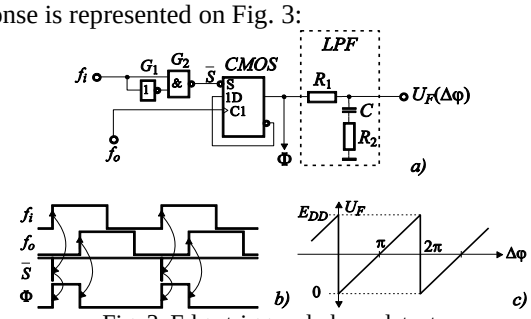


Fig. 3. Edge triggered phase detector

The triggering event in this detector is the rising edge of the input signals. The f_i signal is applied to a forming circuit which generates a short negative pulse after each rising edge of the signal [3]. This pulse sets the flip-flop. The f_o signal is applied to the clock input of the flip-flop and resets it when the rising edge arrives. As it can be seen from the timing diagrams the output frequency Φ is equal to the frequency of the input signals. The phase response is with sawtooth form and linear region from 0 to 2π. To work in the middle of the phase response, the steady phase difference between the input signals should be π. If the input signal f_i is lost the detector works in the middle of the linear region of the phase response.

The most serious drawback of these two-phase detectors is the relatively narrow linear range of the phase response. This leads to a worse interference immunity as the interference injected to the input of the detector can easily get the phase difference outside of the linear region [4].

The more advanced phase detectors are sensitive to the frequency difference of the input signals.

In their phase response there is a constant component outside of the linear range [2]. This gives the ability to tell which one of the input frequencies is higher. These detectors are used in PLL systems with wider working frequency range. On the startup the detector gives information what is the difference of the two frequencies and the proper code for the DCO synchronization is generated. These detectors are having the widest linear range in the phase response. The logical diagram of such detector is represented on Fig. 4. Its output signals are connected to a charge pump and a passive integrator for demonstration of the phase response [5].

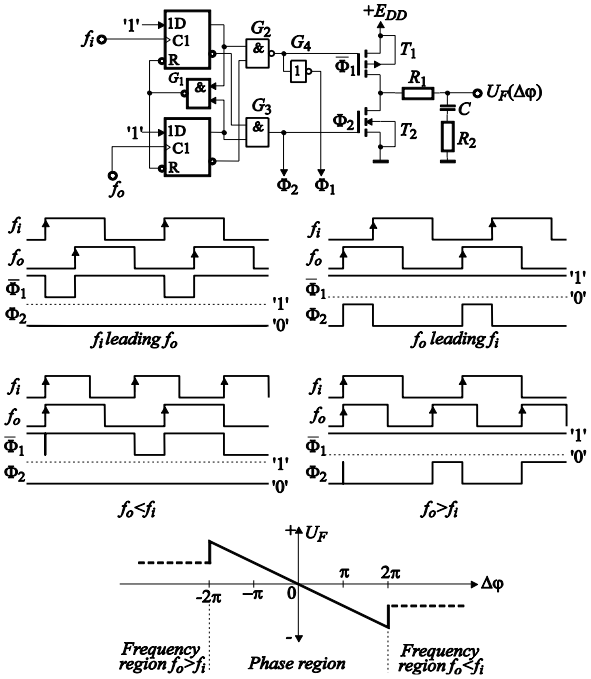


Fig. 4. Edge triggered phase-frequency detector

The compared signals f_i and f_o set on their rising edge two D-type flip-flops. The logical element NAND (G1) generates low level when both of the flip-flops are in set condition. This low level resets the flip-flops simultaneously. The output of the first flip-flop (Φ1 signal) is in high state if f_o leads f_i. The duration of the high state is equal to the duration between the rising edges of the two signals. If the f_i signal leads f_o the output of the second flip-flop (Φ2 signal) is high for the time difference between the rising edges. In the diagram there is shown a schematic of a charge pump and integrator for the demonstration purposes of the phase response. The detector has linear range of the phase response form -2π to 2π. In order to work in the middle of the linear range the phase difference between the input signals should be 0.

IV. Digitally Controlled Oscillators

In the analog PLL a main block is the Voltage Controlled Oscillator (VCO). It receives analog voltage and changes its output frequency depending on that voltage.

It has good parameters, but its price is high and in most of the cases it takes too much space on the Printed Circuit Board (PCB). It is also difficult for integration due to the large inductor on the chip [7]. For that reason, in the all-digital PLLs it is replaced by Digitally Controlled Oscillator. In it, the output frequency is controlled by a binary code. This

function can be achieved by several different methods. The first method is connecting a Digital to Analog Converter to a standard Voltage Controlled Oscillator [2]. It is not very desirable because we keep the high price and the difficulty for integration of the standard VCO. The advantage of the method is the easily controlled frequency step. The other more desirable and frequently used solution in the digital PLL is the high frequency square wave oscillator followed by a programmable frequency divider Fig. 5. It has the advantages of easy integration, lower price and higher stability. The disadvantage is the relatively high frequency of the oscillator required for a small enough frequency step. This leads to higher energy consumption.

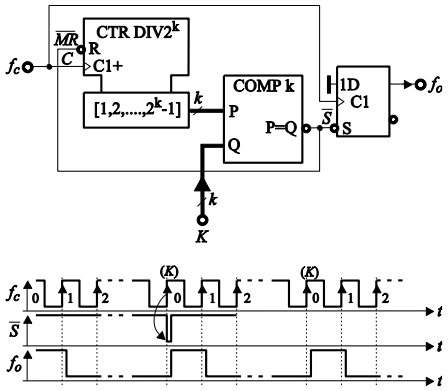


Fig. 5. Programmable frequency divider

The main problem of the concept is the increase of the frequency step when lowering the division coefficient *K*. A way to correct this is to connect a fixed frequency divider before the programmable one [8].

V. Digital Integrators and Filters

In the all-digital PLLs are used the same phase comparators as in the analog systems. The difference is that the analog filter or integrator is replaced with digital integrator or filter. This way the digital control code for the DCO is generated. Here the main logical diagrams and concepts will be discussed. The most common method for converting the time signals at the output of the phase detector to code is with digital integrator. The logical diagram of the integrator is represented on Fig. 6.

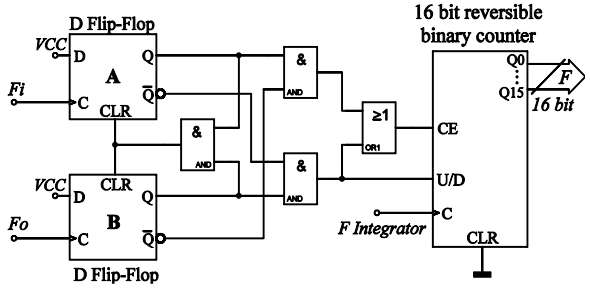


Fig. 6. Phase-frequency detector with digital integrator

It is built with a reversible counter with Clock Enable input and an Up/Down input. The clock frequency applied to the counter *F Integrator* is the equivalent of the time constant of the integrator.

By changing that frequency, the time constant can be changed. The outputs of the phase frequency detector are connected to the Count Enable and the Up/Down inputs of the counter. This way the integrator accumulates n pulses of *F Integrator* the time between the rising edges of the compared signals. The direction of accumulation is determined on which of the signals is leading. This way at the output of the counter is formed a digital code for the DCO [9].

Another version for conversion of the output signal of the phase detector Φ to digital code is proposed on Fig. 7:

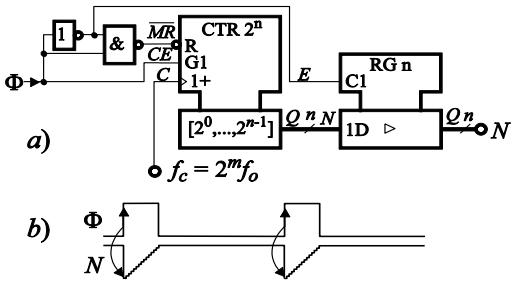


Fig. 7. Time-to-digital converter with parallel register

On the rising edge of the signal Φ the counter is being cleared and enabled and starts counting the pulses of the clock frequency f_c. On the falling edge of the signal the counter is disabled and stops counting [10].The accumulated value is stored to the parallel register after it [2]. This way the control code for the DCO is formed based on the phase difference of the signals. The main difference from the digital integrator is that the value of the counter here is cleared before each counting cycle. That means that in this case there is a fixed relation between time difference and digital code.

In some cases, for the stable operation of the PLL it is required additional filtration of the digital signal before applying to the DCO. It is possible to implement a standard Finite Impulse Response (FIR) filter into the FPGA Fig. 8 [2]:

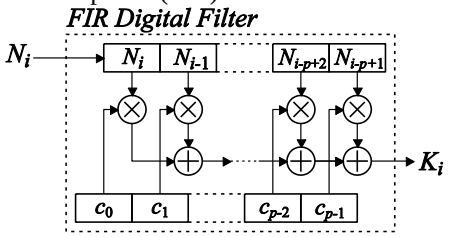


Fig. 8. Digital FIR filter

The logical diagram of the filter consists of a parallel shift register *N* for the samples from the digital integrator or filter. When a new sample enters the register from the left, all the samples inside are shifted with one position to the right. Every sample is multiplied by the corresponding coefficient *c* and the results are summed in order to form the current output sample K_i which is applied to the DCO [6]. One of the common realizations of the principle is the pipelined architecture Fig. 9. The circuit consists of *p* parallel registers connected in series forming one shift register. Before entering the register, the value N_i is divided by *p* into the DIV block.

The N_{i-p}/p sample is subtracted from the current filtered value K_{i-1} into the SUB block. At its output the value $K_{i-1} - N_{i-p}/p$ is formed. The value N_i/p is added to it into the SUM

block. The current sample of the output filtered signal $K_i = K_{i-1} - N_{i-p}/p + N_i/p$ is stored into the output parallel register.

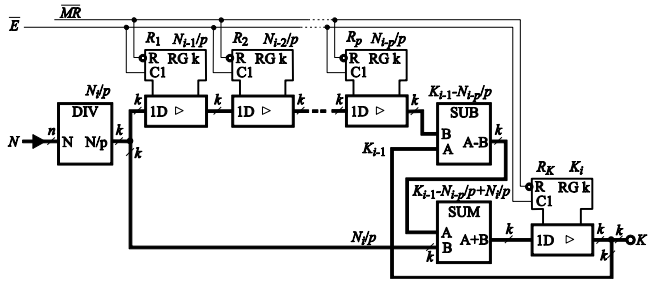


Fig. 9. Digital FIR filter with pipelined architecture

The division of the input samples is easily obtained if the division coefficient p is a power of two. The input sample is just shifted p bits toward the lower significant bits while the more significant bits are loaded with zeroes Fig. 10:

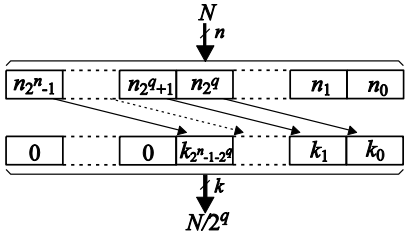


Fig. 10. Division by 2^q coefficient

The impulse response and the amplitude-frequency response of the filter with $p = 16$ is shown on Fig. 11:

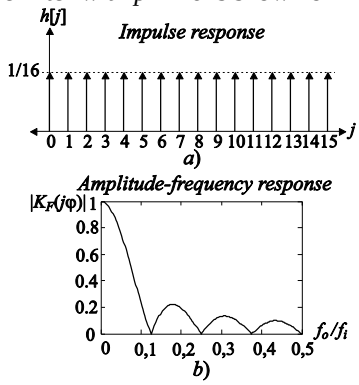


Fig. 11. Impulse and amplitude-frequency response of the filter

VI. Tasks for the Training

Several different tasks are developed for the education on all digital PLL:

- A digital PLL is built with phase comparator with XOR gate followed by digital integrator. The digital code from the integrator is applied to the DCO built with a programmable frequency divider. The circuit is tested with the digital oscilloscope, and the starting frequency of the DCO from which a successful synchronization is achieved is examined. The stability of the generated signal is observed, and the results are discussed.

- In the PLL from the previous task the phase detector is replaced with the phase-frequency detector from Fig. 7. The starting frequency of the DCO from which a quick synchronization is achieved is tested. The stability of the generated signal is observed. The results are discussed.

- To the PLL is added the digital FIR filter from Fig. 9. The system is tested for different values of multiplication of the reference frequency. The stability of the generated signal is observed with the digital oscilloscope. The time for synchronization for different frequencies is observed. The results are discussed.

VII. Conclusion

The developed digital functional equivalent modules for PLL give the students the opportunity to build all digital Phase Locked Loop on programmable logic device. The developed tasks offer the students the ability to design several different PLL topologies and to compare the performance of the different concepts. This way they can gain the knowledge on the PLL basics with the role of every block for the different PLL final parameters.

The use of FPGA programmable logic device gives the opportunity for every student to test a different circuit combination and even personal ideas for circuit improvements.

References

[1] D. Badarov, G. Mihov, "Development and Implementation of Digital Phase Locked Loop on Xilinx FPGA", Proceedings of XXV International Scientific Conference Electronics, pp. 80-83, 2016.

[2] G. Mihov, "Phase and frequency synchronizers", Digital Electronics, Technical University - Sofia, pp. 201-219, October 2010.(in bulgarian)

[3] S.R. Al-Araji, Z.M. Hussain, M. A. Al-Qutayri, "Digital Phase Lock Loops Architectures and Applications", Springer US, pp. 15-30, 2006.

[4] Y. R. Shayan, T. Le-Ngoc, "All digital phase-locked loop: concepts, design and applications", IEE Proceedings, Vol.136, Pt. F, No. 1, pp. 1-59, February 1989.

[5] V. Kratyuk, P. K. Hanumolu, U. Moon, K. Mayaram, "A Design Procedure for All-Digital Phase-Locked Loops Based on a Charge-Pump Phase-Locked-Loop Analogy", IEEE Transactions on Circuits and Systems - II: Express Briefs, Vol. 54, No. 3, pp. 247-251, March 2007

[6] M. Saber, Y. Jitsumatsu, M. T. A. Khan, "A Simple Design to Mitigate Problems of Conventional Digital Phase Locked Loop", Signal Processing: An international journal (SPIJ), Volume (6), Issue (2), pp. 65-77, 2012

[7] A. Djemouai, M. A. Sawan, M. Slamani, "New Frequency-Locked Loop Based on CMOS Frequency-to-Voltage Converter: Design and Implementation", IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS - II: Analog and Digital Signal Processing, vol. 48, No. 5, pp. 441-449,May 2001.

[8] C. Albea, D. Puschini, S. Lesecq, E. Beigné, P. Vivet. "Architecture and Control of a Digital Frequency-Locked Loop for Fine-Grain Dynamic Voltage and Frequency Scaling in Globally Asynchronous Locally Synchronous Structures", Journal of Low Power Electronics, American Scientific Publishers, pp.328-340, 2011

[9] D. Fernandez, S. Manandhar, "Digital Phase Locked Loop", pp. 1-59, December 2003.

[10] Xilinx Spartan-3A and Spartan-3A DSP Libraries Guide for Schematic Designs. UG614, (v 13.1), March 2011.

The Optimization of the STOI Algorithm Parameters in Presence of the White Gaussian Noise (WGN)

Zoran Milivojević[1], Dijana Kostić[1], and Zoran Veličković[1]

Abstract –**The first part of the paper describes the *Short Time Objective Intelligibility* - STOI algorithm, which makes an objective evaluation of intelligibility, as well as an algorithm for estimating the optimal parameters of the STOI algorithm, N and β. Second part of the paper, described an experiment that examines the intelligibility of the sentences formed from the *Serbian Matrix Sentence Test* - SMST base using: a). subjective MOS (*Mean Opinion Score*) test and b). objective test STOI algorithm. Subsequently, a comparative analysis of the results were made and it was determinate an analytical formula which connecting optimal pairs (*N*, *β*). Using the results of the MOS test as a reference, a mean absolute error of estimation of the STOI algorithm was determined.**

Keywords – **Intelligibility, Objective Measure, STOI, MOS.**

I. INTRODUCTION

When designing audio systems for speech transfer, one of the most important features should be predicted - intelligibility of speech, taking care about one-third octave band. In speech technology, term "intelligibility", is ranging from 0 to 100% intelligibility. Intelligibility of phoneme identification by listeners is a direct measure of intelligibility. The intelligibility of speech is the property of a speech signal that can be measured using: subjective and objective methods.

Subjective methods of evaluation intelligibility are realized with the MOS test. These tests are often long-lasting and cost very well. They are realized using different sets of word (logatoms [1], Phonetic Balansed [2]...) and sentences (everyday [3] and matrix [4]).

Objective methods for evaluation intelligibility implies speech tests are carried out using a computer. Parameters such as: AI [5], SII [6], STI [7], CSTI [8] are used for testing. They are suitable for evaluation of intelligibility of speech for several types of degradation: reverberation, additive noise, filtering and clipping.

AI is an articulation index developed in the AT & Bell Labs [5]. Based on AI the speech intelligibility index - SII was developed [6]. Following the index SII, Houtgast and Steeneken developed a speech transfer index - STI [7], which can predict the intelligibility of reverberated speech and nonlinearity of distortion.

However, if we want to make an objective evaluation of intelligibility using a method where the disorder signal is processed by some kind of time-frequency variable (TF) gain function, these are not appropriate methods.

In the paper [9], STOI algorithm for objective testing of sentences was demonstrated. The work is based on the analysis of the speech signal in the time frequency domain (TF). The efficiency of the STOI algorithm was tested under the conditions of a superposed noise: car, caffe and bottle.

In this paper, the effectiveness of the STOI algorithm was analyzed in evaluation the intelligibility of sentences spoken in the Serbian language in the presence of the variable level of the White Gaussian Noise (WGN). The authors created an algorithm for determining the optimal parameters of the STOI algorithm (Nopt, βopt). The experiment was realized at the Collage of Applied Science in Niš in which it was performed: a). subjective MOS test and b). an objective test using the STOI algorithm. The experiment was realized for SNR = {- 7, -5, -2, 0, 2} dB. The tests were implemented on students of the Collage of Applied Tehnical Science of Niš, gender structure: 10 male and 10 female, age from 19 to 26 years. The subjective MOS test determined the intelligibility of sentences spoken in the Serbian language and was used as a reference for the purpose of comparing the results obtained with the STOI algorithm. Comparative analysis of the results determined the optimal parameters of the STOI algorithm Nopt, βopt, as well as their analytical dependence. In the end, a mean absolute error of estimation of the STOI algorithm was determined, for measuring the results obtained by the MOS test .

The organization of work is the follow. Section II describes the STOI algorithm. Section III describes the algorithm for evaluation of optimal parameters. Section IV describes an experiment of evaluation intelligibility and performes comparative analysis of the results. Section V is a conclusion.

II. STOI ALGORITHM

The STOI algorithm is described in [9] and consists of the following steps:

Input: x - clean speech signal, y - speech signal with superimposed noise,

Output: d - intelligibility.

Step 1: TF decomposition of the x and y signals.

Step 2: Forming frames of length 256 and modification using the Hann - windowed.

Step 3: Removing the silence regions.

Step 4: Find the frame with the maximum energy of signal x.

Step 5: Finding the k^{th} DFT-bin of the m^{th} frame. The standard of one-third octave is defined as:

[1]Zoran Milivojević is with the College of Applied Technical Sciences in Niš, Aleksandra Medvedeva, Niš 18000, Serbia, E-mail: zoran.milivojevic@vtsnis.edu.rs.

[1]Dijana Kostić is with the College of Applied Technical Sciences in Niš, Aleksandra Medvedeva, Niš 18000, Serbia, E-mail: koricanac@yahoo.com

[1]Zoran Veličković is with the College of Applied Technical Sciences in Niš, Aleksandra Medvedeva, Niš 18000, Serbia, E-mail: zoran.velickovic@vtsnis.edu.rs.

$$X_j(m)=\sqrt{\sum_{k=k_1(j)}^{k_2(j)-1}\left|\hat{x}(k,m)\right|^2}\,, \qquad (1)$$

where k_1 and k_2 represented the boundaries of one-third octave.

Step 6: Forming a short-time envelope vector of clean signal:

$$x_{j,m}=\left[x_j(m-N+1), X_j(m-N+2),..., X_j(m)\right]^T, \qquad (2)$$

where is N = 30, which corresponding to a frame length of 384 ms, for fs = 10 kHz.

Step 7: The normalization and clipping $\boldsymbol{y}$ signal

$$\bar{y}_{j,m}(n)=\min\left(\frac{\|x_{j,m}\|}{\|y_{j,m}\|}y_{j,m}(n),\left(1+10^{-\beta/20}\right)x_{j,m}(n)\right), \qquad (3)$$

where is β = -15 dB.

Step 8: Defining a measure of intelligibility $d_{j,m}$:

$$d_{j,m}=\frac{\left(x_{j,m}-\mu_{x_{j,m}}\right)^T\left(\overline{y_{j,m}}-\mu_{\overline{y_{j,m}}}\right)}{\left\|x_{j,m}-\mu_{x_{j,m}}\right\|\left\|\overline{y_{j,m}}-\mu_{\overline{y_{j,m}}}\right\|}, \qquad (4)$$

Step 9: The average measure of intelligibility for all bands and frames is d:

$$d=\frac{1}{JM}\sum_{j,m}d_{j,m}\,, \qquad (5)$$

were M represent total number of frames, and J number of one-third octaves.

III. ALGORITHM FOR EVALUATION OF OPTIMAL PARAMETERS

The algorithm for determining the optimal parameters of the STOI algorithm is realized in the following steps:

Input: SNR_{min}, SNR_{max}, ΔSNR, β_{min}, β_{max}, $\Delta\beta$, N_{min}, N_{max}, ΔN, N_r.

Output: trajectory of optimal values y_{fit}.

Step 1: Created the test of sentence, with fixed syntax word structure from the SMST base (signal x_r)

FOR $SNR = SNR_{min} : \Delta SNR : SNR_{max}$

FOR $r = 1 : N_r$

Step 2: Superposed White Gaussian noise, n (σ^2=1, μ=0)

$$x_{1r}=x_r+k_r\cdot n, \qquad (6)$$

where k_r is the coefficient determining the SNR of the speech signal x_{1r}.

Step 3: Realization of a subjective MOS test and generating a evaluation of intelligibility MOSxr.

FOR $\beta = \beta_{min} : \Delta\beta : \beta_{max}$

FOR $N = Nmin : \Delta N : Nmax$

Step 4: Application of STOI algorithm

$d_r(N, \beta)$ = STOI (N, β)

END N;

END β;

END r;

Step 5: Forming of equivalent matrix of intelligibility coefficient:

$$d(N,\beta)=\overline{d_r(N,\beta)}, \qquad (7)$$

Step 6: The mean value of coefficient for STOI intelligibility is:

$$d_{STOI}=\overline{d(N,\beta)}, \qquad (8)$$

and statistical parameters are mean value $\mu=d_{STOI}$ and variance σ^2.

Step 7: Forming matrix error of evaluation:

$$e=\left|d-MOS_{xr}\right|, \qquad (9)$$

Step 8: Generating a trajectory of minimum error:

$$y_k = f(N_{opt}, \beta_{opt}), \qquad (10)$$

Step 9: Generating an equivalent trajectory of optimal values by fitting method:

$$y_{fit}=\beta_{fit}=a_3N^3+a_2N^2+a_1N+a_0\,, \qquad (11)$$

where is a_3=-4.927*10^{-5}, a_2=0.01177, a_1=-0.9856, a_0=2.942.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, an experiment was implemented in order to find the optimal parameters of the STOI algorithm.

A. Experiment

In order to test the effectiveness of evaluation using the STOI algorithm, experiment was implemented at the Collage of Applied Science Niš, where is applied a algorithm described in Section III for the evaluation of intelligibility. An experiment was implemented in order to find the optimal parameters of the STOI algorithm. The input parameters are: a). SNR_{min} = -7dB, SNR_{max} = 2dB, ΔSNR = 2dB, b). N_r = 20, number of test sentences for each value SNR, c). N_{min} = 10, N_{max} = 100, ΔN = 10, β_{min} = -40, β_{max} = -5, $\Delta\beta$ = 5. The output parameters of the algorithm are: a). analytic function of optimal values β_{opt} = y_{fit}, b) the results of subjective test MOS_{xr}, c). the results of objective test d_{STOI} i d). statistical parameter: mean value μ and variance σ^2 for d_{STOI}. Results are shown in tabular and graphical form.

B. The base

A SMST base has been created which contains words (name, verb, number, adjective and object) spoken in Serbian language [10]. Using the algorithm described in section III

(step 1), sentences with a fixed syntax structure were formed. The words are formed by a random selection of the word from the SMST base to the syntax structure: name-verb-number-adjective-object. It is possible to create 100000 different sentences, which made test unpredictabil for the tested group.

C. Test group

A test group was formed from students of the Collage of Applied Science Nis, age from 19 to 26 (μ = 20.85), gender stucture: 10 male and 10 female respondents.

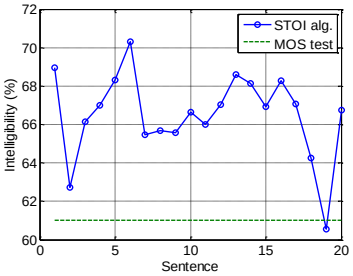
D. The results

Table 1 shows the results of MOS (step 3) and d_{STOI} (step 4) and statistical of mean values μ and variance σ^2 for various *SNR* values.

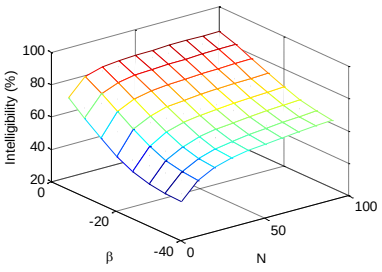
The values of the MOS and STOI coefficients for each sentence are shown graphically for: a) *SNR* = 0 dB (Figs. 1. a) and b) *SNR* = -7 dB (Figs. 2.a). The value of equivalent matrix coefficients $d(N, \beta)$, (step 5) are shown: a) *SNR* = 0 dB (Figs. 1.b) and b) *SNR* = -7 dB (Figs. 2.b). A matrix error of intelligibility evaluation e (step 7) are shown: a) *SNR* = 0 dB (Figs. 1.c) and b) *SNR* = -7 dB (Figs. 2.c). Trajectories of minimal error of intelligibility evaluation $y_k(N, \beta)$, (step 8) are shown on: a) *SNR* = 0 dB (Figs. 1.d) and b) *SNR* = -7 dB (Figs. 2.d). In Fig. 3 are shown: a) trajectory y_k minimal error for *SNR* = {-7, -5, -2, 0, 2} dB and b) equivalent trajectory y_{fit} (step 9). Intelligibility determinated with: a) subjective MOS test, MOS_{xr}, and objective test using STOI algorithm, d_{STOI}, are shown in Fig.4.

TABLE I

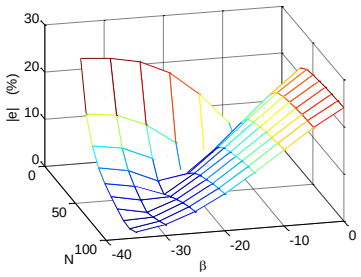
SNR (dB)	MOS_{xr} (%)	σ^2	μ(%)	d_{STOI}
2	71	4.4834	70.1975	70.1975
0	61	4.8871	66.5205	66.5205
−2	57	5.0465	61.5471	61.5471
−5	47	538.5059	45.0886	45.0886
−7	34	364.4125	44.0291	44.0291



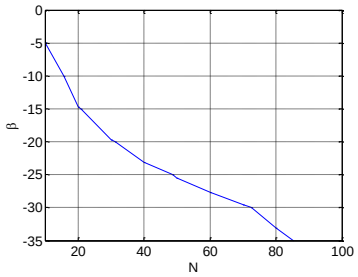
a)



b)

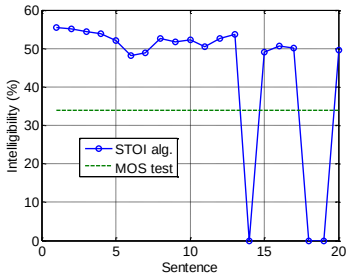


c)

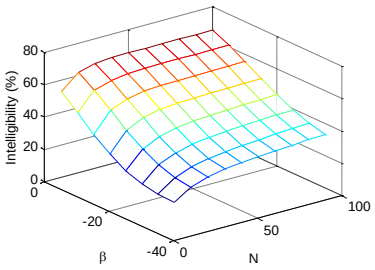


d)

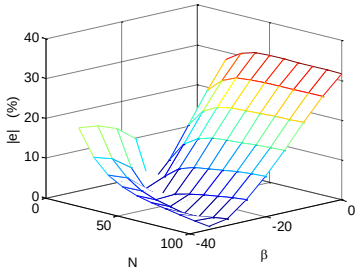
Fig. 1. SNR=0 dB: a) intelligibility of sentence x_r (r=1,…,20), coefficient of STOI algorithm and MOS evaluation, b) matrix of mean value d, c) matrix of error **e** and d) trajectory of minimal error of evaluation intelligibility y_k.



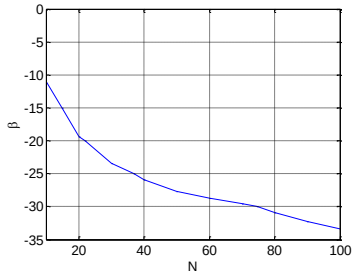
a)



b)



c)



d)

Fig. 2. SNR=-7dB: a) intelligibility of sentence xr (r=1,…,20), coefficient of STOI algorithm and MOS evaluation, b) matrix of mean value d, c) matrix of error e and d) trajectory of minimal error of evaluation intelligibility yk.

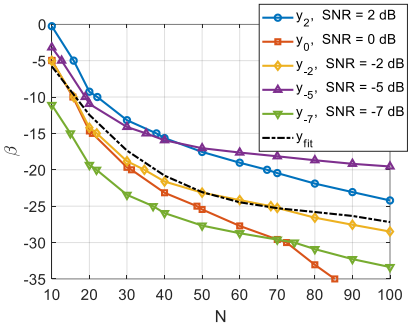


Fig. 3. Trajectry of minimal error of inteligibility evaluation for various levels of SNR

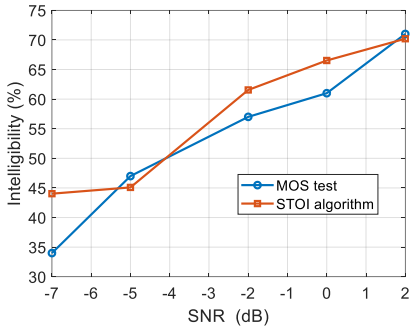


Fig. 4. Intelligibility determinated by: a) MOS test and b) STOI algorithm.

E. Analysis of results

By analysis of results shown in Table I and Figs. 1 ÷ 4, it can be concluded that:

a). there are many numbers of pairs (N_{opt}, β_{opt}), which can analytical described with Eq. (11) and graphically in Fig. 3 with trajectory;

b). taking into considering the results of the subjective MOS test as a reference, the mean absolute error of the intelligibility evaluation determined by the STOI algorithm (Table I, Fig. 4) is:

$$e_{SNR} = \frac{1}{5}\sum_{SNR=-7dB}^{2dB}\left|MOS_{XR}(SNR) - d_{STOI}(SNR)\right| = 4.5622\,\%$$

which is considered a good evaluation.

V. CONCLUSION

The paper presents an algorithm for determining the optimal parameters of the STOI algorithm in the presence of WGN. A detailed analysis of the results of the conducted experiment showed that there are an infinitely large number of values of optimal parameters and their analytical formula is determined. Taking into account the results of the subjective MOS test as a reference, it has been shown that the mean absolute error of evaluation intelligibility of the STOI algorithm is e_{SNR} = 4.5622%. This result points to the quality of the STOI algorithm and makes recommendations for its application when evaluating speech intelligibility.

REFERENCES

[1] D. Kostić, Z. Milivojević, V. Stojanović, "The Evaluation of Speech Intelligibility in the Orthodox Church on the Basis of MOS Test Intelligibility Logatom Type CCV", ICEST 2016, Ohrid, Macedonia, pp. 153-156, 2016.

[2] K. Kruger, K. Gough, P. Hill, “A Comparison of Subjective Speech Intelligibility Test in Reverberant Environments”, Canadian Acoustic vol 19 no 4, pp. 23 - 4, 1991.

[3] R. Plomp, A.M. Mimpen,” Improving the Reliability of Testing the Speech Reception Threshold for Sentences”, Audiology, vol 18, pp. 43 - 52, 1979.

[4] B. Hagerman, “Sentences for Testing Speech Intelligibility in Noise”, Scand Audio, vol. 11, pp. 79-87, 1982.

[5] N. R. French, J. C. Steinberg, “Factors Governing the Intelligibility of Speech Sounds”, J. Acoust. Soc. Amer., vol. 19, no. 1, pp. 90–119, 1947.

[6] Methods for Calculation of the Speech Intelligibility Index, S3.5-1997, ANSI, New York, 1997

[7] H. J. M. Steeneken, T. Houtgast, “A Physical Method for Measuring Speech-Transmission Quality”, J. Acoust. Soc. Amer., vol. 67, no. 1, pp. 318–326, 1980

[8] R. L. Goldsworthy, J. E. Greenberg, “Analysis of Speech-Based Speech Transmission Index Methods with Implications for Nonlinear Operations”, J. Acoust. Soc. Amer., vol. 116, no. 6, pp. 3679–3689, 2004.

[9] C. H. Tall, R. C. Hendriks, R. Heusdens, J. Jensen, ”An Algorithm for Intelligibility Prediction of Time-Frequency Weighted Noise Speech”, EEE Transactions on audio, speech, and language processing, vol 19, no. 7, september, 2011.

Z. Milivojević, D. Kostić, Z. Veličković, D. Brodić, “Serbian Sentence Matrix Test for Speech Intelligibility Measurement in Different Reverberation Conditions,” UNITEH Gabrovo II,*pp.* 173-178, 2016.

Correlation Analysis of Analog and Digital Signals in MATLAB Environment

Lyubomir Laskov[1], Veska Georgieva[2] and Kalin Dimitrov[3]

Abstract – **This paper presents an approach for studying the correlation analysis of analog and digital signals with various shape forms. An algorithm and its implementation in MATLAB environment have been developed. The implemented software allows studying the correlation of different analog and digital signals and the influence of signals parameters on its correlation functions. The presented approach can be used in engineering education for studying this process.**

Keywords – **signal processing, correlation analyses, auto correlation function, cross correlation function, computer simulation, MATLAB.**

I. INTRODUCTION

The correlation analysis of the signals is a basic part in the theory of the telecommunications. The correlation function quantifies the relation between two signals at different time points. Two signals are uncorrelated if their scalar product (their common energy) is zero. Depending on the signals, there are two types of correlation function: auto-correlation function and cross-correlation function. Auto-correlation function gives the relation of one signal at different moments of time. The cross-correlation function gives the relation between two different signals at different moments of time. Correlation analysis is widely used to detect signals in high level of noise or by low power signals. Another application of correlation analysis is packet synchronization, when the specific combination of bits is searched, used as marker for packet beginning.

The correlation analysis is studying in the course "Signals and Systems" at the undergraduate course in specialties "Telecommunications", "Computer and Software Engineering" and "Electronics" at Technical University of Sofia. During the seminar exercises, the students will have the opportunity to study the auto-correlation function (ACF) of different periodic and aperiodic analog and digital signals, as well as cross-correlation function (CCF) between some of them. Students will explore the influence of various parameters of the signals on their ACF and CCF.

We propose to use a program code in the MATLAB environment in order to generate signals with maximum number of different shape forms and to change as much as possible signal parameters [1]. The MATLAB environment has a definite advantage over other software products and other programming languages, because of built-in functions [2], [3]. This would considerably facilitate students in their work. Writing source code in MATLAB is preferable to the MATLAB Simulink because of the additional features that provide students with ability to change signals in a way that MATLAB Simulink does not provide [3], [4]. Writing the source code also leads to the requirement that the students will understand the functions describing the calculation of the correlation functions of signals.

[1]Lyubomir Laskov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: laskov@mail.com.

[2]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.

[3]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.

II. PROBLEM FORMULATION

The main tasks, solved in the correlation analysis of signals are associated with the variety of signals as types (analog and discrete, periodic and aperiodic) and shapes (sinusoidal, rectangular, triangular, sawtooth, Gaussian, etc.).

Auto-correlation function (ACF) shows the relation between a signal and its time-shifted copy. In the case of periodic analog signals it is defined as follows:

$$\psi(\tau) = \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} S(t)S(t-\tau)dt \qquad (1)$$

where $S(t)$ is the signal, and τ is a parameter showing the time slot on which the copy of the signal has been moved $S(t-\tau)$. For aperiodic analog signals, the expression for ACF has the following form:

$$\psi(\tau) = \int_{-\infty}^{\infty} S(t)S(t-\tau)dt \qquad (2)$$

In the cross-correlation function (CCF), the two signals, which are compared, are different, thus the CCF shows the relation between one signal $S_1(t)$ and time-shifted copy of another signal $S_2(t-\tau)$, where τ is a time shift interval. Cross-correlation function of periodic analog signals is defined as follows:

$$\psi_{1,2}(\tau) = \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} S_1(t)S_2(t-\tau)dt \qquad (3)$$

where $S_1(t)$ and $S_2(t)$ are signals whose relation is quantified. For aperiodic analog signal, CCF is described by the following expression:

$$\psi_{1,2}(\tau) = \int_{-\infty}^{\infty} S_1(t)S_2(t-\tau)dt \qquad (4)$$

If the signals are discrete, it is necessary to modify the above given expressions, by replacing the integral with summation and "τ" with "nT". The expression for the Discrete Auto-Correlation Function (DACF) for periodic signals is following:

$$\psi(nT) = \frac{1}{NT} \sum_{k=-\frac{NT}{2}}^{\frac{NT}{2}} S(kT)S\left[(k-n)T\right] \qquad (5)$$

where *S(kT)* is the signal, n is the number of samples, on which the copy of the signal has been shifted *S[(k-n)T]*, and N is number of samples for one period of the signal. For aperiodic signals, the expression for DACF has the following form:

$$\psi(nT) = \sum_{k=-\infty}^{\infty} S(kT)S\left[(k-n)T\right] \qquad (6)$$

Discrete Cross-Correlation Function (DCCF) for periodic discrete signals is defined by:

$$\psi_{1,2}(nT) = \frac{1}{NT} \sum_{k=-\frac{NT}{2}}^{\frac{NT}{2}} S_1(kT)S_2\left[(k-n)T\right] \qquad (7)$$

where S_1(kT) and S_2((k-n)T) are signals whose relation is quantified. For aperiodic signals, DCCF is defined as follows:

$$\psi_{1,2}(nT) = \sum_{k=-\infty}^{\infty} S_1(kT)S_2\left[(k-n)T\right] \qquad (8)$$

Often, instead of ACF and CCF, their normalized values are used. They represent the ratio of the current value of the ACF or CCF and its maximum value. Their advantage is that they reduce the impact of impulse noise on signal recognition.

The main tasks, which can be solved by computer simulations, are following:

1. Create an appropriate model for a presentation of the signals.
2. Create algorithms for calculation of the analog and digital ACF and CCF.
3. Investigate the influence of the waveforms and signal parameters on the ACF and CCF.
4. Present the obtained results graphically.

III. Basic Algorithm for Correlation Analysis

The block diagram of the proposed algorithm for correlation analysis of the analog and discrete signals is shown in Fig. 1.

Initially the global parameters and variables are set. Then it is necessary to select the type (analog or discrete) of the signals, whose ACF and CCF will be calculated. Signals with different shape forms have different specific parameters (basic parameters such as period, amplitude, etc. as well as specific parameters for the particular signal). Depending on the signals type, there are specific parameters to be set (for analog ones - the variables in which the signals expressions will be stored, and for digital ones - the number of samples, etc.). In addition to the type of signal, it is necessary to know whether it is periodic or not. Depending on this, different models of signals will be created and different expressions for determining ACF and CCF will be used.

Considering all the parameters of the signals, students must create a mathematical description of the researched signals (the studies of the following analog and discrete, periodic and aperiodic signals are envisaged: sinusoidal, rectangular, triangular, sawtooth, and Gaussian, as well as digital sequences).

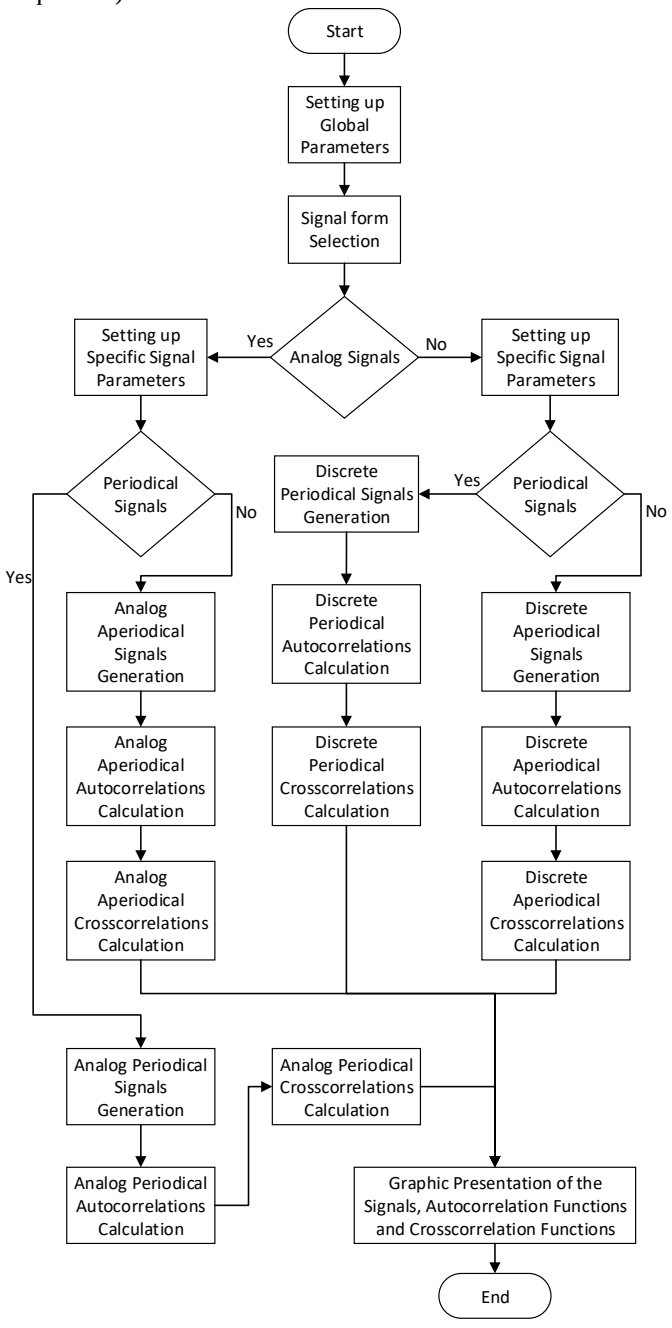


Fig. 1. Block diagram of the basic algorithm

Once the signal models are created, it is necessary to create an algorithm, which calculates the auto-correlation functions of both signals as well as the cross-correlation functions between them. For the different types of signals, the mathematical expressions describing ACF and CCF are different, so the used mathematical models will differ. The discrete part of the algorithm allows to be calculated ACF and CCF, not only for discrete signals but also for digital sequences.

The implemented software modules for calculation of the correlation functions of periodic and aperiodic signals allow creating a graphical representation of the signals, their auto-correlation and cross-correlation functions. This gives possibilities for investigation and visualization of the impact of certain signal parameters of the signal and its correlation functions.

IV. Experimental part

The formulated problems are solved by computer simulation in MATLAB environment.

Some results from simulations of non-periodic analog, periodic discrete and periodic digital signals are shown in next figures bellow. The single analog rectangular impulse and its auto-correlation functions are presented in Fig. 2. In the first row of Fig. 3, the two mentioned above signals are shown, assuming that the signal displayed in red is shifted over time. The cross-correlation functions, which is obtained by offsetting the corresponding signal is shown in the second row of Fig. 3.

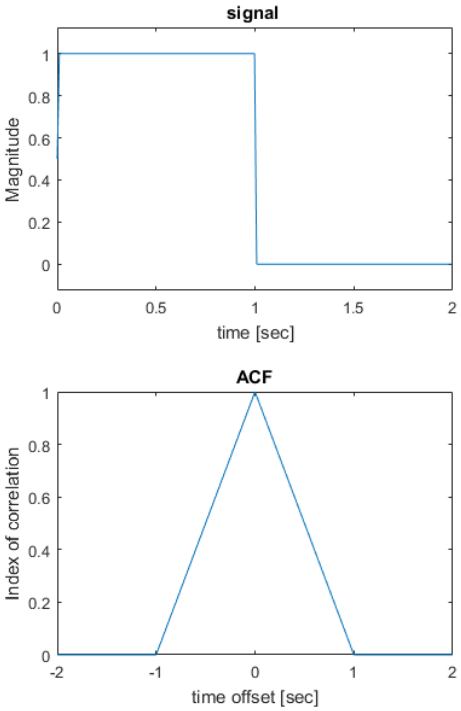


Fig. 2. Analog rectangular pulse and its ACF

The discrete sequence of rectangular impulses and its auto-correlation function are shown in Fig 4. Two discrete periodical signals (rectangular impulses and sawtooth impulses), where with red is marked the signal that is shifting in time are presented in the first row of Fig. 5. The second row of Fig. 6 shows the cross-correlation function when the corresponding signal is time shifted.

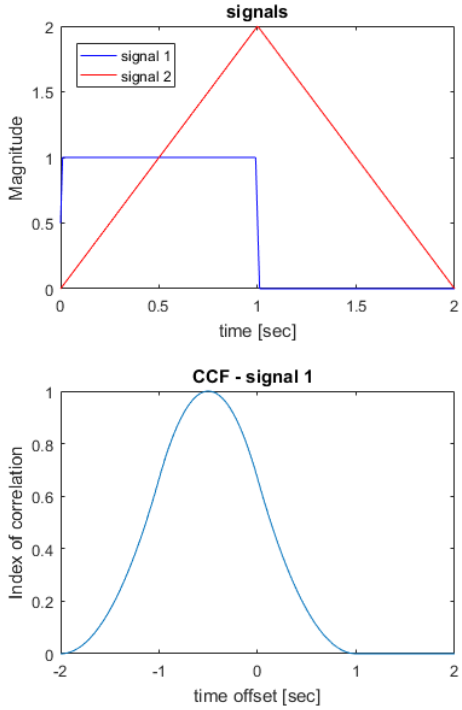


Fig. 3. Analog rectangular pulse and triangle pulse and their CCF

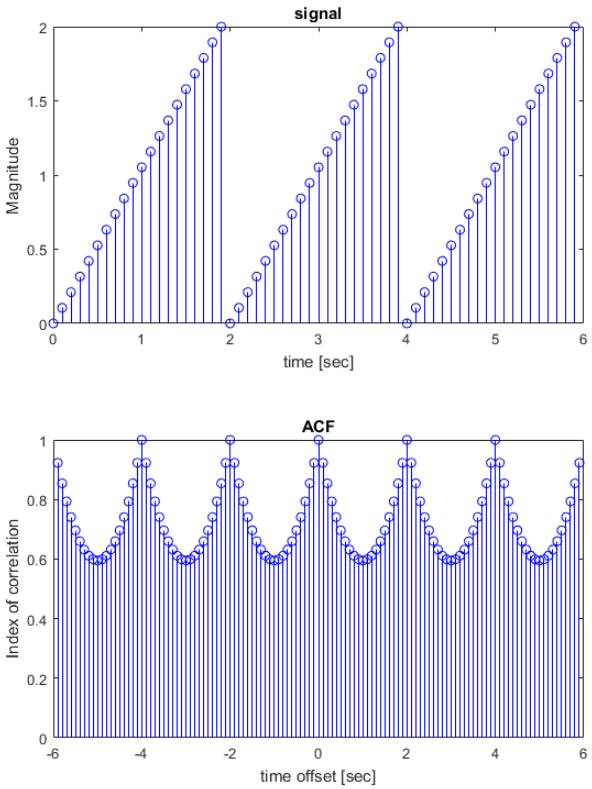


Fig. 4. Discrete sawtooth pulse train and its ACF

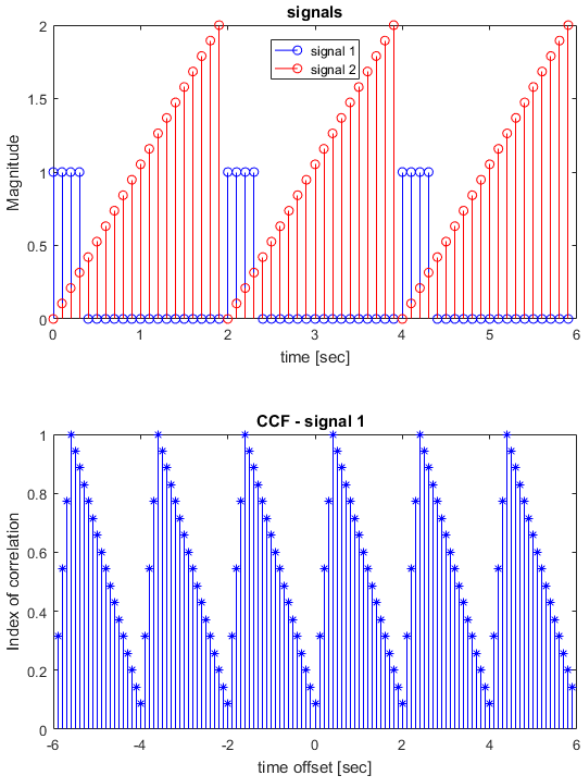


Fig. 5. Discrete rectangular pulse train and sawtooth pulse train and their CCF

In Fig. 6, a period of periodic digital sequence 10101111 and its auto-correlation function are shown. In the first row of Fig. 7, a period of the following two 8 bit digital sequences is shown: 10101111 and 01000011. The second row of this figure shows the cross-correlation function between the selected digital sequences, when shifted in time is second sequences (marked in red on the first row).

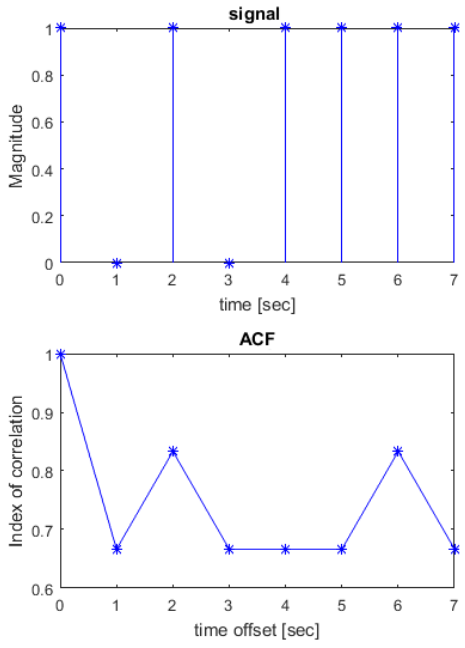


Fig. 6. Digital Auto-Correlation Function for digital sequence

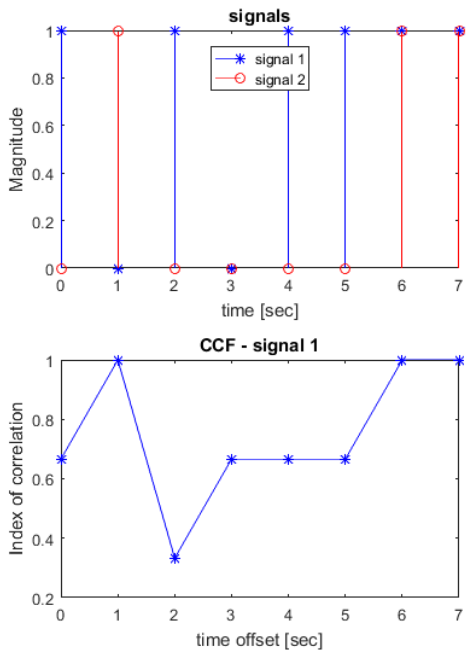


Fig. 7. Two digital sequences and their CCF

V. Conclusion

In this paper, we propose an approach for studying on correlation functions of periodic and aperiodic analog and digital signals. An algorithm and its implementation in MATLAB environment have been developed. The students can create models of signals with different type, shape forms and different parameters. They will be able to analyze influence of different signal parameters on the corresponding auto-correlation and cross-correlation functions. Our next work will be focused on further developing the proposed approach to conducting web distance learning exercises.

References

[1] MATLAB, User's Guide, www.mathwork.com
[2] L. Chaparro, *Signal and Systems using MATLAB*, Academic Press (Elsevier), 2011.
[3] S. Karris, *Signal and Systems with MATLAB; Computing and Simulink; Modeling*, Fifth edition, Orchard Publications, 2012.
[4] V. Georgieva, P. Petrov, *Signals and systems - manual for laboratory exercises*, King, 2016 (In Bulgarian).
[5] E. Ferdinandov, *Signals and systems*, Siela, Sofia, 1999 (In Bulgarian).
[6] S. Donevska, B. Donevsky, *Advanced Engineering Mathematics*, TU-Sofia, Sofia, 2014.
[7] G. Nenov, *Signals and systems*, Novi Znania, Sofia, 2008 (In Bulgarian).

A laboratory kit for an in-depth study of PLLs and PLL frequency synthesizers

Ludwig Lubich[1]

Abstract – **A laboratory kit for an in-depth study of PLLs is presented. It is mainly oriented (but not limited) to PLL applications for frequency synthesis. It can be used by students in wireless communications and electronics. Suggestions for demonstrations and experiments using the described laboratory kit are given.**

Keywords – **PLL, laboratory kit, laboratory session, education.**

I. Introduction

Phase-locked loops (PLLs) are widely employed in telecommunications, computers and other electronic equipment. They can be used for frequency synthesis, demodulation, clock recovery, distribution of precisely timed clock pulses in digital circuits, etc. [1]. Frequency synthesizers are indispensable in modern radio communications. Presently, the type of frequency synthesizers most used is based on PLLs [2]. Therefore, the good understanding of PLLs is an important educational aim in the education in electrical and telecommunication engineering.

Currently, computer simulations tend to replace the studying of physically realized devices. Although giving practically unlimited opportunities for studying, computer simulation inevitably leads to some degree of abstraction. This can have adverse effects on the understanding of the studied object and the students' motivation. Pedagogical experience shows that sometimes educators overlook the fact that students' abstract reasoning is not developed to the level, typical for a researcher. The result is mechanically performed actions without understanding their purpose and the practical significance of the results obtained. Without renouncing computer simulations in any respect, we reckon that the use of physically realized laboratory kits is not outdated and can be a valuable part of engineering education.

An example of a popular ready-made laboratory kit allowing the studying of PLLs (in addition to many other topics) is Analog System Lab Kit PRO from Texas Instruments [3]. Unfortunately, its capabilities for studying PLLs are too limited and the helpfulness of using it is questionable.

Many universities use their own laboratory kits or the students themselves build PLLs on breadboards, using popular PLL ICs, usually CD4046, as in [4]. Such laboratory kits typically have limited capabilities for experiments and are mainly oriented to low-frequency applications, whereas nowadays radiofrequency synthesis tends to be among the most important PLL applications.

Our goal was to develop a laboratory kit for an in-depth study of PLLs with a focus on their application in (but not limited to) frequency synthesis for radio communications. The ability for performing wide range of instructive experiments without needing expensive laboratory equipment was also among the primary considerations. The proposed laboratory kit has been implemented and used in the Department of Radio Communications an Video Technologies in the Technical University of Sofia.

In the next Section, the developed laboratory kit is presented. In Section III demonstrations and a laboratory session plan are suggested.

II. Laboratory kit description

A slightly simplified schematic diagram of the laboratory kit is presented in Fig. 1.

The core of the PLL is TSA5511, a popular synthesizer IC from NXP Semiconductors. Although it is relatively old, similar ICs are offered for new designs, for example ADF4002 [5] from Analog Devices. A very useful feature of TSA5511 for our purposes is its ability to route the two phase-frequency detector (PFD) inputs to two of the general purpose IC outputs/inputs.

The main loop filter is made up of the built-in amplifier, selectable RC networks for frequency dependent negative feedback and some additional components which are not shown for the sake of clarity. In positions 1 and 2 of switch S4, the recommended filter topology is selected. In position 1, the resistor is variable, whereas in position 2 it is fixed at a nearly optimal value and the capacitances are nearly an order of magnitude smaller than in position 1. In position 3, the minimum number of components that ensure an acceptable stability margin are used. In position 4, the phase margin is nearly zero (but the system is still stable). In positions 1-4, the loop filter acts as an analog memory retaining the steady state control voltage and its gain (theoretically) approaches infinity at DC, whereas in position 5 the loop filter acts as a usual low-pass filter having a finite gain from DC to a certain cut-off frequency. In order to demonstrate an unstable loop, an additional low-pass filter can be included by setting S2 in the upper position.

The voltage controlled oscillator (VCO) is implemented as an LC oscillator with a pair of varactors for frequency control. The tuning sensitivity can be chosen from two values. It is useful for some experiments to be able to change of the VCO free running frequency (FRF). Typically, in practical implementations, this is performed by adjusting the inductance in the oscillator LC tank. In our case this is not very convenient. For this reason, an auxiliary VCO control

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

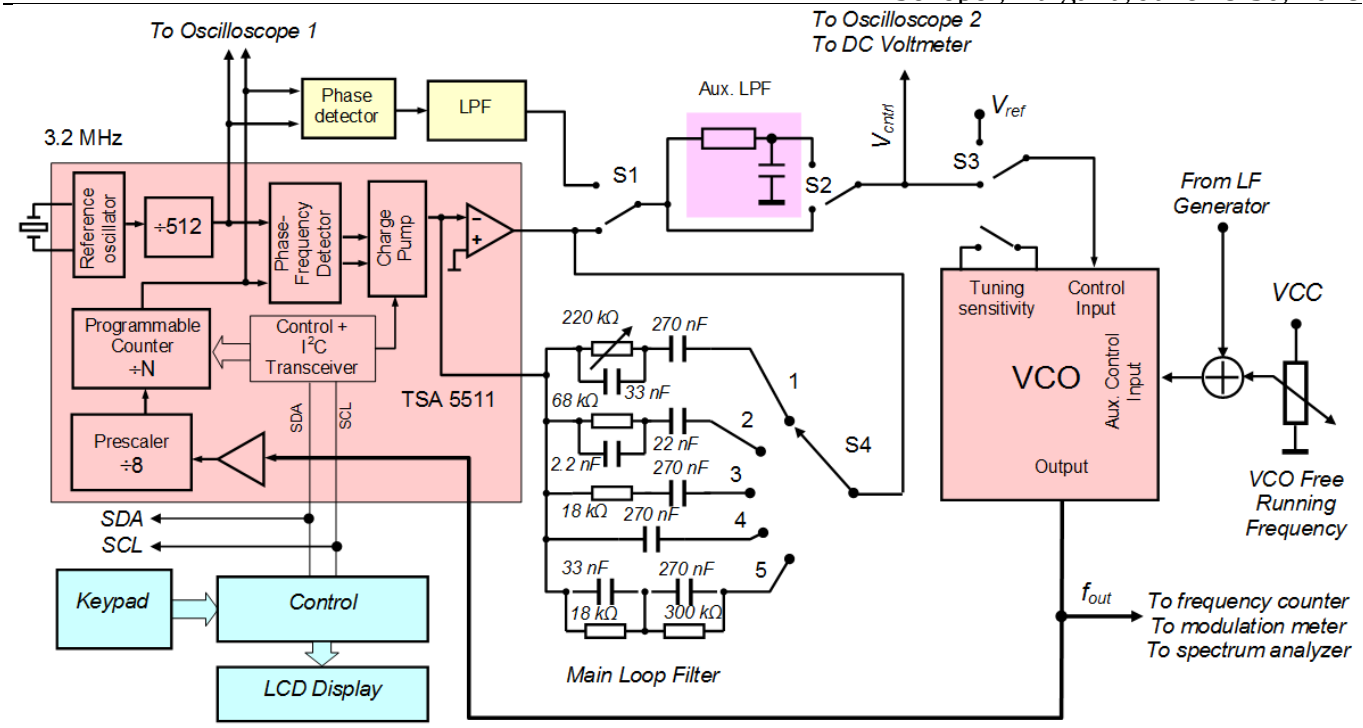


Fig. 1. Simplified schematic diagram of the laboratory kit

input is provided. It allows to change the VCO FRF in a relatively wide range. In addition, it allows to modulate the VCO FRF by an laboratory LF generator. This is useful for two purposes: (1) To examine the ability for frequency modulation of the synthesized oscillation and (2) To examine the ability of the PLL to suppress the VCO phase noise components at different offset frequencies. It is known that phase noise measurement is a rather complicated task for the students, which requires special theoretical training and/or expensive equipment [6], [7]. In our case, an intentional sinusoidal modulation of the VCO FRF (and the corresponding phase modulation) can imitate the phase noise components at particular frequency offsets. Since the modulation index can be set at a sufficiently high value, no sensitive equipment or advanced techniques are needed for the measurement. Naturally, the modulation index should not be too large to maintain the linearity.

In order to demonstrate the benefits of using PFDs in PLLs, an auxiliary path is implemented, in which a plain phase detector (PD) is employed. When S1 is in the upper position the PLL uses this path.

The control unit has the following functions: To enable the setting the frequency division ratio *N1* of the programmable counter; to enable the selecting of the charge pump current (TSA5511 offers 50 μA and 220 μA); to enable the setting of a second division ratio *N2* and automatically alternate the programmable counter division ratio between *N1* and *N2* in a periodic manner. The last capability of the control unit is useful when the transient behavior of the PLL is examined. In this case, the control voltage is observed by Oscilloscope 2. The periodic repetition of the transients makes it possible to see almost immediately the results caused by the changes in the loop filter and also enables the use of oscilloscopes without a storage capability.

III. Suggested Laboratory Session Plan and Demonstrations

A. Steady state of the PLL

First, the students are asked to compose an equation based on the laboratory kit diagram, in which the synthesized frequency is unknown, and to solve it in order to obtain the relationship between the division ratio *N* and the synthesized frequency. Based on the result they should also determine the frequency step of the synthesizer.

Next, the laboratory kit is powered up with the switches in the positions depicted in Fig. 1. The control unit automatically sets the default frequency division ratio *N1*=2000. Then the students are asked to check whether the frequency counter readout agrees with the theoretical prediction.

Further on, the students are asked to look at Oscilloscope 1 and to say if the frequencies of the phase detector inputs are really equal to each other and if it is possible to detect a very small frequency difference (if there is such). Then the teacher can make the following demonstration: He sets S3 in the upper position (so that the automatic frequency control is terminated) and one of the two oscillograms begins to move. After that, he tries to stop it by adjusting the VCO FRF, which fails, making it clear that even the slightest frequency difference can be detected as an oscillogram motion. Then he restores the control loop and the oscillogram freezes. Now it is the proper time to draw the students' attention to the fact that the steady state frequency error in PLLs is zero [8], [1]. This

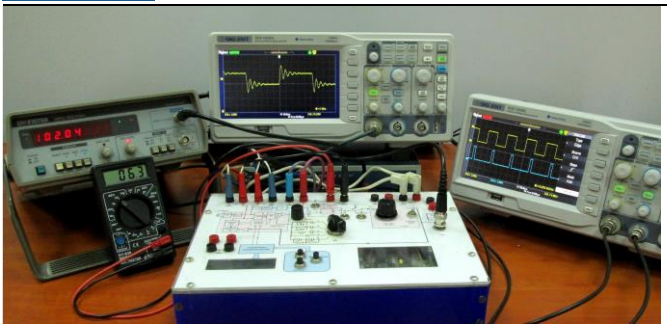


Fig. 2. The proposed kit in use (at the moment of taking the photo, the variable resistor in the loop filter is not adjusted to its optimal value.

is especially necessary if the students are not trained in control system theory. Often, when learning negative feedback systems in Signal and Systems course, telecommunication students do not consider them in an automatic control perspective. Hence, they do not readily recognize that there is a non-zero steady state error in a typical negative feedback control system and are not impressed by the zero steady state error in the PLLs. Then the teacher can explain that the zero steady state frequency error of PLLs, combined with the impeccable maintaining of the intended frequency division ratios of the digital frequency dividers, makes the output frequency strictly related to the reference oscillator frequency.

Then the students examine the dependency of the output frequency and the control voltage on the VCO FRF with a fixed frequency division ratio *N*. The FRF is varied approximately from 70 to 130 MHz. In order to measure the VCO FRF, the switch S3 should be set in the upper position. (For the purpose of this laboratory session, "VCO FRF" should be understood as the VCO frequency when the control voltage is equal to a fixed reference value.) The relationships are graphically presented. The students are asked to designate on the graphs the VCO FRF range, corresponding to the in-lock condition and to explain briefly the behavior of the PLL system.

The next task is to determine the synthesized frequency range of the synthesizer, provided that the VCO FRF by design is 100 MHz for example, which under normal operating conditions can change within a certain range, e. g. ±1 MHz. Based on the findings in the previous task, the students should be able to deduce what the worst conditions for the synthesis of the minimum and maximum frequencies are. Then, after setting the corresponding VCO FRFs they should pick out experimentally the minimum and maximum *N* values, for which the in-lock condition is still maintained. For the minimum and maximum synthesized frequencies, the corresponding control voltages should also be written down in the lab record. Then the students should propose ways for modifying (widening/narrowing and translation up/down) the synthesized frequency range of a PLL synthesizer.

Finally, the students should find the synthesized frequency errors for a few frequency division ratios and identify the cause for these errors. It is now the appropriate time to remember that PLLs themselves have zero steady state frequency error, that the digital frequency dividers strictly maintain the intended division ratios and that the output frequency is strictly related to the reference frequency. Based on these considerations, the students are expected to conclude that the error is caused by the reference frequency inaccuracy.

Steady state phase errors. Although in the context of frequency synthesis, the phase error has little significance, it can be important for other PLL applications. The students can change the frequency division ratio *N* and the VCO FRF and observe the steady state phase error for the five S4 positions on Oscilloscope 1. They can see that the phase error is also zero (besides the zero frequency error) for positions 1-4 [1]. For position 5, there is a nonzero phase error which changes depending on the VCO FRF and the *N* value.

B. Dynamical behavior

The **transient behavior** can be examined first.

The teacher should outline the reasons why transient behavior is important, especially if the students are not familiar with the concepts of control theory [8]. It should be emphasized, among other things, that the look of the transient response is informative about the stability margins of the system [1]. Then the control unit should be put in the "Transient" mode and a second frequency division ratio *N2* should be set (for example *N2*=2050 if *N1*=2000). This ensures an automatic alternation of the programmable counter division ratio between *N1* and *N2* in a periodic manner. Then the teacher explains the oscillogram on Oscilloscope 2: What part of the oscillogram corresponds to a steady state, in what moment the *N* value is changed, and (approximately) when the corresponding new frequency settles.

Then the students should examine the PLL transient response for the two charge pump currents with S4 in position 1 when the resistor value is changed. They should sketch out the transient responses for the optimum and for one markedly inappropriate value of the resistor for each charge pump current setting and write down their durations. Transient responses for positions 2 and 3 should also be examined for each charge pump current setting. Finally, the obtained transient response durations should be compared and conclusions should be drawn and shortly written in the lab record.

Next, the transient response for position 4 should be observed. It should be sketched and its duration should be measured and written down. At this point the students should be able to explain what the function of the loop filter components in positions 1, 2 and 3 from theoretical and from practical point of view is.

A stable vs. an unstable loop. In the next experiment the frequency division ratio is fixed and an additional low-pass filter is included in the loop. It could presumably further reduce the reference spurs in the output oscillation. In reality, the PLL becomes unstable. The students have studied about stability in mathematics, circuit theory and analog electronics. However they are often not aware of the exceptional practical significance of this concept, erroneously thinking that it relates to the consistency of the performance or endurance. Therefore, it is necessary that they observe an unstable system at least once in order to realize that a system is not operational at all if it is not stable. Additionally, it should be explained

that PLLs are inherently prone to instability as a result of the accumulated phase lags caused by the integration (which models the VCO behavior with respect to the phase in the linear PLL model [1]) and by the low-pass filter in the loop.

Acquisition. When the phase-frequency detector is used, it greatly facilitates the acquisition, as a result of which the corresponding phenomena cannot be observed. When the plain phase detector is used (S1 in the upper position), the typical beat frequency oscillation in the control voltage can be seen on Oscilloscope 2. Initially, the VCO FRF is changed (up or down) until the PLL unlocks. Then the VCO FRF is slowly changed towards approaching the acquisition range. It can be observed that the beat frequency decreases, the voltage swing increases and finally the PLL locks.

Examining the reference spurs in the VCO oscillation. Although a spectrum analyzer would be useful for this purpose, the reference spurs can be roughly estimated by measuring the reference component in the control voltage. It would be advantageous in this case, if Oscilloscope 2 has a good Y channel sensitivity, e. g. 1 mV per division or even less. With the Y input AC coupled and a sufficiently high Y gain setting, the reference component in the control voltage should be evaluated qualitatively and quantitatively for the 5 positions of S4. Then the reference spur level can be roughly estimated using the VCO sensitivity.

A special attention should be paid to the reference components in position 5, when the loop filter is a usual low-pass filter which no longer retains the steady state control voltage. This case is especially important for understanding the benefits of using a PD output configuration with a charge pump.

Frequency modulation capability examination. In this case an LF sine generator and a modulation meter should be connected according to Fig. 1. The sine magnitude is chosen so as to obtain an output frequency deviation of a few tens of kilohertz when the frequency is of the order of kilohertz. Then the dependency of the frequency deviation on the modulation frequency is examined. Alternatively the control voltage swing can be measured on Oscilloscope 2 if a modulation meter is not available. In this case, it would be advantageous if the modulation signal is also observed on the second Y channel. For sufficiently low modulation frequencies, the loop succeeds to track the VCO frequency variations and produces corresponding control voltage variations which suppress the modulation. When the modulation frequency rises enough, the control voltage swing decreases and the loop no longer counteracts the modulation successfully.

This experiment also makes possible the evaluation of the transfer function of the PLL for the VCO phase noise components if some additional calculations are performed.

C. PLL ranges.

PLL ranges [1] can be examined by changing the VCO FRF. When the PFD is used, there is no difference between the hold-in and the acquisition range from the observer's point of view: As a result of the sensitivity of the PFD to the frequency differences, the VCO frequency is pulled towards its in-lock value and the acquisition happens readily. When the plain PD is used, the students can observe a fairly large difference between the hold-in range and the acquisition range. Additionally, they can observe in this case an inclination of the PLL to false lock conditions where the frequencies on the PD inputs are related by ratios of small integers.

D. Theory vs. experiment.

It would be very instructive in an advanced course if the students compose the linear model of the loop and calculate the building block parameters based on the laboratory kit data. The VCO sensitivity can be easily evaluated experimentally. Then the students can predict the loop behavior theoretically (or by simulations) and compare the predictions with the experimental results.

IV. CONCLUSION

A laboratory kit for an in-depth study of PLLs was designed and implemented. The experience gathered during its use showed that the intended goals were reached. The operating capabilities of the kit satisfy the requirements of the undergraduate curriculum in Telecommunication Engineering and also offers additional opportunities for further studying PLLs for students with special interest in the area of radio communication devices.

REFERENCES

[1] Egan, W. F., "Phase-Lock Basics", John Wiley & Sons, 2008.

[2] Chenakin, A., "Frequency Synthesis: Current Status and Future Projections", Microwave Journal, April 2017

[3] Analog System Lab Kit PRO Manual, Texas Instruments, June 2012.

[4] The University Of British Columbia, Department of Electrical and Computer Engineering, ELEC 391 – Electrical Engineering Design Studio II, "Lab Assignment 3 – Phase-Locked Loops", Summer 2017. [Online]. Available: http://courses.ece.ubc.ca/elec391/Lab3-PLL.pdf (Accessed 09-March-2018).

[5] ADF4002 datasheet, Analog Devices, Last Content Update: 23 Feb. 2017.

[6] Rohde, U. L., A. K. Poddar, "Phase Noise Measurements and System Comparisons", Microwave Journal, April 2013.

[7] Gheen, K., Agilent Technologies, "Phase Noise Measurement Methods and Techniques", Aerospace & Defense Symposium, 2012.

[8] Åström, K. J., R. M. Murray "Feedback Systems. An Introduction for Scientists and Engineers", Princeton University Press, 2008.

AUTHOR INDEX